Supplemental Material for “Treatment Effects in Bunching Designs: The Hours Impact of the Federal Overtime Rule”

Leonard Goff

For Online Publication
Last updated: December 22, 2021

Contents

1 An equilibrium search model of hours and wages

1.1 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. Let there be a large number N_w of workers and large number N_f of firms, and define $m = N_w/N_f$.[1] 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 (we take for granted here that firms only post job offers that are preferred to unemployment, which will indeed be 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

$$F(z, h) := P_{ZH}(u(Z_j, H_j) \leq u(z, h)) \tag{1}$$

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

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

of employed workers that are at inferior firms to (z,h). In a steady state, note that $G(z,h)$ 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 can derive the steady-state labor supply function facing each firm:

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

Eq. (2) is analogous to the baseline Burdett and Mortensen model, with the quantity $F(z,h)$ playing the role of the firm-level CDF of wages in 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

$$\pi(z,h) = N(z,h)(p(h)-z) = m\delta\lambda\cdot\frac{p(h)-z}{(\delta+\lambda(1-F(z,h))^2} \tag{3}$$

where the function $p(h)$ corresponds to $e(h)-\psi$, 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^*:

$$F(z,h) = 1+\frac{\delta}{\lambda}-\sqrt{\frac{m\delta}{\lambda}\cdot\frac{p(h)-z}{\pi^*}} \tag{4}$$

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. (4) only needs to hold at (z,h) that are actually chosen by firms in equilibrium

It follows from Eqs. (4) and (2) that we can rank firms in equilibrium by $F(z,h)$ and by size according to the quantity $z-p(h)$. Note that since Eq. (2) 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. (3).

To fully characterize the equilibrium P_{ZH}, we begin by arguing that for a strictly quasi-concave utility function u, workers cannot be indifferent between more than two points that (z,h) share a value of $z-p(h)$. 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 continuously along a line that that keeps $z-ph$ constant from a given point (z,h) in the support of P_{ZH}. Since $F(z,h)$ is constant along this line, we must have from the definition of $F(z,h)$ that either utility is constant or that P_{ZH} has no additional mass along the line. However, we cannot be moving along an indifference curve, 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.:

$$(z(t),h(t))=\text{argmax}_{z,h}u(z,h)\text{ s.t. }z-p(h)=F^{-1}(t)$$

where $F^{-1}(t)=F(z(t),h(t))$, viewed as a function of t. Suppose instead that $u(z(t),h(t))<\max_{(z,h):z-p(h)=F^{-1}(t)}u(z,h)$. 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 1 depicts the path $\{(z(t),h(t))\}_{t\in[0,1]}$ for a case in which

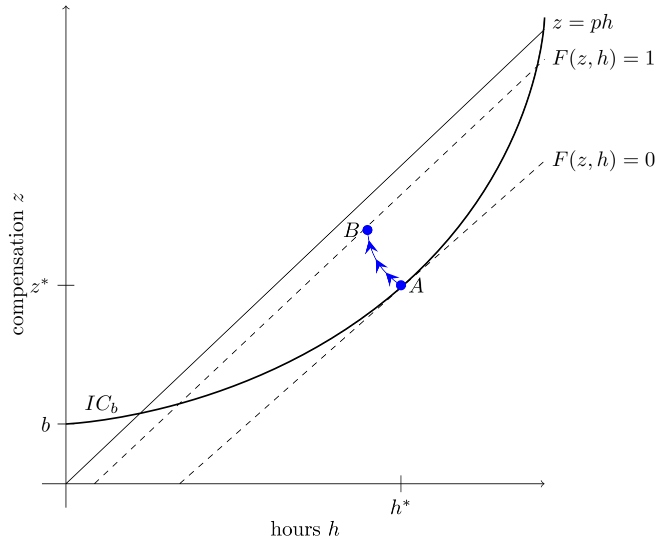

Figure 1: The support of the equilibrium distribution of compensation-hours offers (z, h) lies along the arrowed (blue) curve AB. Figure shows the case of perfect hours substitutability $p(h) = ph$. Plain curve IC_b is the indifference curve passing through the unemployment point $(b, 0)$. The least 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 beginning 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. (4) at point B. Note that this line still lies below the zero profit line $z = ph$, as firms make positive profit. Curve AB 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 north-west from A. This will be the case in the numerical calibration, in which we take preferences to follow the Stone-Geary functional form.[2] 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.

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 1. 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 to have positive employment. We 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,

[2] A CES generalization of Stone-Geary preferences would also result 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.

eventually passing $p'(h)$ at 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 $(z^*, h^*) = (z(0), h(0))$ in order to derive an expression for the $F(z,h) = 1$ line, representing the most desired firms in equilibrium. Using that $\pi^* = \pi(z^*, h^*)$, we can rewrite Equation (4) as:

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

The firms at point B in Figure 1 thus solve $z - p(h) = \left(\frac{\delta}{\delta+\lambda}\right)^2 (z^* - p(h^*))$. Equilibrium profits are

$$\pi^* = m(p(h^*) - z^*) \cdot \frac{\lambda/\delta}{(1 + \lambda/\delta)^2}$$

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 $z = p(h)$.[3]

1.2 Effects of FLSA policies

Now consider the introduction of a minimum wage, which introduces a floor on the hourly wage $w := y/h$. We 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 2 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 1, 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 2 shows how equilibrium is further affected if in addition to a binding minimum wage, premium pay is required at a higher minimum wage $1.5\underline{w}$ 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

[3] 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.

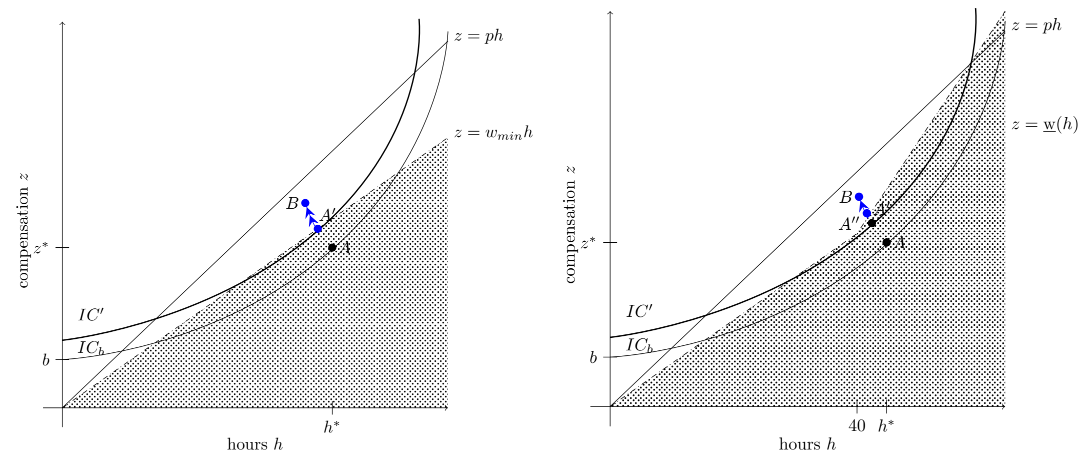

Figure 2: Left panel shows the support of the equilibrium distribution of compensation-hours offers (z, h) under a binding minimum wage. The compensation hours pairs that do meet $\underline{\mathrm{w}}$ are indicated by the shaded region. The lowest-wage offer in the economy moves from point A in the unregulated equilibrium to the point A' on the minimum wage line $z = \underline{\mathrm{w}}h$ 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 1 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 $\underline{\mathrm{w}}(h)$.

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: $\underline{\mathrm{w}}(h) := \underline{\mathrm{w}}h + 1(h > 40)(h - 40)\underline{\mathrm{w}}/2$.

1.3 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. Using this result, we calibrate the model to consider the effects of FLSA policies on earnings and hours.

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

$$\begin{pmatrix} z(t) \\ h(t) \end{pmatrix} = \begin{pmatrix} p\beta\gamma_h + (1-\beta)\gamma_z - \beta z - \beta \left(1 - \frac{t}{1+\frac{\delta}{\lambda}}\right)^2 \cdot (ph(0) - z - z(0)) \\ \beta\gamma_h + \frac{1-\beta}{p}(\gamma_z + z) + \frac{(1-\beta)}{p} \left(1 - \frac{t}{1+\frac{\delta}{\lambda}}\right)^2 \cdot (ph(0) - z - z(0)) \end{pmatrix}$$

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:*

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

We calibrate the model focusing on a lower-wage labor market where productivity is a constant $p = \$15$. We allow non-wage costs of $z = \$100$ a week, with the value based on estimates of benefit costs in the low-wage labor market.[4] We take $b = \$250$ corresponding to unemployment benefits that can be accrued at zero weekly hours of work.[5]. 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 model.

To calibrate the preference parameters, we 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 $\beta = 0.85$. We 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. We 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.[6] The value of γ_h plays a central role in setting the location of the hours distribution that we focus on. Again, the model should be interpreted as for a specific homogeneous labor market, which we take here to be full-time low wage workers in the US. We ignore taxation in the calibration.

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.

[4]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 $102.425 \approx 100$.

[5]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 $b \approx \$22,880 \cdot 0.6/52.25 = \$263$

[6] 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.

Table 1 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 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			$\underline{w} = 7.25$	$\underline{w} = 7.25$ & *OT*	$\underline{w} = 12$ & *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 1: 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 1 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.

2 Additional identification results for the bunching design

This section presents several additional sufficient conditions for point or partial identification in the bunching design, which may be applicable in various circumstances. Throughout this section, I assume that Y_0 and Y_1 admit a density everywhere so there is no counterfactual bunching at the kink. However, the results in this section can still be applied given a known $p = P(Y_{0i} = Y_{1i} = k)$ by trimming this from the observed bunching and re-normalizing the distribution $F(y)$, as described in Section 4.3.

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

2.1 Parametric approaches with constant treatment effects

To generalize the notion of constant treatment effects $\Delta_i = \Delta$, let us for any strictly increasing and differentiable transformation $G(\cdot)$ define for each unit i:

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

For example, with G equal to the logarithm function, Δ_i^G becomes proportional to a reduced form elasticity measuring the percentage change in $y_i(\mathbf{x})$ when moving from constraint B_{1i} to B_{0i}. This notion of treatment effects facilitates comparison with existing work, because familiar models predict that while Δ_i is heterogeneous δ_i^G is homogeneous when G is the natural logarithm function. For simplicity of notation, let us denote δ_i^G by δ_i when G is the natural logarithm. 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, if utility follows the iso-elastic quasi-linear form of Equation (4), we have that

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

for all units i.

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 it is strictly increasing, and thus when Δ_i^G is homogeneous for some G we have that

$$\mathcal{B} = P\left(G(Y_{0i}) \in \left[G(k), G(k) + \delta^G\right]\right)$$

by Proposition 1. We can also identify the density functions f_0^G of $G(Y_{0i})$ and f_1^G of $G(Y_{1i})$ to the left and right of $G(k)$, respectively. Given that the function $G(\cdot)$ is strictly increasing, we may also write the bunching condition as

$$\mathcal{B} = P(Y_{0i} \in [k, k + \Delta]) \text{ where } \Delta = G^{-1}\left(G(k) + \delta^G\right) - k \tag{5}$$

which defines a pseudo-parameter Δ that plays the same role as Δ would in a setup in which we assume a constant treatment effects in levels $\Delta_i = \Delta$. For example, the constant elasticity model motivates $G = \ln$ and hence $\Delta = k(e^\delta - 1)$. Note that if Δ 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, since for example this implies that $f_0^G(G(k)+\delta) = f_1^G(G(k))$ but not that $f_0(k+\Delta) = f_1(k)$. Lemma A.6 in Appendix B shows that Theorem 6 becomes exact without the uniform density assumption – making precise the idea that it is justified under a "small-kink" approximation.

Recall from Section A.2 that when Δ_i is homogeneous and $f_0(y)$ is locally uniform in the missing region $[k, k + \Delta]$, we have that

$$\Delta = \mathcal{B}/f_0(k) \tag{6}$$

and thus with constant effects in logs $\delta_i = \delta$, we can identify δ as $\ln\left(1 + \mathcal{B}/\{kf_0(k)\}\right)$. Taking the approximation $\ln(1 + x) \approx x$ and defining $\epsilon = \ln((1 - \tau_0)/(1 - \tau_1))\delta = -\ln(1 - (\tau_1 - \tau_0)/(1-\tau_0))\delta$ motivated by the iso-elastic model, we obtain $\epsilon \approx (\tau_1-\tau_0)/(1-\tau_0)\cdot\mathcal{B}/\{kf_0(k)\}$, c.f. Equations (1)-(2) in Kleven (2016).

This represents the simplest and most basic point identification result for the bunching design, and might be motivated by the idea that the kink is small, and a smooth density is locally uniform. Equation 7 generalizes naturally to a setting with heterogeneous treatment effects, as we shall see in the next section. [7] However, the uniform density assumption/approximation underlying Equation 6 may be hard to motivate in empirical settings where the kink is not small (e.g. $\tau_0 \not\approx \tau_1$), and the density away from the kink does not appear to be uniform. Thus Saez (2010) instead assumes that $f_(y)$ is linear in the missing region $[k, k+\Delta]$. We can phrase his identification result as a special case of the following:

[7] Note that the same "small-kink" approximation might be used to motivate instead the expression:

$$\delta^G = \mathcal{B}/f_0^G(k) = G'(k)\cdot\mathcal{B}/f_0(k), \tag{7}$$

which is evidently inconsistent with Equation (6) when $G'(k) \neq 1$. This illustrates the point that assuming $f_0(y)$ is constant on the region $[k, k + \Delta]$ is not the same as assuming that $f_0^G(y)$ is constant on $[G(k), G(k) + \delta^G]$ when G is non-linear.

Proposition 1 (identification by linear interpolation, à la Saez 2010). *If $\delta_i^G = \delta^G$ 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 4. □

In particular, if we assume the iso-elastic utility model Equation (4) then we have:

$$\mathcal{B} = \frac{\Delta}{2}\left\{\lim_{y\uparrow k} f(y) + \frac{k}{k+\Delta}\lim_{y\downarrow k} f(y)\right\} = \frac{k}{2}\left(\left(\frac{\rho_0}{\rho_1}\right)^{\epsilon}-1\right)\left(\lim_{y\uparrow k} f(y) + \left(\frac{\rho_0}{\rho_1}\right)^{-\epsilon}\lim_{y\downarrow k} f(y)\right) \tag{8}$$

which can be solved for ϵ by the quadratic formula, and serves as the main estimating equation from Saez (2010). Thus the empirical approach of that paper be seen as applying a result justified in a much more general model than the iso-elastic utility function assumed therein.[8]

While Proposition 1 constitutes a straightforward solution to the identification problem, the linearity assumption may like uniformity be falsified by visual inspection. For example, if we believe that $f_0(y)$ is continuously differentiable and treatment effects in levels are homogeneous (i.e. G is the identity function), then the linear interpolation used by Proposition 1 cannot hold unless $\lim_{y\downarrow k} f'(y) = \lim_{y\uparrow k} f'(y)$. Otherwise, f_0 would have to have a kink at one of the endpoints of the missing region. This limitation of Proposition 1 can be seen as a result of the fact that it only uses information about $f_0(y)$ at two points and ignores it everywhere else. 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 missing region of unknown width Δ. This technique has the added advantage of accommodating diffuse bunching, for which the relevant $\mathcal{B}$ is the "excess-mass" around k rather than a perfect point mass at k.

When bunching is exact, as in the overtime setting, the polynomial approach can be seen as a special case of the following result:

Proposition 2 (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 $\delta_i^G = \delta^G$ for some G and CONVEX and CHOICE hold. Then, provided that*

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 $y \in [k, k+\Delta]$,*

[8] 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 $\delta^G = \mathcal{B}/\frac{1}{2}\left\{\lim_{y\uparrow k} f^G(y) + \lim_{y\downarrow k} f^G(y)\right\}$.

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

$$f(y) = \begin{cases} g(y;\theta_0) & y < k \\ g(y+\Delta(\theta_0);\theta_0) & y > k \end{cases} \tag{9}$$

Proof. See Section 4. □

The standard approach of fitting a high-order polynomial to $f_0(y)$ can satisfy the assumptions of Proposition 2, since polynomial functions are analytic everywhere. Proposition 2 yields an identification result that can justify an estimation approach similar to one often made in the literature, based on Chetty et al. (2011).[9] 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 2. 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.

2.2 Nonparametric approaches with constant treatment effects

The additional assumptions from the preceding section have allowed for point-identification of causal effects under an assumption of constant 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 constant, is linear, or fits a parametric family 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. In this section, we'll see that the assumptions about $f_0(y)$ can be made non-parametric, at the expense of replacing point identification by the identification of bounds on Δ.

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. In our more general common-treatment effects framework:

Proposition 3 (partial identification from monotonicity). *Suppose that $\Delta_i = \Delta$ and $f_0(y)$ is monotonic in the interval $y \in [k, k+\Delta]$, and CONVEX and CHOICE hold. Suppose*

[9] The technique proposed by Chetty et al. (2011) in fact ignores the shift term $\Delta(\theta)$ in Equation (9), a limitation discussed by Kleven (2016). A more robust estimation procedure for parametric bunching designs could be based on iterating on Equation (9) after updating $\Delta(\theta)$, until convergence. I do not pursue this in the present paper.

that $F_1(y)$ and $F_0(y)$ are twice continuously differentiable. Then:

$$\Delta \in \left[\frac{\mathcal{B}}{max\{f_-, f_+\}}, \frac{\mathcal{B}}{min\{f_-, f_+\}}\right]$$

where the density limits $f_- := \lim_{y\uparrow k} f(y)$ and $f_+ := \lim_{y\downarrow k} f(y)$ are identified from the data.

Proof. Monotonicity of $f_0(y)$ implies that $f_0(y) \in [\min\{f_0(k), f_0(k+\Delta)\}, \max\{f_0(k), f_0(k+\Delta)\}]$ for all $y \in [k, k+\Delta]$. Homogeneous treatment effects implies that $f_0(k+\Delta) = f_1(k)$, and by continuous differentiability of F_1 and F_0 we have that $f_0(k) = f_-$ and $f_1(k) = f_+$. □

A version of Proposition 3 that allows heterogeneous treatment effects is presented in Section 2.3. 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 possible shape constraint). Log-concavity of $f_0(y)$ may be considered a natural assumption in the sense that it nests many common parametric distributions, including for example the uniform, normal, exponential extreme value and logistic, among others.[10]

Proposition 4 (bounds from log-concavity). *Suppose that $\Delta_i = \Delta$ and $f_0(y)$ is log-concave in the interval $y \in [k, k+\Delta]$ and differentiable at k and $k+\Delta$. Then, under CONVEX, CHOICE, and CONT:*

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

where

$$\Delta^U = \mathcal{B} \cdot \frac{\ln(f_+) - \ln(f_-)}{f_+ - f_-}$$

and

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

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

Proof. See Supplemental Material Section 4 and Figure 3. □

Intuition for Proposition 4 is provided in Figure 3. 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, and the corresponding integrals can be computed analytically.

[10] 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.

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 (examples of log-convexity include Pareto and certain gamma, Weibull and F distributions). Another approach is to note that locally, any well-behaved density will be either log-concave or log-convex. Thus we might justify that *one* of the two assumptions holds so long as the missing region Δ is small enough that whatever the global behavior of $f_0(y)$ is, the density does not switch between log-convexity and log-concavity within it. Under this assumption, we'll have that $\Delta \in [\min\{\Delta^U, \Delta^L\}, \max\{\Delta^U, \Delta^L\}]$, which allows the data to dictate which assumption appears more appropriate.

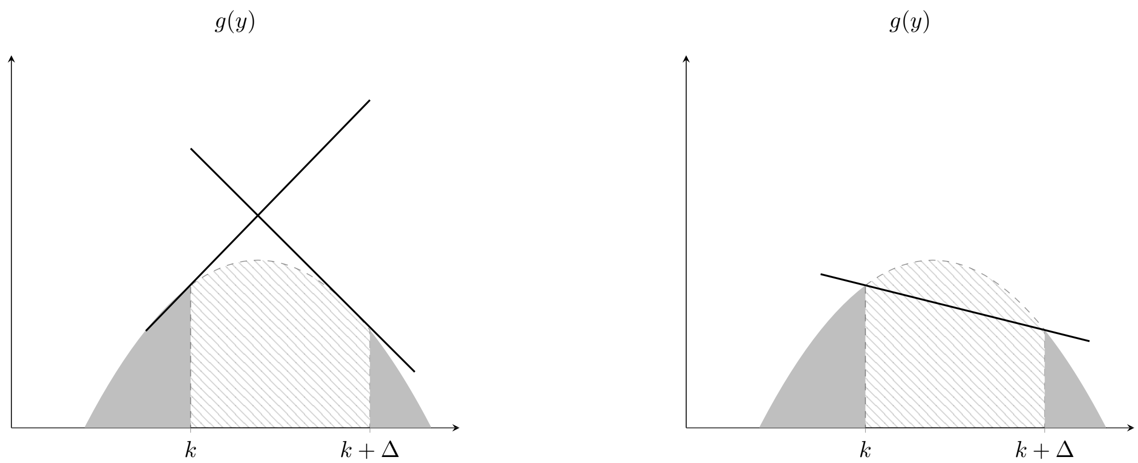

Figure 3: The left and right panels of this figure depict intuition for the lower and upper bounds on Δ in Proposition 4. See text for description.

A weaker assumption that log-concavity is bi-log-concavity, as discussed in Section 4.[11] In Proposition 5, we demonstrate that an assumption that both Y_0 and Y_1 are bi-log concave delivers bounds on Δ:[12]

Proposition 5 (bounds from bi log-concavity). *If both Y_{0i} and Y_{1i} are bi-log concave (in the regions $[k, k+\Delta]$ and $[k-\Delta, k]$ respectively), then under CHOICE and CONVEX:*

$$\Delta \in \mathcal{D}_0 \cap \mathcal{D}_1$$

where

$$\mathcal{D}_0 = \left[\frac{F_0(k)}{f_0(k)} \ln\left(1 + \frac{\mathcal{B}}{F_0(k)}\right), \ \frac{F_0(k)-1}{f_0(k)} \ln\left(1 + \frac{\mathcal{B}}{F_0(k)-1}\right)\right]$$

[11] BLC states that for a random variable with distribution $F(y)$, both $\log F(y)$ and $\log(1-F(y))$ are concave functions. Since a "typical" CDF (e.g. normal) is convex (e.g. below it's median) and then concave (e.g. above it's median), BLC essentially states that the convex part is not *too* convex that it can't be undone by the concave log function, and the concave part is not *too* concave that the convexity of $1 - F(y)$ in the corresponding region can't also be undone by the logarithm. Unlike log-concavity, which implies that a random variable's density is unimodal, bi log-concavity is compatible with the distribution having any number of modes.

[12] A result from Saumard (2019) demonstrates that if treatment effects are independent of one potential outcome and log-concavely distributed, then bi log-concavity of one potential outcome implies bi-log concavity of the other.

and

$$\mathcal{D}_1 = \left[\frac{F_1(k)}{f_1(k)}\ln\left(1+\frac{\mathcal{B}}{F_1(k)}\right),\ \frac{F_1(k)-1}{f_1(k)}\ln\left(1+\frac{\mathcal{B}}{F_1(k)-1}\right)\right]$$

Proof. Theorem 1 of Dümbgen et al. (2017) implies that

$$(1-F_0(k))\left(1-e^{-\frac{f_0(k)}{1-F_0(k)}\Delta}\right) \leq \mathcal{B} \leq F_0(k)\left(e^{\frac{f_0(k)}{F_0(k)}\Delta}-1\right)$$

Since each bound is strictly increasing in Δ, they can each be inverted to obtain bounds on the treatment effect Δ. This leads to $\mathcal{D}_0$, and a similar argument applies for $\mathcal{D}_1$. □

I conclude this section by observing that the results presented here that assume constant treatment effects often remain interpretable when this assumption fails. In particular, define Δ_0^* and Δ_1^* to be the quantile treatment effects evaluated at the points $F_1(k)$ and $F_0(k)$, respectively: $\Delta_0^* := Q_0(F_1(k))-Q_1(F_1(k)) = Q_0(F_1(k))-k$ and $\Delta_1^* := Q_0(F_0(k))-Q_1(F_0(k)) = k - Q_1(F_0(k))$. When treatment effects are a constant Δ and both Y_{0i} and Y_{1i} have positive density at k, we have that $\Delta_0^* = \Delta_1^* = \Delta$. However, even when treatment effects are not constant these two parameters satisfy "constant Δ" analogues of the bunching condition $\mathcal{B} = P(Y_{0i} \in [k, k+\Delta_i])$; in particular: $P(Y_{0i} \in [k, k+\Delta_0^*]) = P(Y_{1i} \in [k-\Delta_1^*, k]) = \mathcal{B}$. Thus, Δ_0^* and Δ_1^* represent pseudo-parameters defined by the relation $P(Y_{0i} \in [k, k+\Delta_i])$, when the econometrician assumes that $\Delta_i = \Delta$.[13] For example, in Proposition 5, $\mathcal{D}_0$ remains an identified set for Δ_0^* and $\mathcal{D}_1$ remains an identified set for Δ_1^* even when the assumption of $\Delta_i = \Delta$ fails. This echoes the robustness of the results that assume rank invariance, as described in Figure A.2.

2.3 Alternative identification strategies with heterogeneous treatment effects

A standard argument in the literature (e.g. Saez 2010 and Kleven and Waseem (2013)) is to allow heterogeneous treatment effects under a uniform density approximation. If a kink is very small, then this might be justified as an approximation by saying that Δ_i must be small for all individuals, then invoking smoothness assumptions on $f(\Delta, y)$ (see the corollary to Proposition 6 below). My results in Section 4 move beyond the need to approximate the kink as small, however I show here how an analog of this result can be stated in my generalized bunching design framework. 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,* $P(\Delta_i \geq 0 | Y_{0i} = k) = P(\Delta_i \geq 0 | Y_{1i} = k) = 1$.

[13] In the context of the general problem of comparing two populations, Doksum (1974) shows that Δ_0^* is also the unique functional of F_0 and F_1 that maps a value y of Y_0 to a value $y + \Delta(y)$ of Y_1 while respecting the following two properties: i) the mapping commutes with increasing transformations of $Y_1 = Y_0 + \Delta$; and ii) if $F_1(y) = F_0(y)$ globally, then $\Delta(y) = 0$ for all y (a similar argument applies for Δ_1^*).

Proof. Suppose $Y_{0i} = k$ and $\Delta_i < 0$, so that $Y_{1i} > k$. The proof of Proposition 1 shows that if $Y_{0i} \leq k$ then $Y_i = Y_{0i}$, so we must have that $Y_i = k$. However it also shows that $Y_{1i} \geq k$ implies that $Y_i = Y_{1i}$, so $Y_i > k$, a contradiction. An analogous argument holds when $Y_{1i} = k$. □

Proposition 6 (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 value of Δ with support: 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

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

using Lemma POS in the last step. The inequalities are equalities under CONVEX. □

Analogous assumptions on the joint distribution of Δ_i and Y_{1i} would justify replacing $\lim_{y \uparrow k} f(y)$ with $\lim_{y \uparrow k} f(y)$ in Proposition 6. Lemma SMALL in Appendix B formalizes the idea that the uniform density approximation from Proposition 6 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 7 (monotonicity with heterogeneous treatment effects). *Assume CONVEX and CHOICE, and suppose the joint density $f_0(\Delta, y)$ of Δ_i and Y_{0i} exists and 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 the joint density $f_1(\Delta, y)$ of Δ_i and Y_{1i} exists and is weakly decreasing on the interval $y \in [k, k + \Delta]$ 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. In the first case, for example:

$$\mathcal{B} \;=\; \int_0^\infty d\Delta \int_k^{k+\Delta} dy \cdot f_0(\Delta, y) \leq \int_0^\infty \Delta f_0(\Delta, k) d\Delta = f_0(k)\tau_0$$

following the proof of Proposition 6 and using Lemma POS. Note that for each Δ, $f_1(\Delta, y) = f_0(\Delta, y+\Delta)$, which can be used to derive the bound for τ_1. The reverse case is analogous $\square$

This result implies that when treatment effects are statistically independent of Y_0: $\Delta_i \perp Y_{0i}$, the bounds $\left[\frac{\mathcal{B}}{\max\{f_-, f_+\}}, \frac{\mathcal{B}}{\min\{f_-, f_+\}}\right]$ from Proposition 3 that assume constant treatment effects are also valid for the average treatment effect $\mathbb{E}[\Delta_i] = \tau_0 = \tau_1$ when treatment effects are heterogeneous.

Other approaches to identification with heterogeneous treatment effects are possible when the researcher observes covariates X_i that are unaffected by a counter-factual 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 $y = k$ (available upon request). 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).[14]. 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 a very strong assumption.

2.4 Additional bunching design examples from the literature

Below I discuss two additional examples that fit into the general kink bunching design framework described in Section 4. The first is the classic labor supply example, where quasi-concavity of $u_i(c, y)$ can arise from increasing opportunity costs of time allocated to labor as in an iso-elastic model. In the second example, firms are again the decision makers but now the "running variable" y is a function of two underlying choice variables z.

[14]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.

Example 1: Labor supply with taxation

Individuals have preferences $\tilde{u}_i(c, y) = u(c, y, \epsilon_i)$ 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. $u(c, y, \epsilon)$ is taken to be strictly quasi-concave in (c, y) as the opportunity cost of leisure rises with greater earnings, and monotonically increasing in consumption. Now let $u_i(t) = \tilde{u}_i(y - t, y)$ 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 $B(y) = \tau_0 y + 1(y \geq k)(\tau_1 - \tau_0)(y - k)$, with the kink arising from an increase in marginal tax rates from τ_0 to $\tau_1 > \tau_0$ at $y = k$. Note that in this example c represents a good, rather than a bad, from the perspective of the decision-maker.

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 where the tax regime depends on reported profits as a fraction of output: $y = \frac{r - \hat{w}}{r} = 1 - \frac{\hat{w}}{r}$:

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

which has a kink in both r and $\hat{w}$ when $y(r, \hat{w}) = k = \tau_r/\tau_\pi$. In this case, $B_0(r, \hat{w}) = \tau_r r$, 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 $(r, \hat{w})$. In this setting, the functions B_{0i}, B_{1i} and y_i are all common across firms.

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

$$u_i(T, r, \hat{w}) = r - w_i(r) - g_i(\hat{w} - w_i(r)) - T \tag{10}$$

where $g_i(\cdot)$ represents costs of tax evasion by misreporting costs. This specification of $u_i(T, r, \hat{w})$ 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: 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 indices implicit:

$$\begin{aligned}
\Delta &\approx \left.\frac{dy(r,\hat{w})}{\hat{w}}\right|_{(r_0,\hat{w}_0)} \Delta_{\hat{w}i} + \left.\frac{dy(r,\hat{w})}{r}\right|_{(r_0,\hat{w}_0)} \Delta_r \\
&= \frac{\hat{w}_0}{r_0^2}\Delta_r - \frac{1}{r_0}\left(\Delta_{w(r)} + \Delta_{(\hat{w}-w(r))}\right) \\
&\approx \frac{\hat{w}_0}{r_0^2}\Delta_r - \frac{1}{r_0}\left(w'(r_0)\Delta_{ri} + \Delta_{(\hat{w}-w(r))}\right) \\
&= \frac{1}{r_0}\left\{(1 - Y_0 - w'(r_0))\Delta_r \Delta_{\hat{w}}\right\} \approx \frac{1}{r}\left\{-k\Delta_r - \Delta_{(\hat{w}-w)}\right\} \\
&\approx \frac{1}{r_0}\left\{-\frac{\tau_r}{\tau_\pi} \cdot r\epsilon^r \frac{d(1-\tau_E)}{\tau_E} - \Delta_{\hat{w}i}\right\} = \frac{\tau_r^2}{\tau_\pi}\epsilon^r - \frac{\Delta_{(\hat{w}-w)}}{r_0}
\end{aligned} \tag{11}$$

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 (11) 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.

	p=0		p from PTO	
	Bunching	Elasticity	Net Bunching	Elasticity
Accommodation and Food Services	0.036	[-0.059, -0.060]	0.036	[-0.059, -0.060]
(N=69427)	[0.029, 0.044]	[-0.073, -0.073]	[0.029, 0.044]	[-0.073, -0.073]
Administrative and Support	0.062	[-0.102, -0.106]	0.009	[-0.014, -0.017]
(N=49829)	[0.051, 0.074]	[-0.125, -0.125]	[0.005, 0.013]	[-0.020, -0.020]
Construction	0.139	[-0.190, -0.180]	0.029	[-0.034, -0.043]
(N=136815)	[0.128, 0.149]	[-0.218, -0.218]	[0.022, 0.035]	[-0.043, -0.043]
Health Care and Social Assistance	0.051	[-0.085, -0.095]	0.005	[-0.008, -0.010]
(N=13951)	[0.034, 0.069]	[-0.135, -0.135]	[0.000, 0.010]	[-0.018, -0.018]
Manufacturing	0.137	[-0.158, -0.127]	0.018	[-0.018, -0.020]
(N=112555)	[0.126, 0.148]	[-0.177, -0.177]	[0.016, 0.021]	[-0.022, -0.022]
Other Services	0.160	[-0.120, -0.123]	0.037	[-0.024, -0.033]
(N=19263)	[0.132, 0.188]	[-0.167, -0.167]	[0.024, 0.049]	[-0.034, -0.034]
Professional, Scientific, Technical	0.136	[-0.140, -0.160]	0.010	[-0.009, -0.013]
(N=47705)	[0.117, 0.155]	[-0.175, -0.175]	[0.003, 0.016]	[-0.014, -0.014]
Real Estate and Rental and Leasing	0.187	[-0.250, -0.230]	0.097	[-0.115, -0.133]
(N=13498)	[0.141, 0.234]	[-0.355, -0.355]	[0.060, 0.135]	[-0.177, -0.177]
Retail Trade	0.129	[-0.256, -0.238]	0.032	[-0.055, -0.066]
(N=56403)	[0.112, 0.146]	[-0.359, -0.359]	[0.024, 0.040]	[-0.084, -0.084]
Transportation and Warehousing	0.091	[-0.124, -0.161]	0.015	[-0.019, -0.031]
(N=25926)	[0.070, 0.111]	[-0.167, -0.167]	[0.009, 0.022]	[-0.029, -0.029]
Wholesale Trade	0.126	[-0.212, -0.163]	0.046	[-0.067, -0.068]
(N=66678)	[0.110, 0.141]	[-0.248, -0.248]	[0.037, 0.055]	[-0.088, -0.088]
All Industries	0.116	[-0.179, -0.168]	0.027	[-0.037, -0.043]
(N=630217)	[0.112, 0.121]	[-0.190, -0.190]	[0.024, 0.029]	[-0.041, -0.041]

Table 2: Estimates of ϵ in the iso-elastic model based on assuming $h_{0it} = \eta_{it}^{-\epsilon}$ is bi-log-concave, by industry. 95% bootstrap confidence intervals in gray brackets, clustered by firm.

3 Additional figures and tables

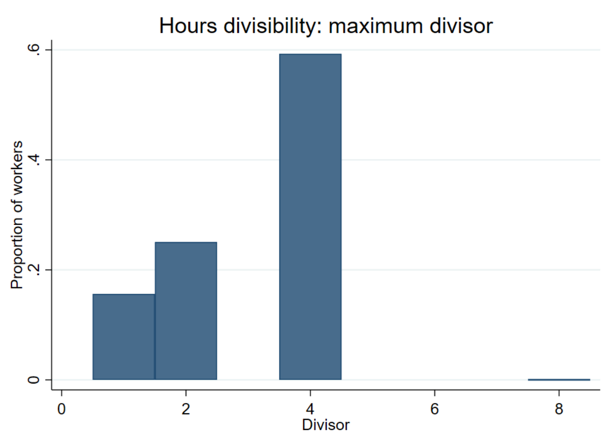

Figure 4: Distribution of the largest integer $m = 1 \ldots 10$ that maximizes the proportion of worker i's paychecks for which hours are divisible by m. This can be thought of as the granularity of hours reporting for worker i.

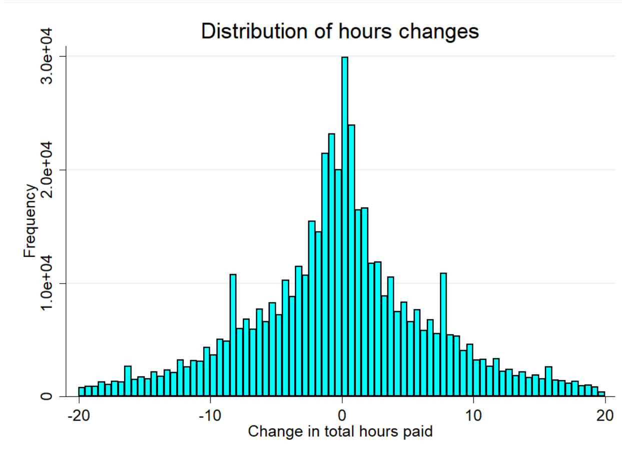

Figure 5: Distribution of changes in total hours between subsequent pay periods (truncated at -20 and 20).

4 Additional proofs

4.1 Proof of Lemma 3

This mostly follows the proof in Kasy (2017) adapted to our setting in which y is one-dimensional. As in the proof of Lemma 2 I leave k implicit in the functions $Y_i(\rho, k)$ and $Y(\rho, k, \epsilon)$, as k remains fixed throughout. One additional subtlety concerns the possibility of a point mass in the distribution of each $Y_i(\rho)$ at k^*. Note that Assumption SMOOTH implies a continuous density $f_\rho(y)$ for all $\rho \in [\rho_0, \rho_1]$ and $y \neq k^*$, which is also continuously differentiable in ρ. We define $f_\rho(k^*) = \lim_{y \to k} f_\rho(y)$ in the case that $p > 0$.

Consider any bounded differentiable function $a(y)$ having the property that $a(k^*) = 0$, and note that we may write $A(y) := \frac{d}{d\rho}\mathbb{E}[a(Y_i(\rho))]$ in two separate ways. Firstly:

$$A(y) = \frac{d}{d\rho} \int dy \cdot f_\rho(y) \cdot a(y) = \int dy \cdot a(y) \cdot \frac{d}{d\rho} f_\rho(y) \tag{12}$$

and secondly:

$$A(y) = \frac{d}{d\rho}\mathbb{E}[a(Y_i(\rho, \epsilon_i))] = \int dF_\epsilon(\epsilon) \frac{d}{d\rho} a(Y(\rho, \epsilon)) = \int dF_\epsilon(\epsilon) a'(Y(\rho, \epsilon)) \cdot \partial_\rho Y(\rho, \epsilon) \tag{13}$$

The first representation integrates over the distribution of $Y_i(\rho)$, while the second integrates over the distribution of the underlying heterogeneity ϵ_i. In both cases we are justified in swapping the integral and derivative by boundedness of $a(y)$.

Continuing with Eq. (13), we may apply the law of iterated expectations over values of

$Y(\rho, \epsilon)$, and then integrate by parts:

$$\begin{aligned} A(y) &= \int dy f_\rho(y) a'(y) \int dF_{\epsilon|Y(\rho,\epsilon)=y} \partial_\rho Y(\rho, \epsilon) \\ &= \int dy f_\rho(y) a'(y) \cdot \mathbb{E}\left[\frac{\partial Y(\rho,\epsilon)}{\partial \rho} \middle| Y(\rho,\epsilon) = y \right] \\ &= -\int dy \cdot a(y) \cdot \frac{\partial}{\partial y} \left\{ f_\rho(y) \cdot \mathbb{E}\left[\frac{\partial Y(\rho,\epsilon)}{\partial \rho} \middle| Y(\rho,\epsilon) = y \right] \right\} \end{aligned}$$

where we've assumed the density $f_\rho(y)$ vanishes at the limits of y. Comparing with Eq. (12), we see that for this to be true of any bounded differentiable function a (satisfying $a(k^*) = 0$, we must have

$$\frac{d}{d\rho} f_\rho(y) = -\frac{\partial}{\partial y} \left\{ f_\rho(y) \cdot \mathbb{E}\left[\frac{\partial Y(\rho,\epsilon)}{\partial \rho} \middle| Y(\rho,\epsilon) = y \right] \right\}$$

point-wise for all $y \neq k^*$.

Now consider $y = k^*$. First note that

$$\frac{d}{d\rho} f_\rho(k^*) = \frac{d}{d\rho} \lim_{y \to k^*} f_\rho(y) = \lim_{y \to k^*} \frac{d}{d\rho} f_\rho(y) = -\lim_{y \to k^*} \frac{\partial}{\partial y} \left\{ f_\rho(y) \mathbb{E}\left[\frac{\partial Y(\rho,\epsilon)}{\partial \rho} \middle| Y(\rho,\epsilon) = y \right] \right\}$$

where we can interchange the limit and derivative by the Moore-Osgood theorem, since $\frac{d}{d\rho} f_\rho(y)$ is uniformly bounded over $\rho \in [\rho_1, \rho_0]$ by Assumption SMOOTH. Furthermore, for all $y \neq k^*$: $\mathbb{E}\left[\frac{\partial Y(\rho,\epsilon)}{\partial \rho} \middle| Y(\rho,\epsilon) = y \right] = \mathbb{E}\left[\frac{\partial Y(\rho,\epsilon)}{\partial \rho} \middle| Y(\rho,\epsilon) = y, K_i^* = 0 \right]$, and the latter of these is continuously differentiable at all y (including $y = k^*$) by item 3 of Assumption SMOOTH. Thus:

$$\frac{d}{d\rho} f_\rho(k^*) = -\frac{\partial}{\partial y} \left\{ f_\rho(k^*) \cdot \mathbb{E}\left[\frac{\partial Y(\rho,\epsilon)}{\partial \rho} \middle| Y(\rho,\epsilon) = k^*, K_i^* = 0 \right] \right\}$$

since $f_\rho(y)$ is also continuously differentiable at $y = k^*$, by SMOOTH and the definition of $f_\rho(k^*)$ as $\lim_{y \to k^*} f_\rho(y)$.

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

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

where the third equality applies the identity $1 = P(Y_{0i} \leq k) + P(Y_i(\rho) \leq k \leq Y_i(\rho+\delta)) + P(Y_{1i} > k)$ 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.

4.3 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 $\Delta_i(\rho, \rho', \epsilon) := Y(\rho, \epsilon) - Y(\rho', \epsilon)$.

$$
\begin{aligned}
\lim_{\rho' \downarrow \rho} \frac{P(Y_i(\rho) \leq k \leq Y_i(\rho')) - p(k)}{\rho' - \rho} &= \lim_{\rho' \downarrow \rho} \frac{P(Y_i(\rho) \in [k, k + \Delta(\rho, \rho')_i]) - p(k)}{\rho' - \rho} \\
&= \lim_{\rho' \downarrow \rho} \frac{P(Y_i(\rho) \in (k, k + \Delta(\rho, \rho')_i])}{\rho' - \rho} \\
&= \lim_{\rho' \downarrow \rho} \frac{1}{\rho' - \rho} \int_0^\infty d\Delta \int_k^{k+\Delta} dy \cdot f_{\Delta(\rho,\rho'),Y(\rho)}(\Delta, y) \\
&= \lim_{\rho' \downarrow \rho} \int_0^\infty d\Delta \int_k^{k+\Delta} dy \cdot \frac{f_{\Delta(\rho,\rho'),Y(\rho)}(\Delta, k) + (y-k) r_{\Delta(\rho,\rho'),Y(\rho)}(\Delta, k, y)}{\rho' - \rho}
\end{aligned} \tag{14}
$$

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:

$$\begin{aligned}\text{RHS of above} &\le \lim_{\rho'\downarrow\rho}\int_0^{\xi(\rho,\rho')} d\Delta\frac{\xi(\rho,\rho')}{\rho'-\rho}\int_k^{k+\xi(\rho,\rho')} dy\cdot\left|r_{\Delta_i(\rho,\rho'),Y_i(\rho)}(\Delta,y)\right| \\ &= \lim_{\rho'\downarrow\rho}\frac{\xi(\rho,\rho')}{\rho'-\rho}\cdot\lim_{\rho'\downarrow\rho}\int_0^{\xi(\rho,\rho')} d\Delta\int_0^{\xi(\rho,\rho')} dy\cdot\left|r_{\Delta_i(\rho,\rho'),Y_i(\rho)}(\Delta,k+y)\right| \quad (15)\end{aligned}$$

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 $\rho\in[\rho_0,\rho_1]$:

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

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 $||\epsilon-\epsilon'||<\delta$:

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

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 $n\in\mathbb{N},\epsilon\in E$.

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

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. (15) exists and is finite, completing the proof that it evaluates to zero.

Given that the second term in Eq. (14) is zero, we can simplify the remaining term as:

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

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 $\mathbb{E}\left[M|\, Y_i(\rho) = k\right] = M < \infty$.

4.4 Proof of Appendix 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 $h = \left[\frac{\phi}{e_0}\cdot\frac{\beta}{\beta-\eta}\right]^{1/\eta}$.

4.5 Proof of Appendix C Proposition 3

Note: this proof follows the notation of Y_i from Appendix A, rather than h_{1it} from Appendix C 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:

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

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:

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

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

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

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:

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

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}.

4.6 Proof of Appendix C Theorem 2*

Note: this proof follows the notation of Y_i from Appendix A, rather than h_{1it} from Appendix C 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

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

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:

$$\begin{aligned}
&-\frac{1-F_{0|T=1}(k)}{f_{0|T=1}(k)(\mathcal{B}^*)}\int_{F_{0|T=1}(k)}^{F_{0|T=1}(k)+\mathcal{B}^*} \ln\left(\frac{1-u}{1-F_{0|T=1}(k)}\right)du \\
&\qquad\qquad - \frac{F_{1|T=1}(k)}{f_{1|T=1}(k)(\mathcal{B}^*)}\int_{F_{1|T=1}(k)-(\mathcal{B}^*)}^{F_{1|T=1}(k)} \ln\left(\frac{v}{F_{1|T=1}(k)}\right)dv \\
&= g'(F_{0|T=1}(k), f_{0|T=1}(k), \mathcal{B}^*) + h'(F_{1|T=1}(k), f_{1|T=1}(k), \mathcal{B}^*)
\end{aligned}$$

where

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

and

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

We have then that $E[Y_{0i} - Y_{1i}|Y_i = k, T_i = 0] \in [\Delta_k^L, \Delta_k^U]$, where:

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

and

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

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:

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

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

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

As shown by Dumgen 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. Let

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

Then

$$P(Y_i < k) = P(Y_{0i} \le k \text{ and } T_i = 1) + A$$

and

$$P(Y_i > k) = P(Y_{1i} > k \text{ and } T_i = 1) + B$$

Meanwhile:

$$\begin{aligned} P(Y_i \le k) - p = P(Y_i \le k \text{ and } K_i^* = 0) &= P(Y_i \le k \text{ and } T_i = 1) + A \\ &= P(Y_{1i} \le k \text{ and } T_i = 1) + A \end{aligned}$$

and

$$\begin{aligned} P(Y_i \ge k) - p = P(Y_i \ge k \text{ and } K_i^* = 0) &= P(Y_i \ge k \text{ and } T_i = 1) + B \\ &= P(Y_{0i} > k \text{ and } T_i = 1) + B \end{aligned}$$

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.

4.7 Proof of Appendix E 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 $h = \left[\frac{\phi}{e_0} \cdot \frac{\beta}{\beta-\eta}\right]^{1/\eta}$.

4.8 Proof of Supplemental Material Proposition 1

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 $\mathcal{B} = \frac{\Delta}{2}\left(f_0(k) + f_0(k+\Delta)\right)$. Although $f_0(k) = \lim_{y\uparrow k} f(y)$ by CONT, $f_0(k+\Delta)$ is not immediately observable. However:

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

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.

4.9 Proof of Supplemental Material Proposition 2

We seek a Δ such that for some θ_0:

$$\mathcal{B} = \int_{\tilde{k}}^{k+\Delta} g(y;\theta_0)dy \tag{16}$$

and

$$f(y) = \begin{cases} g(y;\theta_0) & y < k \\ g(y+\Delta;\theta_0) & y > k \end{cases} \tag{17}$$

and

$$g(y;\theta_0) > 0 \text{ for all } y \in [k, k+\Delta] \tag{18}$$

Recall from Equation (5) that $\Delta = G^{-1}\left(G(k)+\delta\right) - k$ 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 (16), 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. (16) for a given θ_0, which is unique for each permissible value θ_0 by the positivity condition (18). To satisfy (17), we would also need

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

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 (19), 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 $f_0(y)$.

4.10 Proof of Supplemental Material Proposition 4

We first prove the lower bound. Let $g(y) := \ln f_0(y)$. Concavity of $g(y)$ means that for any $\theta \in [0,1]$:

$$g\left((1-\theta)k + \theta(k+\Delta)\right) \geq (1-\theta)g(k) + \theta g(k+\Delta)$$

Then:

$$\begin{aligned}
\mathcal{B} &= \int_k^{k+\Delta} e^{g(x)} dx = \Delta \int_0^1 e^{g((1-\theta)k+\theta(k+\Delta))} d\theta \\
&\geq \Delta \int_0^1 e^{(1-\theta)g(k)+\theta g(k+\Delta)} d\theta \\
&= \Delta e^{g(k)} \int_0^1 e^{\theta(g(k+\Delta)-g(k))} d\theta \\
&= \Delta f_0(k) \cdot \left. \frac{e^{\theta(g(k+\Delta)-g(k))}}{g(k+\Delta)-g(k)} \right|_0^1 = \Delta f_0(k) \cdot \frac{\frac{f_1(k)}{f_0(k)} - 1}{g(k+\Delta)-g(k)} \\
&= \Delta \frac{f_1(k) - f_0(k)}{\ln(f_1(k)) - \ln(f_0(k))} := \Delta^U
\end{aligned}$$

where the change of variables implies that $dx = -kd\theta + (k+\Delta)d\theta = \Delta d\theta$ and we've used that $e^{g(k+\Delta)} = f_0(k+\Delta) = f_1(k)$.

We now turn to the lower bound. Log concavity of $f_0(k)$ means that $g(y)$ is concave, and thus $g(k+x) \leq g(k) + g'(k)x$ for all $x \in [0, \Delta]$. Then:

$$\begin{aligned}
\mathcal{B} &= \int_0^\Delta e^{g(k+x)} dx \leq e^{g(k)} \int_0^\Delta e^{g'(k)x} dx \\
&= \frac{f_0(k)}{g'(k)} \left(e^{g'(k)x} \Big|_0^\Delta \right) \\
&= \frac{f_0(k)^2}{f_0'(k)} \left(e^{\frac{f_0'(k)}{f_0(k)}\Delta} - 1 \right)
\end{aligned}$$

where we've used that $g'(k) = \frac{f_0'(k)}{f_0(k)}$. Inverting this expression for Δ leads to $\Delta \geq \Delta_0^L$, where

$$\Delta_0^L = \frac{f_0(k)}{f_0'(k)} \ln \left(1 + \frac{\mathcal{B}}{f_0(k)^2} f_0'(k) \right)$$

(the inequality has the same direction regardless of the sign of $f_0'(k)$). Replacing f_0 with f_1 and extrapolating from the right would a second lower bound $\Delta \leq \Delta_1^L$, where

$$\Delta_1^L = \frac{-f_1(k)}{f_1'(k)} \ln \left(1 - \frac{\mathcal{B}}{f_1(k)^2} f_1'(k) \right),$$

based on the inequality $g(k+x) \leq g(k+\Delta) - g'(k+\Delta)(\Delta - x)$ for all $x \in [0, \Delta]$.

However, neither Δ_0^L or Δ_L^1 is a sharp lower bound for Δ, because assuming log-concavity holds the bounds cross within the missing region at the value $k + X$ such that $g_0 + g_0'X + g_1'(\Delta - X) = g_1$, or:

$$X = \Delta \frac{g_1'}{g_1' - g_0'} - \frac{g_1 - g_0}{g_1' - g_0'},$$

where $g_0 = \ln f_0(k) = \ln f_-$, $g_1 = \ln f_0(k+\Delta) = \ln f_+$, $g_0' = \ln f_0'(k) = \ln f_-'$ and $g_1' = \ln f_0'(k+\Delta) = \ln f_+'$.

So:

$$
\begin{aligned}
\mathcal{B} &= \int_0^{\Delta} e^{g(k+x)}dx = \int_0^{X} e^{g(k+x)}dx + \int_X^{\Delta} e^{g(k+x)}dx \\
&\le \int_0^{X} e^{g(k)+g'(k)x}dx + \int_X^{\Delta} e^{g(k+\Delta)-g'(k+\Delta)(\Delta-x)}dx \\
&= e^{g_0}\int_0^{X} e^{g_0' x}dx + e^{g_1}\int_X^{\Delta} e^{-g_1'(\Delta-x)}dx \\
&= \frac{e^{g_0}}{g_0'}\left(e^{g_0' x}\Big|_0^X\right) - \frac{e^{g_1}}{g_1'}\left(e^{-g_1' x}\Big|_0^{\Delta-X}\right) \\
&= \left(\frac{e^{g_1}}{g_1'} - \frac{e^{g_0}}{g_0'}\right) + \frac{e^{g_0}}{g_0'}e^{g_0' X} - \frac{e^{g_1}}{g_1'}e^{g_1'(X-\Delta)} \\
&= \left(\frac{e^{g_1}}{g_1'} - \frac{e^{g_0}}{g_0'}\right) + e^{\Delta\frac{g_0' g_1'}{g_1'-g_0'}}\left(\frac{e^{g_0}}{g_0'}e^{-\frac{g_1-g_0}{g_1'-g_0'}g_0'} - \frac{e^{g_1}}{g_1'}e^{-\frac{g_1-g_0}{g_1'-g_0'}g_1'}\right) \\
&= \left(\frac{e^{g_1}}{g_1'} - \frac{e^{g_0}}{g_0'}\right) + e^{\Delta\frac{g_0' g_1'}{g_1'-g_0'}}\left(\frac{e^{g_0}}{g_0'}e^{(g_0-g_1)\frac{g_0'}{g_1'-g_0'}} - \frac{e^{g_1}}{g_1'}e^{(g_0-g_1)\left(1+\frac{g_0'}{g_1'-g_0'}\right)}\right) \\
&= \left(\frac{e^{g_1}}{g_1'} - \frac{e^{g_0}}{g_0'}\right) + e^{\frac{\Delta}{\frac{1}{g_0'}-\frac{1}{g_1'}}}e^{g_0+(g_0-g_1)\frac{g_0'}{g_1'-g_0'}}\left(\frac{1}{g_0'} - \frac{1}{g_1'}\right)
\end{aligned}
$$

where we've used that $X - \Delta = \Delta\frac{g_0'}{g_1'-g_0'} - \frac{g_1-g_0}{g_1'-g_0'}$. Solving for Δ:

$$
\begin{aligned}
\Delta &= \left(\frac{1}{g_0'} - \frac{1}{g_1'}\right)\ln\left\{\frac{\mathcal{B} + \frac{e^{g_0}}{g_0'} - \frac{e^{g_1}}{g_1'}}{e^{g_0+(g_0-g_1)\frac{g_0'}{g_1'-g_0'}}\left(\frac{1}{g_0'} - \frac{1}{g_1'}\right)}\right\} = \left(\frac{1}{g_0'} - \frac{1}{g_1'}\right)\ln\left\{\frac{\mathcal{B} + \frac{e^{g_0}}{g_0'} - \frac{e^{g_1}}{g_1'}}{e^{\frac{g_0 g_1' - g_1 g_0'}{g_1'-g_0'}}\left(\frac{1}{g_0'} - \frac{1}{g_1'}\right)}\right\} \\
&= \left(\frac{1}{g_0'} - \frac{1}{g_1'}\right)\left\{\ln\left(\mathcal{B} + \frac{e^{g_0}}{g_0'} - \frac{e^{g_1}}{g_1'}\right) + \frac{g_1 g_0' - g_0 g_1'}{g_1' - g_0'} - \ln\left(\frac{1}{g_0'} - \frac{1}{g_1'}\right)\right\} \\
&= \left(\frac{1}{g_0'} - \frac{1}{g_1'}\right)\left\{\ln\left(\mathcal{B} + \frac{e^{g_0}}{g_0'} - \frac{e^{g_1}}{g_1'}\right) + \frac{\frac{g_1}{g_1'} - \frac{g_0}{g_0'}}{\frac{1}{g_0'} - \frac{1}{g_1'}} - \ln\left(\frac{1}{g_0'} - \frac{1}{g_1'}\right)\right\} \\
&= \left(\frac{f_0}{f_0'} - \frac{f_1}{f_1'}\right)\ln\left(\frac{\mathcal{B} + \frac{f_0^2}{f_0'} - \frac{f_1^2}{f_1'}}{\frac{f_0}{f_0'} - \frac{f_1}{f_1'}}\right) + \frac{f_1}{f_1'}\ln f_1 - \frac{f_0}{f_0'}\ln f_0 \qquad (20)
\end{aligned}
$$

4.11 Details of calculations for policy estimates

4.11.1 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 $\mathbb{E}\left[\ln Y_{0i} - \ln Y_{1i}|Y_i = k, K_i^* = 0\right]$.

4.11.2 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:

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

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

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

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

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

4.11.3 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 $k' < k^*$:

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

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

4.11.4 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

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

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:

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

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.*

For any $y' > y_0$:

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

and for any $y' < y_0$:

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

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

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

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:

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

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.

4.12 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 6 below.

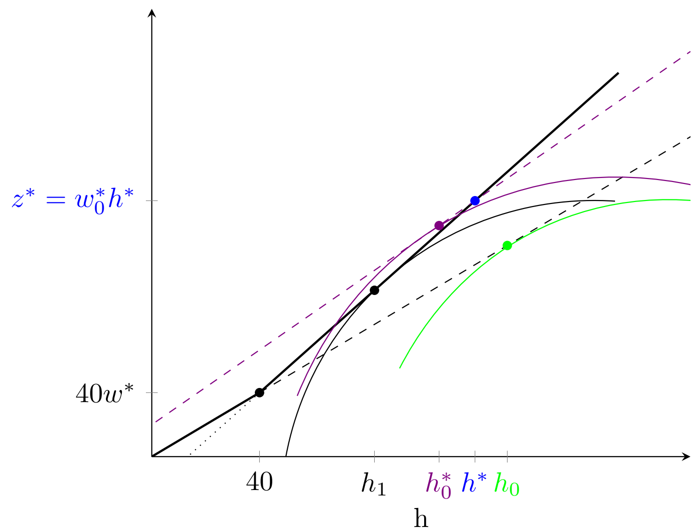

Figure 6: Depiction of h^*, h_0, h_0^* and h_1 for a single fixed worker that works overtime at h_1 hours this week. Their realized wage w^* has been set to yield earnings z^* based on anticipated hours h^* given the FLSA kink. In a world without the FLSA, the worker's wage would instead be $w_0^* = z^*/h^*$, 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

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

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

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

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

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

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

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

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.

DOKSUM, K. (1974). "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case". *Annals of Statistics* 2 (2). Publisher: Institute of Mathematical Statistics, pp. 267–277.

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. (2017). "Who wins, who loses? Identification of the welfare impact of changing wages". *Working Paper*, 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 (June), pp. 435–464.

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

MANNING, A. (2003). *Monopsony in Motion: Imperfect Competition in Labor Markets*. OCLC: 876212374. Princeton: Princeton University Press.

SAEZ, E. (2010). "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2 (3). ISBN: 1945-7731 _eprint: arXiv:1011.1669v3, pp. 180–212.

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

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