Online Appendices for “Treatment Effects in Bunching Designs: The Impact of Mandatory Overtime Pay on Hours”

Leonard Goff

Last updated: May 13, 2025

Contents

B 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 also allowing a distinction that is in some cases useful (e.g. in the overtime setting, models in which hours of pay for work differ from actual hours of work).

B.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 B.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 $B_i(\mathbf{x}) := \max\{B_{0i}(\mathbf{x}), B_{1i}(\mathbf{x})\}$ denote the actual constraint function that applies to z. A

[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.

budget constraint like Eq. $z \geq B_i(\mathbf{x})$ is typically "kinked" because while the function $B_i(\mathbf{x})$ 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 I 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.

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 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 I.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. B.3), bunching at 40 hours will be informative about the effect of

[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 $B_i(x)$ 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 I.5 for details).

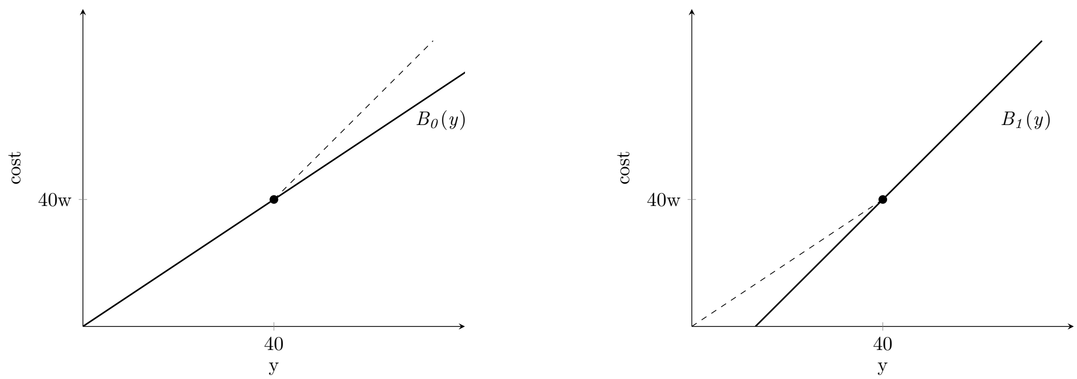

Figure 1: Definition of counterfactual cost functions B_0 and B_1 that firms could have faced, absent the overtime kink. Regardless of what choice variables are in $\mathbf{x}$, these functions only depend on $y_i(\mathbf{x})$ and are thus depicted as a function of y. Dashed lines show the rest of actual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

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 I.5).

B.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 $Y_{B_0 i}$ and $Y_{B_0 i}$, respectively. In this notation, the actual outcome Y_i observed by the econometrician is equal to $Y_{B_i i}$.

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.

Note: Some departures from CONVEX are allowable without compromising it's main implication for the bunching-design, which is given in Lemma B.1 below. If B_0 and B_1 are linear in $\mathbf{x}$ and the constraint $z \geq B(\mathbf{x})$ 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 F). 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

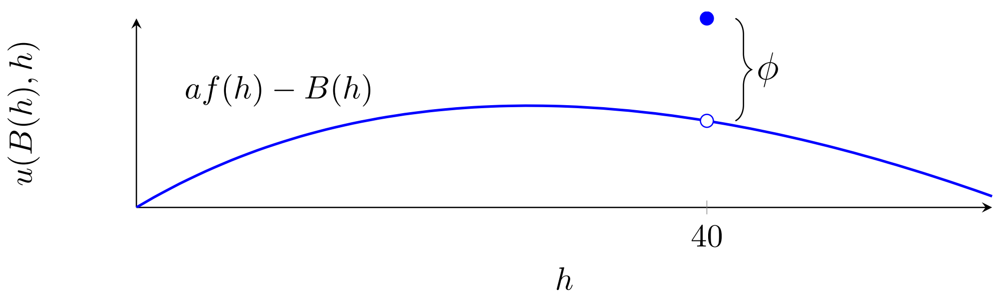

Figure 2: An example of preferences that satisfy CONVEX but are not strictly convex, cf. Eq. (2).

with an assumption that u_i are strictly decreasing in z (see Eq. (7) in the Proof of Lemma B.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 G 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 B.1 and B.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").

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

B.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 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 $z = B(h)$, 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

[5] Appendix F 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 B 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.

$\rho < 1 + 1/\epsilon$, 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.

B.4 Observables in the kink bunching design

Lemma B.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 B.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 A. □

Lemma B.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 B.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 B.1 gives the relationship between this bunching probability and treatment effects, which holds in a weakened form when CONVEX is replaced by WARP:

Proposition B.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 D. □

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 B.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:

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

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

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

Proof. See Appendix D. □

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 B.1 and B.2 we have that $\Delta \geq \mathcal{B}/f_0(k)$ under WARP and CHOICE.

B.5 Treatment effects in the bunching design

Proposition B.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 I.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 B.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

[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 $log(h)$.

(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.

Additional identification results for the bunching design:

While Theorem 1 of Section 4 develops the treatment effect identification result used to evaluate the FLSA, Appendix I 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 I 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.

Assumption RANK and the sign of treatment effects:

Another important point regarding Assumption RANK is that it does not require $Y_{0i} \geq Y_{1i}$ for all units i. Figure 3 shows an example in which $Y_1 = 2Y_0 - k$, so that $Y_1 < Y_0$ when $Y_0 < k$ and $Y_1 > h_0$ when $Y_0 > k$. For simplicity there is no bunching at the kink in this example, provided that Y_0 has a continuous marginal distribution around k. Note that from Lemma B.1, we can write $\mathcal{B} = P(Y_1 \leq k, Y_1 \leq Y_0) - P(Y_0 \leq k, Y_1 \leq h_0)$,[10] which when combined with $\mathcal{B} = P(Y_1 \leq k) - P(Y_0 \leq k)$ (c.f. Eq. (6) in the main text) in turn implies that

$$P(Y_0 \leq k, Y_1 > Y_0) = P(Y_1 \leq k, Y_1 > Y_0) \quad (4)$$

[10] I thank an anonymous referee for pointing out this alternative expression for $\mathcal{B}$, which holds provided that (Y_0, Y_1) is absolutely continuous.

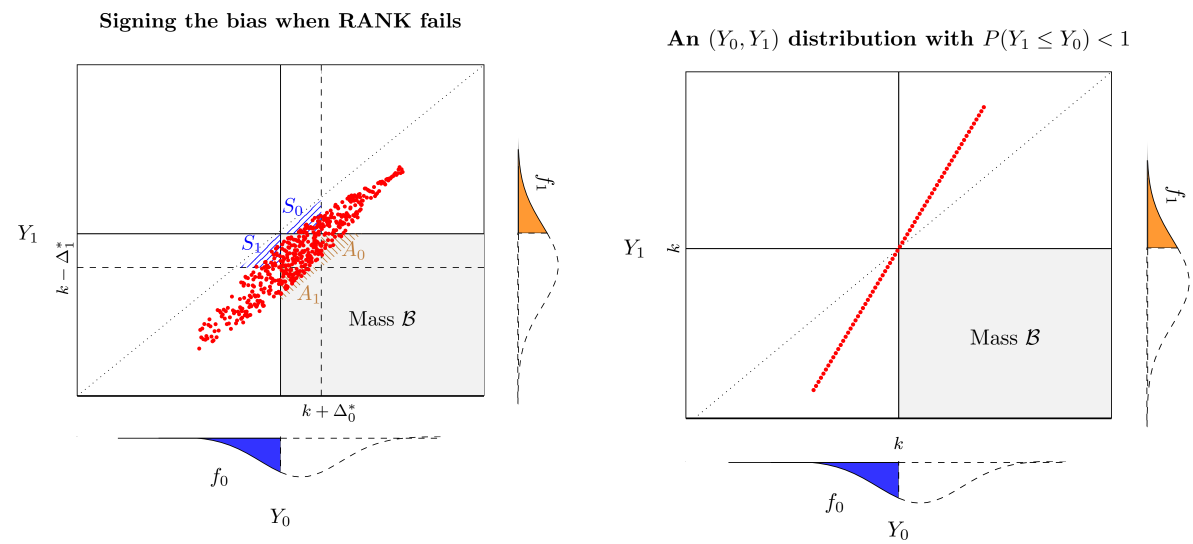

Figure 3: *Left:* When Assumption RANK fails, the average $\mathbb{E}[Y_{0i}|Y_{0i} \in [k, k+\Delta_0^*]]$ will include the mass in the region S_0, who are not bunchers (NE lines) but will be missing the mass in the region A_0 (NW lines) who are. This causes an under-estimate of the desired quantity $\mathbb{E}[Y_{0i}|Y_{1i} \leq k \leq Y_{0i}]$. 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 $\mathbb{E}[Y_{1i}|Y_{1i} \leq k \leq Y_{0i}]$. *Right:* A distribution of (Y_0, Y_1) such that $P(Y_0 > Y_1) > 0$ and $P(Y_1 > Y_0) > 0$. The thin black dots reflect the 45 degree line. Note that while there is mass on either side of the 45 degree line, there is no mass in the NW quadrant of the figure, which would violate Assumption CONVEX.

This implication is certainly satisfied if $Y_0 \geq Y_1$ with probability one, since then both sides are equal to zero. This is the case in for example the isoelastic model, given a positive elasticity. More generally however, Eq. (4) simply says that the mass above the 45 degree line in the western half of Figure 3 is equal to the mass above the 45 degree line in the southern half of it. Any joint distribution for which there is no mass in the NW quadrant—consistent with assumption CONVEX—will satisfy (4), for example the case depicted in Figure 3.

B.6 Identification of the buncher ATE by polynomial extrapolation

Following the discussion in Section 4.3.4 let us assume that for each $d \in \{0,1\}$, $Q_d^{(m+1)}(u)$ exists for some m where $Q_d^{(n)}$ denotes the n^{th} derivative of $Q_d(u)$. We begin with the case without counterfactual bunchers for ease of notation, but we will see that extending to that case is straightforward.

Assume that $\sup\{|Q_d^{(m+1)}(u)| : u \in [F_0(k), F_1(k)]\} \leq M$ for some constant M, where Q_d^m is the m^{th} derivative of Q_d. Consider first $d=0$. Recall that since $\mathcal{B} = F_1(k) - F_0(k)$, we can write the bunching region as $[F_0(k), F_0(k)+\mathcal{B}]$. By Taylor's theorem we have for any $v \in [0,\mathcal{B}]$ that:

$$Q_0(F_0(k)+v) = \sum_{n=0}^{m} Q_0^{(n)}(F_0(k))) \cdot \frac{v^n}{n!} + Q_0^{(n+1)}(F_0(k))) \cdot \frac{v^{m+1}}{(m+1)!}$$

for some $u^* \in [F_0(k), F_0(k)+v]$. By the Lipschitz assumption on $Q_0^{(n+1)}$, this implies:

$$Q_{0,n}(v) - \frac{v^{m+1}}{(m+1)!}M \leq Q_0(F_0(k)+v) - k \leq Q_{0,n}(v) + \frac{v^{m+1}}{(m+1)!}M$$

where $Q_{0,n}(v) := \sum_{n=1}^{m} Q_0^{(n)}(F_0(k))) \cdot \frac{v^n}{n!}$ and I've used that $Q_0(F_0(k)) = k$. Similarly for $d=1$, for any $v \in [0,\mathcal{B}]$: $Q_{1,n}(v) - \frac{v^{m+1}}{(m+1)!}M \leq Q_1(F_1(k)-v) - k \leq Q_{1,n}(v) + \frac{v^{m+1}}{(m+1)!}M$ where $Q_{1,n}(v) := \sum_{n=1}^{m} Q_1^{(n)}(F_1(k))) \cdot \frac{(-v)^n}{n!}$.

The buncher ATE is:

$$\Delta_k^* = \frac{1}{\mathcal{B}}\int_{F_0(k)}^{F_1(k)} \{Q_0(u) - Q_1(u)\} \cdot du = \frac{1}{\mathcal{B}}\int_0^{\mathcal{B}} \{Q_0(F_0(k)-v) - Q_1(F_1(k+v)\} dv \in [\Delta_k^{LM}, \Delta_k^{UM}]$$

where $\Delta_k^{UM} = \frac{1}{\mathcal{B}}\int_0^{\mathcal{B}}\{Q_{0,n}(v) - Q_{1,n}(v) + 2\frac{v^{m+1}}{(m+1)!}M\} \cdot dv$ and $\Delta_k^{LM} = \frac{1}{\mathcal{B}}\int_0^{\mathcal{B}}\{Q_{0,n}(v) - Q_{1,n}(v) - 2\frac{v^{m+1}}{(m+1)!}M\} \cdot dv$. Meanwhile:

$$\begin{aligned}
\frac{1}{\mathcal{B}}\int_0^{\mathcal{B}} \{Q_{0,n}(v) - Q_{1,n}(v)\} \cdot dv &= \sum_{n=1}^{m} \{Q_0^{(n)}(F_0(k))) + (-1)^{n+1} \cdot Q_1^{(n)}(F_1(k)))\} \cdot \frac{1}{\mathcal{B}}\int_0^{\mathcal{B}} \frac{v^n}{n!} \cdot dv \\
&= \sum_{n=1}^{m} \{Q_0^{(n)}(F_0(k))) + (-1)^{n+1} \cdot Q_1^{(n)}(F_1(k)))\} \cdot \frac{1}{\mathcal{B}} \left\{ \left.\frac{v^{n+1}}{(n+1)!}\right|_0^{\mathcal{B}} \right\} \\
&= \sum_{n=1}^{m} \frac{\mathcal{B}^n}{n+1!} \cdot \{Q_0^{(n)}(F_0(k))) + (-1)^{n+1} \cdot Q_1^{(n)}(F_1(k)))\} \cdot
\end{aligned}$$

and the remainder term is $\frac{1}{\mathcal{B}}\int_0^{\mathcal{B}} 2\frac{v^{m+1}}{(m+1)!}M \cdot dv = \frac{2M}{\mathcal{B}(m+1)!}\int_0^{\mathcal{B}} v^{m+1} = \frac{2M}{(m+1)!} \cdot \left\{ \left.\frac{v^{m+2}}{(m+2)}\right|_0^{\mathcal{B}} \right\} = M\frac{2\mathcal{B}^{m+1}}{(m+2)!}$.

Using that $Q_d^{(1)}(u) = \frac{1}{f_d(Q_d(u)))}$ he derivatives $Q_d^{(n)}(F_d(k))$ can be worked out from the density of f_d and it's derivatives at the kink:

$$Q_d^{(1)} = \frac{1}{f_d(k)} \qquad Q_d^{(2)} = -\frac{f_d'(k)/f_d(k)}{f_d(k)^2} \qquad Q_d^{(3)} = \frac{3(f_d'(k)/f_d(k))^2 - (f_d''(k)/f_d(k))}{f_d(k)^3} \qquad \text{etc.}$$

Notice that these derivatives take the form of the ratio of a numerator that is invariant to the overall "scale" of f_d divided by a denominator of $f_d(k)^n$. We can rewrite $\frac{1}{\mathcal{B}}\int_0^{\mathcal{B}}\{Q_0(v) - Q_1(v)\} \cdot dv$ as

$$\begin{aligned}
&\frac{1}{2}\left(\frac{\mathcal{B}}{f_0(k)}\right) - \frac{1}{3!}\left(\frac{\mathcal{B}}{f_0(k)}\right)^2 \left\{\frac{f_0'(k)}{f_0(k)}\right\} + \frac{1}{4!}\left(\frac{\mathcal{B}}{f_0(k)}\right)^3 \left\{3\left(\frac{f_0'(k)}{f_0(k)}\right)^2 - \frac{f_0''(k)}{f_0(k)}\right\} \\
&\quad + \frac{1}{2}\left(\frac{\mathcal{B}}{f_1(k)}\right) + \frac{1}{3!}\left(\frac{\mathcal{B}}{f_0(k)}\right)^2 \left\{\frac{f_0'(k)}{f_1(k)}\right\} + \frac{1}{4!}\left(\frac{\mathcal{B}}{f_1(k)}\right)^3 \left\{3\left(\frac{f_1'(k)}{f_1(k)}\right)^2 - \frac{f_1''(k)}{f_1(k)}\right\} + O(\mathcal{B}^4)
\end{aligned}$$

If there is counterfactual bunching, then conditioning on the $K^*=0$ subsample introduces a common factor of $1/(1-p)$ for all densities and density derivatives, as well as for the conditional bunching probability $\mathcal{B} \to \frac{\mathcal{B}-p}{1-p}$ (see proof of Theorem 1 for details). The factor of $1/(1-p)$ cancels out everywhere in the above expression so one can simply replace $\mathcal{B}$ by the net bunching $\mathcal{B}-p$ to accommodate counterfactual bunching $p = P(K^*=1)$. Our central estimate of the buncher ATE then becomes:

$$\frac{1}{2}\left\{\left(\frac{\mathcal{B}-p}{f_0(k)}\right)+\left(\frac{\mathcal{B}-p}{f_1(k)}\right)\right\}-\frac{1}{3!}\left\{\left(\frac{\mathcal{B}-p}{f_0(k)}\right)^2\frac{f_0'(k)}{f_0(k)}-\left(\frac{\mathcal{B}-p}{f_0(k)}\right)^2\frac{f_1'(k)}{f_1(k)}\right\}$$
$$+\frac{1}{4!}\left\{\left(\frac{\mathcal{B}-p}{f_0(k)}\right)^3\left(3\left(\frac{f_0'(k)}{f_1(k)}\right)^2-\frac{f_0''(k)}{f_0(k)}\right)+\left(\frac{\mathcal{B}-p}{f_1(k)}\right)^3\left(3\left(\frac{f_1'(k)}{f_1(k)}\right)^2-\frac{f_1''(k)}{f_1(k)}\right)\right\}+O((\mathcal{B}-p)^4)$$

This expression makes clear that this approach is more accurately thought of as an expansion in powers of $(\mathcal{B}-p)/f_0(k)$ and $(\mathcal{B}-p)/f_1(k)$ than it is over the bunching probability $\mathcal{B}-p$.

B.7 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:

$$B_i(y;\rho,k) = \rho w_i y - k w_i(\rho-1) \tag{5}$$

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 $kw_i(\rho-1)$ that keeps $B_i(k;\rho,k)$ invariant as ρ is changed.

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 (5) satisfies these requirements with $y_i(h) = h$.[11]

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 $z \geq B_i(\mathbf{x};\rho,k)$. A natural restriction in the overtime setting that is that the function $Y_i(\rho,k)$ does not depend on k, and some of the results below will

[11] As an alternative example, I construct in Appendix I.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.

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 (5), 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.*

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 I.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 B.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.[12]

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 B.2 (bunching expressed in terms of marginal responsiveness). *Assume CHOICE, SMOOTH and WARP. Then:*

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

with equality under CONVEX.

Proof. See Appendix D. □

The main tool in establishing Lemma B.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:*

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

[12] Blomquist et al. (2021) derive the special case of Lemma B.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.

Proof. See Appendix D. □

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 Proposition I.4) 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:

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

By Lemma B.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 B.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 B.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 B.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 B.2 and B.3 together imply Theorem 2 (see Appendix A for proof).*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 $\frac{dY_i(\rho)}{d\rho}$.

C Motivating the bi-log-concavity assumption

C.1 As an extrapolation assumption

The polynomial estimation approaches of Saez (2010) and Chetty et al. (2011) can be thought of as extrapolating the exact curve of a polynomial fit to the observable distribution to point identify ϵ, in the iso-elastic model. An alternative is to extrapolate features of the observed density without extrapolating it's exact functional form. For example, Bertanha et al. (2023) proposes computing the maximum derivative of the density of $\ln(h)$ for $h \neq k$ and assuming that this Lipshitz bound also holds across the missing region depicted in 4. Similarly, Blomquist et al. (2021) propose bounding the *level* of the density of $\ln(h)$ to be within the convex hull of the left and right limits of the density of $\ln(h)$, expanded by a specified constant σ. Taking $\sigma = 1$ nests the non-parametric shape constraint of imposing monotonicity of the density of $\ln(h)$.

The logic of verifying BLC of h_0 to the left of the kink to motivate BLC of h_0 across the unobserved region $[k, k + \Delta_0^*]$ (and analagously for h_1, looking to the right of the kink) as demonstrated in Figure 1, can be described in similar terms. Focusing on the case with no counterfactual bunchers ($p = 0$) for simplicity, Theorem 1 assumes that $\ln F_0(h)$ and $\ln(1 - F_0(h))$ are both concave on the interval $[k, k + \Delta_0^*]$. Assuming this CDF is twice differentiable, this is equivalent to:

$$\sup_{h \in [k, k+\Delta_0^*]} \max\{d^2/dh^2 \ln F_0(h), d^2/dh^2 \ln(1 - F_0(h))\} \leq 0 \tag{6}$$

If h_0 is in fact BLC to the left of the kink, then

$$\sup_{h<k} \max\{d^2/dh^2 \ln F_0(h), d^2/dh^2 \ln(1 - F_0(h))\} \leq 0$$

and hence a sufficient condition for (6) is that

$$\begin{aligned}\sup_{h \in [k, k+\Delta_0^*]} & \max\{d^2/dh^2 \ln F_0(h), d^2/dh^2 \ln(1 - F_0(h))\} \\ & \leq \sup_{h<k} \max\{d^2/dh^2 \ln F_0(h), d^2/dh^2 \ln(1 - F_0(h))\}\end{aligned}$$

Similarly, if BLC of h_1 is verified for values to the right of the kink, then a sufficient condition for the assumption required by Theorem 1 is that

$$\begin{aligned}\sup_{h \in [k-\Delta_1^*, k]} & \max\{d^2/dh^2 \ln F_1(h), d^2/dh^2 \ln(1 - F_1(h))\} \\ & \leq \sup_{h>k} \max\{d^2/dh^2 \ln F_1(h), d^2/dh^2 \ln(1 - F_1(h))\}\end{aligned}$$

In this way, the BLC assumptions made by Theorem 1 can be thought of as extrapolating the extreme value of a property (or properties) of the distribution of F_d from a region in which that property is observed, to an unobserved region corresponding to the bunchers. While Blomquist et al. (2021) extrapolates the maximum/minimum levels of the density right next to the kink, and Bertanha et al. (2023) the magnitude of it's derivative across all point away from the kink, a sufficient condition for my result in Theorem 1 is to extrapolate the maximum of the second derivative of both $\ln F_d$ and $\ln(1 - F_d)$, for each of $d \in \{0, 1\}$.

C.2 Bi-log-concavity in terms of hazard functions

The partial identification result of Theorem 1 hinges on the assumption that the distribution of counterfactual hours h_{0it} and h_{1it} (among units it that are not counterfactual bunchers) are both bi-log-concave (BLC). In this section, I decompose this assumption into two parts and describe how each part arises naturally as a property of the distribution of working hours.

Consider a random variable with CDF $F(h)$ admitting of a density $f(h)$. BLC is equivalent to the following:

1. the hazard rate function $f(h)/(1-F(h))$ is (weakly) increasing in h
2. the reverse hazard rate function $f(h)/F(h)$ is (weakly) decreasing in h

These can be derived by observing that the derivative of $\log F(h)$ is $f(h)/F(h)$ and that $-f(h)/(1-F(h))$ is the derivative of $\log(1-F(h))$—see Dümbgen et al. (2017) for details. Note that while BLC is introduced in Section 4 as a property applying to the whole support of a random variable, the Theorem 1 bounds on the buncher ATE only in face require these properties to hold for h_0 on the interval $[k, k+\Delta_0^*]$ and on the interval $[k-\Delta_1^*, k]$ for h_1 (as described in the proof).

The property of an increasing hazard rate arises in reliability theory, which often models the aging properties of a system over time. Consider a very simple model in which workers continue working until they "fail" at some stochastic number H of work hours. The hazard rate $f(h)/(1-F(h))$ then captures the probability that the worker fails after h hours given that they have not failed yet ($H > h$). That the instantaneous probability of failure for a system of age h would be increasing in h is a natural notion of *wear* (e.g. of a machine), and is often referred to as the *increasing failure rate* or IFR property (Barlow et al., 1996). While we might view the above as a model of worker fatigue that manifests as a dichotomous notion of "failure", the next section shows how IFR also emerges in a more realistic model in which worker productivity declines gradually and stochastically over time.

The second aspect of bi-log-concavity is that the reversed hazard rate $f(h)/F(h)$ is weakly decreasing in h, referred to as a *decreasing reverse hazard rate* or DRHR. Block et al. (1998) show that any non-negative random variable must be DRHR at least somewhere in its support. The BLC assumption made by Theorem 1 requires something stronger: that counterfactual hours be DRHR (in addition to being IFR) across a *particular* region near the kink. Like IFR, one can characterize DRHR in terms of failure times: DRHR holds iff the time $h-H$ that has elapsed by some moment h since failure at H (given that $H \leq h$) is increasing in h in the sense of stochastic dominance (Gupta and Nanda, 2001). This is an intuitive property, but could fail to hold if the density of H increases too rapidly at some h. The model in the next section provides primitive conditions that rule this out.[13]

[13] Another way to motivate the DRHR property of work hours is to note that if there are T total hours in a week, the number of non-work hours L is $T-H$, and to observe that H is DRHR if and only if L is IFR.

While the above considerations may lend plausibility to the IFR and DRHR properties of counterfactual working hours h_0 and h_1 by giving them intuitive interpretations, they fall short of providing explicit sufficient "economic" conditions for them both. The next section does so, by modeling the working hours as chosen optimally according to a worker productivity that is generated by an underlying process with an assumed Markovian structure.

C.3 A model of hours with stochastic shocks to productivity

Recall Equation (**??**) of Section 4, which provides intuition for firms' optimal choices of hours in terms of the marginal product of an hour of labor from unit it. In a model in which firms maximize the net revenue $\pi_{it}(h) = f_{it}(h) - B_{kit}(h)$ from worker i in week t, where $f_{it}(\cdot)$ is a revenue production function with respect to hours, then counterfactual choices can be written as $h_{0it} = MPH_{it}^{-1}(w_{it})$ and $h_{1it} = MPH_{it}^{-1}(1.5w_{it})$. Here $MPH_{it}(h) = \frac{d}{dh}f_{it}(h)$ can be thought of as worker i's instantaneous hourly productivity at hour h within the week, and w_{it} is their straight wage.

Within a model of this form, we can motivate BLC of h_{1it} and h_{0it} among a set of ex-ante identical workers that experience different realizations of a common stochastic process generating the function f. Assume these workers share a straight wage $w_{it} = w$, and are not counterfactual bunchers in the language of Section 4.3 (thus conditioning on $K_{it}^* = 0$ will be kept implicit). Consider a single fixed week t which I suppress for now in the notation.

All workers have a common productivity $MPC_i(0) = p_0$ when they are "fresh" and have not yet worked any hours this week. At each moment in continuous time, a worker's hourly productivity either stays the same or drops by a discrete amount. Let $\{p_j\}_{j=0,1,\ldots}$ be a decreasing sequence that denotes hourly productivity after j productivity drops. This function of j is assumed common to all workers i.

We'll see that bi-log-concavity of $MPH_i^{-1}(w)$ for any w then follows when the timing of these drops has a simple Markovian structure. In particular, assume that the probability of j increasing by one in a small timespan around h hours depends only on j and is independent of h and the past trajectory of productivity. This is a reasonable assumption if what matters for the future evolution of worker fatigue is that worker's current level of fatigue, rather than how many hours they have been working so far *per-se*.

Since MPH_i is weakly decreasing in h for all i, we can define an inverse MPH function as $MPH_i^{-1}(w) = \inf\{h : MPH_i(h) \leq w\}$. The RHS of this expression is referred to as a *first-passage time*, a random variable whose distribution is often of interest in the reliability theory literature. We can understand the first passage time $MPH_i^{-1}(w)$ as the first time h that a worker's fatigue j has accumulated to j^*, the smallest j such that $p_j \leq w$. (i.e. it is no longer profitible for the worker to continue working at wage w).

Thus DRHR of H can be interpreted as saying that the failure rate of "leisure" is increasing: the probability that L lies in a infinitesimal neighborhood of ℓ, given that $L > \ell$, is an increasing function of ℓ.

Kijima (1998) shows that if a continuous-time Markov chain on the positive integers can only increase or decrease by one unit at a time, the distribution of first passage times from zero to any given level j^* satisfies both IFR and DRHR, and is hence BLC.[14] Recall that roughly speaking, these properties mean that the density of first passage times can neither rise nor fall too abruptly at any one point h. To get some intuition for this result, let $j_i(h)$ be the number of productivity drops worker i has received in the first h hours of work. Then the (time homogeneous) Markov property implies that transitions into state j^* satisfy:

$$P(j_i(h+s) = j^* | j_i(h) = j^* - 1) = s \cdot \lambda_{j^*-1} + o(s)$$

for some set of rate parameters λ_{j^*}, and any h. Since $j_i(h)$ must first pass through $j^* - 1$ to arrive at j^*, this implies that the density of first passage times evaluated at h is equal to $P(j_i(h) = j^* - 1) \cdot \lambda_{j^*-1}$. Since the factor λ_{j^*-1} is common to all h, the density of first-passage times can only have a "spike" or a "hole" at h if the function $P(j_i(h) = j^* - 1)$ has a corresponding spike or hole at h. But the Markov structure constrains the form of $P(j_i(h) = j^* - 1)$ in a way that rules this out (see e.g. Taylor and Karlin (1994)).

While results of Kijima (1998) show that the first passage times in this model satisfy both components of BLC, Keilson (1971) calculates the distribution of first-passage times explicitly. When a continuous-time Markov chain on the integers cannot increase or decrease by more than one unit at a time, it is referred to as a "birth-death" processes. Keilson (1971) shows that first passage times to j^* for birth-death processes are distributed as a convolution of j^* exponential densities. The resulting density is log-concave, a special case of BLC.[15]

Now let us bring back the index t for the week of paycheck unit it. In the spirit of Section 4 one might view the above as a model of *scheduled* hours that the firm chooses given worker at the beginning of week t, if the firm is aware of that worker's realization of the productivity process. This might be reasonable if the workers' production function f_{it} is the same each week t, and the firm is able to quickly learn it upon hiring the worker. Alternatively, we can view the above model as describing shocks to productivity that are not yet revealed to firms at the beginning of the week. At each moment *of each week*, workers receive a possible shock to productivity and the firm decides whether to keep the worker engaged in labor or

[14] Kijima (1998) predates the introduction of the term *bi-log-concavity* by Dümbgen et al. (2017), so he does not use this term.

[15] Technically, the results of both Kijima (1998) and Keilson (1971) require there to be a non-zero probability of the fatigue process decreasing by one unit from any state j, with some transition rate $\mu_j > 0$. The above model is instead a "pure-birth" process in which fatigue only ever increases. We can obtain the desired result with $\mu_j = 0$ however by considering a sequence of Markov processes characterized by downward transition rates $\mu_j^{(n)} > 0$ where $\lim_{n\to\infty} \mu_j^{(n)} = 0$. Since CDFs of corresponding first passage times $h^{(n)}$ are pointwise continuous functions of $\mu_j^{(n)}$, and BLC is preserved under convergence in distribution (Saumard, 2019), it follows that the distribution of hours in the pure-birth model is BLC. A similar construction can be used to accommodate productivity with continuous rather than discrete support, viewing the continuous diffusion process for productivity as the limit of a sequence of birth-death processes. See Keilson (1971) for details.

withdraw them for the week (after withdrawal productivity resets to p_0 for the beginning of week $t+1$). This simple optimal stopping problem admits of the same general solution Eq. (2) considered before, since productivity within the week is declining with probability one.[16]

In practice, few workers work for a single spell during a given week. However, the above model can also be construed as applying to hours h_d within a single "shift" of work occuring on day d. Suppose that after a worker is withdrawn from labor for the day, they rest and productivity resets to p_0 on day $d+1$. Owing to the Markovian property of productivity, the length of each spell h_d within a week will be independent of the others, and the total hours for the week $h = h_1 + h_2 + \ldots h_7$ is distributed as a convolution of log concave densities. Such a convolution is itself log-concave (Saumard and Wellner, 2014), and hence BLC.

D Additional proofs

For ease of exposition, these proofs use the notation of Appendix B, using i rather than it indices.

D.1 Proof of Lemma B.1

The proof proceeds in the following two steps:

i) First, I show that $h_{0i} \leq k$ implies that $h_i = h_{0i}$, and similarly $h_{1i} \geq k$ implies that $h_i = h_{1i}$. This holds under CONVEX but also under the weaker assumption of WARP.

ii) Second, I show that under CONVEX $h_i < k \implies h_i = h_{0i}$ and $h_i > k \implies h_i = h_{1i}$.

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

Proof of i): Let $\mathcal{X}_{0i} = \{\mathbf{x} : h_i(\mathbf{x}) \leq k\}$ and $\mathcal{X}_{1i} = \{\mathbf{x} : h_i(\mathbf{x}) \geq k\}$. If $h_{0i} \leq k$, then by CHOICE $\mathbf{x}_{B_{0i}}$ is in $\mathcal{X}_0$, where for any budget constraint B, $(z_{Bi}, \mathbf{x}_{Bi})$ are the choices the decision-maker would make under B. Since $B_i(\mathbf{x}) = B_{0i}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}_0$, it follows that $z_{B_{0i}i} \geq B_i(\mathbf{x}_{B_{0i}i})$, i.e. the decision-maker's choice under B_0 is feasible under the kinked budget constraint B. Note that $B_i(\mathbf{x}) \geq B_{0i}(\mathbf{x})$ for all $\mathbf{x}$. By WARP then $(z_{B_ii}, \mathbf{x}_{B_ii}) = (z_{B_{0i}i}, \mathbf{x}_{B_{0i}i})$. Thus $h_i = h_i(\mathbf{x}_{B_ii}) = h_i(\mathbf{x}_{B_{0i}i}) = h_{0i}$. So $h_{0i} \leq k \implies h_i = h_{0i}$. By the same logic we can

[16] Hence, facing a weakly convex pay schedule, the firm never has incentive to keep the worker engaged in labor beyond the point at which their hourly productivity first dips below the marginal hourly wage. This leads directly to Equation (2). In this setting, h_{0it} is understood as the hours that the firm would choose if the worker were paid their straight wage for all hours, but faced the same realization of stochastic productivity decline this week as they actually do (and similarly for h_{1it})

show that $h_{1i} \geq k \implies h_i = h_{1i}$.

Proof of ii): For any convex budget function $B(\mathbf{x})$, $(z_{Bi}, \mathbf{x}_{Bi}) = \text{argmax}_{z,\mathbf{x}} \{u_i(z, \mathbf{x}) \text{ s.t. } z \geq B(\mathbf{x})\}$. 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 $\{(z, \mathbf{x}) : z \geq B(\mathbf{x})\}$. Furthermore, by monotonicity of $u(z, \mathbf{x})$ in z we may substitute in the constraint $z = B(\mathbf{x})$ and write

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

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

$$u_i(B(\tilde{\mathbf{x}}), \tilde{\mathbf{x}}) \geq u_i(\theta B(\mathbf{x}) + (1-\theta)B(\mathbf{x}^*), \tilde{\mathbf{x}}) > \min\{u_i(B(\mathbf{x}), \mathbf{x}), u_i(B(\mathbf{x}^*), \mathbf{x}^*)\} = u_i(B(\mathbf{x}), \mathbf{x}) \tag{7}$$

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 $\mathbf{x} \neq \mathbf{x}^*$, if one draws a line between $\mathbf{x}$ and $\mathbf{x}^*$, the function $u_i(B(\mathbf{x}), \mathbf{x})$ is strictly increasing as one moves towards $\mathbf{x}^*$. When $\mathbf{x}$ is a scalar, this argument is used by Blomquist et al. (2015) (see Lemma 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 still holds without monotonicity of $u_i(z, \mathbf{x})$ in z. This is useful for Theorem 1* of Appendix F 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

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

where B_i is the actual, kinked budget constraint faced by i. Let $\mathbf{x}_{ki}$ be the unique maximizer of $u_{B_i i}(\mathbf{x})$, where $h_i = h_i(\mathbf{x}_{ki})$. Suppose that $h_i < k$. Suppose furthermore that $h_{0i} \neq h_i$, with $h_{0i} = h_i(\mathbf{x}_{0i})$ and $\mathbf{x}_{0i}$ the maximizer of $u_{B_{0i} i}(\mathbf{x})$. Note that we must have that $\mathbf{x}_{0i} \notin \mathcal{X}_{0i}$, because $B_{0i} = B_i$ in $\mathcal{X}_{0i}$ so we can't have $u_{B_{0i} i}(\mathbf{x}_{0i}) > u_{B_{0i} i}(\mathbf{x}_{ki})$ (since $\mathbf{x}_{ki}$ maximizes $u_{B_i}(\mathbf{x})$). Thus $h_{0i} > k$.

By continuity of $h_i(\mathbf{x})$, $\mathcal{X}_{0i}$ 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 $h_i \neq k$ and $h_{0i} \neq k$, Eq. (7) then implies that $u_{B_i i}(\tilde{\mathbf{x}}) > u_{B_i i}(\mathbf{x}_{0i})$. Since $u_{B_{0i} i}(\mathbf{x}) = u_{B_i i}(\mathbf{x})$ for all $\mathbf{x}$ in $\mathcal{X}_{0i}$, it follows that $u_{B_{0i} i}(\tilde{\mathbf{x}}) > u_{B_{0i} i}(\mathbf{x}_{0i})$. However, this contradicts the premise that $\mathbf{x}_{0i}$ maximizes $u_{B_{0i} i}(\mathbf{x})$. Thus, $h_i < k$ implies $h_i = h_{0i}$. Figure 4 depicts the logic visually. The proof that $h_i > k$ implies $h_i = h_{1i}$ is analogous.

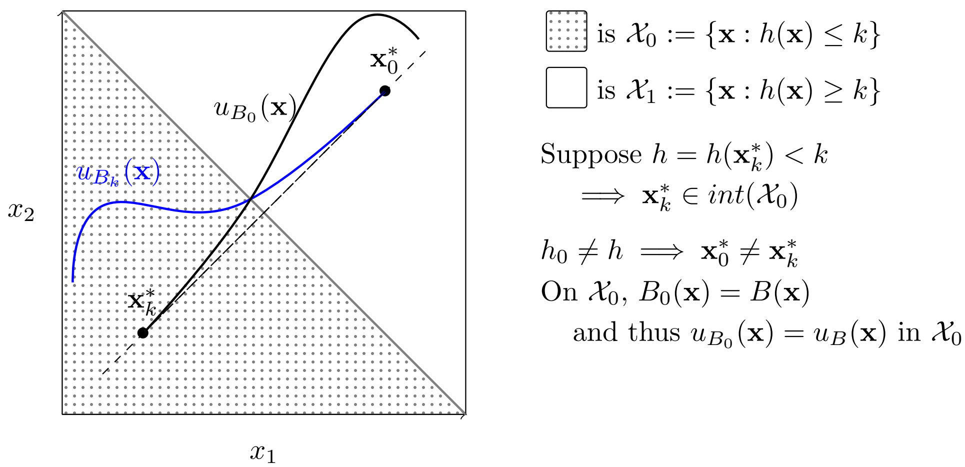

Figure 4: Depiction of the step establishing $(h < k) \implies (h = h_0)$ in the proof of Lemma B.1. Since the result considers a single decision-maker i, I supress this index in the Figure. In this example $z = (x_1, x_2)$ and $y(\mathbf{x}) = x_1 + x_2$. Proof is by contradiction. If $h_0 \neq Y$, then $\mathbf{x}_k^* \neq \mathbf{x}_0^*$, where $\mathbf{x}_k^*$ and $\mathbf{x}_0^*$ are the unique maximizers of $u_B(\mathbf{x})$ and $u_{B_0}(\mathbf{x})$, respectively. By Equation 7, 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 $\mathbf{x}_k^*$. By the assumption that $h < k$, then using continuity of $h(\mathbf{x})$ 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 $h(\mathbf{x}) \leq k$. 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 are equal, and thus so must be the functions u_B and u_{B_0}. Since the same function cannot be both strictly increasing and strictly decreasing, we have obtained a contradiction.

D.2 Sharpness of the bounds in Theorem 1

To see that the bounds $[\Delta_k^L, \Delta_k^L]$ from Theorem 1 are sharp, we need to show that for any $\Delta \in [\Delta_k^L, \Delta_k^L]$, there exists a distribution of potential outcomes consistent with the data and the assumptions of the theorem, for which Δ_k^* is equal to Δ. For simplicity, let us first consider a $\Delta \in \{\Delta_k^L, \Delta_k^L\}$, before considering intermediate values. The approach below bears some similarity to the sharpness proof in Kédagni and Mourifié (2020). Since specifying a joint distribution of (h_0, h_1) is equivalent to specifying the joint distribution of $(h_0, h_1)|K^* = 0$ as well as p (which is known), I focus on the distribution of $(h_0, h_1)|K^* = 0$.

Let us consider the lower bound Δ_k^L first. Let $Q(u)$ denote the quantile function of the data (corresponding to the CDF F). Now specify the marginal distribution of h_0 (conditional on $K^* = 0$) to follow the following quantile function:

$$Q_{0|K^*=0}^L(u) := \begin{cases} Q(u) & \text{if } 0 \leq u \leq F_{0|K^*=0}(k) \\ k + \frac{F_{0|K^*=0}(k)}{f_{0|K^*=0}(k)} \cdot \ln\left(\frac{u}{F_{0|K^*=0}(k)}\right) & \text{if } F_{0|K^*=0}(k) \leq u \leq 1 \end{cases}$$

where I've constructed the extrapolated portion for $u \geq F_{0|K^*=0}(k)$ from the lower bound on $Q^L_{0|K^*=0}$ arising from (A.2). Similarly, construct the marginal distribution of h_1 (conditional on $K^* = 0$) as:

$$Q^L_{1|K^*=0}(u) := \begin{cases} k - \frac{1-F_{1|K^*=0}(k)}{f_{1|K^*=0}(k)} \cdot \ln\left(\frac{1-u}{1-F_{1|K^*=0}(k)}\right) & \text{if } 0 \leq u \leq F_{1|K^*=0}(k) \\ Q(u) & \text{if } F_{1|K^*=0}(k) \leq u \leq 1 \end{cases}$$

using the upper bound on $Q^L_{1|K^*=0}$ arising from (A.2).

It can be readily verified that both $Q^L_{d|K^*=0}$ above are valid quantile functions defined on the unit interval $u \in [0, 1]$: they are increasing and left continuous on $[0, 1]$. Furthermore, $Q^L_{0|K^*=0}(u)$ and $Q^L_{1|K^*=0}(u)$ are locally BLC inside the bunching region by construction, and are also globally BLC provided that $F(h)$ is BLC on the regions $(0, k)$ and (k, ∞).[17]

To build a *joint* distribution of $(h_0, h_1)|K^* = 0$ from the $Q^L_{0|K^*=0}$ and $Q^L_{1|K^*=0}$ functions above, let us impose rank invariance on our constructed distribution. That is, let

$$(h_0, h_1)|K^* = 0 \sim (Q^L_{0|K^*=0}(U), Q^L_{1|K^*=0}(U)) \tag{8}$$

where U is a uniform $[0, 1]$ random variable. Then RANK holds immediately for this distribution.

Note that $Q^L_{0|K^*=0}(u)$ and $Q^L_{1|K^*=0}(u)$ recover the observed distribution $Q(u)$ of h, via Eq. (2). Lastly, we must show that $\Delta^*_k = \Delta^L_k$ when $(h_0, h_1)|K^* = 0$ follows (8). This follows from the same steps used above to prove that $\Delta^*_k \geq \Delta^L_k$ generally, with the weak inequalities replaced as equalities.

To build a distribution of (h_0, h_1) that meets the upper bound Δ^U_k, we proceed analogously. That is, let

$$Q^U_{0|K^*=0}(u) := \begin{cases} Q(u) & \text{if } 0 \leq u \leq F_{0|K^*=0}(k) \\ k - \frac{1-F_{0|K^*=0}(k)}{f_{0|K^*=0}(k)} \cdot \ln\left(\frac{1-u}{1-F_{0|K^*=0}(k)}\right) & \text{if } F_{0|K^*=0}(k) \leq u \leq 1 \end{cases}$$

and

$$Q^U_{1|K^*=0}(u) := \begin{cases} k + \frac{F_{1|K^*=0}(k)}{f_{1|K^*=0}(k)} \cdot \ln\left(\frac{u}{F_{1|K^*=0}(k)}\right) & \text{if } 0 \leq u \leq F_{1|K^*=0}(k) \\ Q(u) & \text{if } F_{1|K^*=0}(k) \leq u \leq 1 \end{cases}$$

and again impose rank invariance as before.

Note that the quantile functions $Q^B_{d|K^*=0}$ for $d \in \{0, 1\}$ and $B \in \{L, U\}$ are valid quantile functions globally across the unit interval, despite the fact that the functions of t defining the

[17] To see this, note that if $Q(u)$ is BLC on $(0, k)$ and (k, ∞), the functions $Q^L_{d|K^*=0}(u)$ are differentiable everywhere on $(0, 1)$, even at the points $F_{0|K^*=0}(k)$ and $F_{1|K^*=0}(k)$. This is because the density associated with the BLC extrapolation is itself continuous at the point of extrapolation (one can see this by differentiating the bounds in (A.1) at $t = 0$). Thus the log of the CDF $F^L_{d|K^*=0}(h)$ corresponding to each $Q^L_{d|K^*=0}$ is piecewise concave and continuous and with no kink at $h = k$, which is thus a concave function globally. The same applies to the log of $(1 - F^L_{d|K^*=0}(h))$.

upper and lower BLC bounds in (A.1) are not valid CDF functions globally in t. While those functions are continuous and increasing for all t, they will exit the unit interval if extrapolated too far in either direction. This does not affect the constructions $Q^B_{d|K^*=0}(u)$ because they are only defined within the unit interval. Intuitively, the BLC extrapolations of $Q(u)$ in quantile space only need to extend across the bunching interval $u \in [F_{0|K^*=0}(k), F_{1|K^*=0}(k)]$, which is an identified subset of the unit interval (note that the $Q^B_{d|K^*=0}$ are defined above to continue the BLC extrapolation beyond that for concreteness, and remain in $[0,1]$).

To show that we can satisfy $\Delta^*_k = \Delta$ also for any intermediate value $\Delta \in (\Delta^L_k, \Delta^U_k)$, we can use the construction $Q^\alpha_{0|K^*=0}(u) = (1-\alpha)\cdot Q^L_{0|K^*=0}(u) + \alpha \cdot Q^U_{0|K^*=0}(u)$ and $Q^\alpha_{1|K^*=0}(u) = (1-\alpha)\cdot Q^L_{1|K^*=0}(u) + \alpha \cdot Q^U_{1|K^*=0}(u)$ for $\alpha \in [0,1]$. Then, if the joint distribution of h_0 and h_1 conditional on $K^* = 0$ is equal to the unique distribution having marginals $Q^\alpha_{0|K^*=0}(u)$ and $Q^\alpha_{1|K^*=0}(u)$ and satisfying rank invariance, note that BLC remains satisfied via (A.2) and (A.3), while the buncher ATE becomes $\Delta^*_k = (1-\alpha)\cdot\Delta^L_k + \alpha\cdot\Delta^U_k$.

D.3 Proof of Theorem 2

Throughout this proof we let $h_i(\rho, k) = h_i(\rho)$, given Assumption SEPARABLE. By Lemmas B.2 and B.3 established in Appendix B, the effect of changing k on bunching is:

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

Turning now to the total effect on average hours.

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

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

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

where I have used Lemma B.2 with the Leibniz rule (establishing Item 3 in Theorem 2) as well as Lemma B.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.

Note: some additional materials are available for reference `here`

Additional References

BARLOW, R., PROSCHAN, F. and HUNTER, L. (1996). *Mathematical Theory of Reliability*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics.

BERTANHA, M., MCCALLUM, A. H. and SEEGERT, N. (2023). "Better bunching, nicer notching". *Journal of Econometrics* 237 (2, Part A), p. 105512.

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.

BLOCK, H. W., SAVITS, T. H. and SINGH, H. (1998). "The Reversed Hazard Rate Function". *Probability in the Engineering and Informational Sciences* 12 (1), 69–90.

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

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.

GUPTA, R. D. and NANDA, A. K. (2001). "Some results on reversed hazard rate". *Communications in Statistics - Theory and Methods* 30 (11), pp. 2447–2457. eprint: `https://www.tandfonline.com/doi/pdf/10.1081/STA-100107697`.

KEILSON, J. (1971). "Log-Concavity and Log-Convexity in Passage Time Densities of Diffusion and Birth-Death Processes". *Journal of Applied Probability* 8 (2), pp. 391–398.

KIJIMA, M. (1998). "Hazard Rate and Reversed Hazard Rate Monotonicities in Continuous-Time Markov Chains". *Journal of Applied Probability* 35 (3), pp. 545–556.

KÉDAGNI, D. and MOURIFIÉ, I. (2020). "Generalized instrumental inequalities: testing the instrumental variable independence assumption". *Biometrika* 107 (3), pp. 661–675. eprint: `https://academic.oup.com/biomet/article-pdf/107/3/661/33658405/asaa003.pdf`.

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

SAUMARD, A. (2019). "Bi-log-concavity: some properties and some remarks towards a multi-dimensional extension". *arXiv:1903.07347*.

SAUMARD, A. and WELLNER, J. A. (2014). "Log-concavity and strong log-concavity: A review". *Statistics Surveys* 8 (none), pp. 45 –114.

TAYLOR, H. and KARLIN, S. (1994). *An Introduction to Stochastic Modeling*. Academic Press.