Additional Material for "Treatment Effects in Bunching Designs: The Impact of Overtime Pay on Hours"

Leonard Goff

Last updated: August 1, 2023

Contents

G Additional empirical information and results

G.1 Estimates imposing the iso-elastic model

This section estimates bounds on ϵ from the iso-elastic model described in Section 4.2, under the assumption that the distribution of $h_{0it} = \eta_{it}^{-\epsilon}$ is bi-log-concave (and linear as in Saez, 2010 as a special case). If h_{0it} is BLC, bounds on ϵ can be deduced from the fact that

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

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

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

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

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

where $F_0(k) = \lim_{h\uparrow 40} F(h)$ and $f_0(k) = \lim_{h\uparrow 40} f(h)$ are identified from the data. The bounds on ϵ estimated in this way are $\epsilon \in [-.210, -.167]$ in the full sample, with all bunching $\mathcal{B}$ attributed to the kink ($p = 0$).

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

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

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

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

where $F_1(k) = F(k)$ and $f_1(k) := \lim_{h\downarrow 40} f(h)$ are identified from the data. Empirically, these bounds are estimated as $\epsilon \in [-.179, -.141]$. Taking the intersection of these bounds with the range $\epsilon \in [-.210, -.168]$ estimated previously, we have that $\epsilon \in [-.179, -.168]$.[1] The

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

identified set is reduced from a length of .043 to .012, a factor of nearly 4. This underscores the importance of using the data from *both* sides of the kink for identification. Since a linear density satisfies BLC, the identification assumption of Saez, 2010, that the density of h_0 is linear, picks a point within the identified set under BLC. Table 6 verifies that this is born out in estimation (with results are expressed there as level effects rather than an elasticity).

As discussed in Section 4.2, a value fo $\epsilon \approx -.175$ is difficult to reconcile with a realistic view of revenue production with respect to hours. Note that if instead of the isoelastic model, production were instead described by a more general separable and homogeneous production function like

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

then treatment effects are $\Delta_{it} = g(1/\eta_{it}) - g(1.5/\eta_{it})$, where $g(m) := (f')^{-1}(m)$ yields the hours h at which $f'(h) = m$. We can then use the fundamental theorem of calculus to express this as $(h_{1it} - h_{0it})/h_{0it} = 1.5^{\bar{\epsilon}_{it}} - 1$ where $\bar{\epsilon}_{it}$ is a unit-specific weighted average of the inverse elasticity of production between $1.5\eta_{it}$ and η_{it}: $\bar{\epsilon}_{it} := \int_{\eta_{it}^{-1}}^{1.5\eta_{it}^{-1}} \lambda(m) \cdot \epsilon(g(m)) \cdot dm$, and $\lambda(m) = \frac{1/m}{\ln 1.5}$ is a positive function integrating to one. Here $\bar{\epsilon}_{it}$ plays the role of an "effective" elasticity parameter that determines the size of treatment effects when the production function is $f(h)$. This suggests that simply generalizing the functional form $f(h)$ is not sufficient to reconcile a realistic picture of production with the data, since the observed bunching still maps to a local average elasticity of $f(h)$. However, the general choice model that allows multiple margins of choice $\mathbf{x}$ can.

G.2 Additional figures and treatment effect estimates

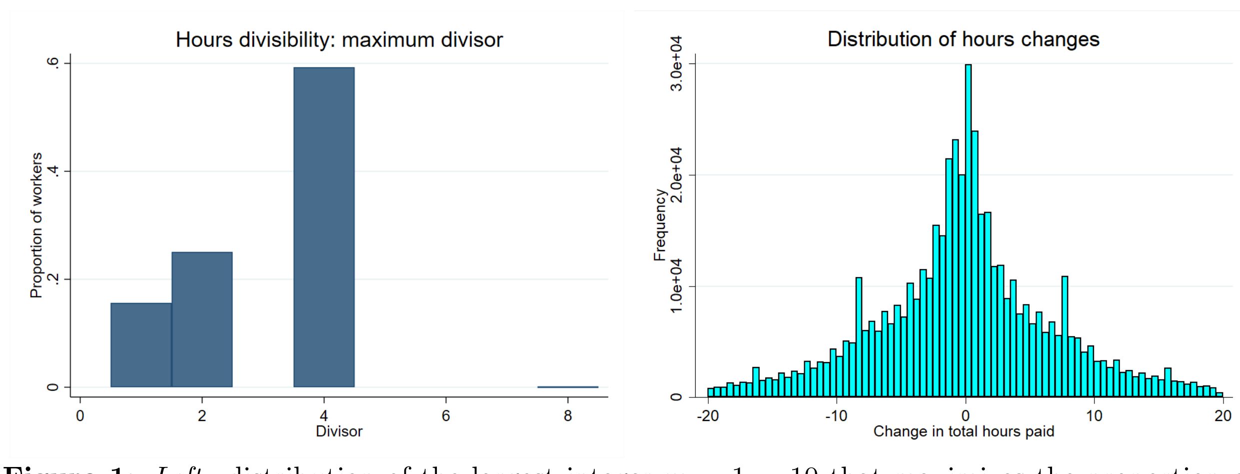

Figure 1: *Left:* distribution of the largest integer $m = 1 \ldots 10$ that maximizes the proportion of worker i's paychecks for which hours are divisible by m. This can be thought of as the granularity of hours reporting for worker i. *Right:* distribution of changes in total hours between subsequent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

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

t statistics in parentheses

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

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

t statistics in parentheses

Table 2: Decomposing variation in total hours. Worker fixed effects and employer by day fixed effects explain about 63% of the variation in total hours.

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

Table 3: Estimates of the hours effect of the FLSA by industry, based on $p = 0$ (left) or p estimated from paid time off (right). Estimates are reported only for industries having at least 10,000 observations. 95% bootstrap confidence intervals in gray, clustered by firm. In the case of Accommodation and Food Services, $P(h_{it} = 40 | \eta_{it} > 0) > \mathcal{B}$, so I take the PTO-based estimate to be $p = 0$.

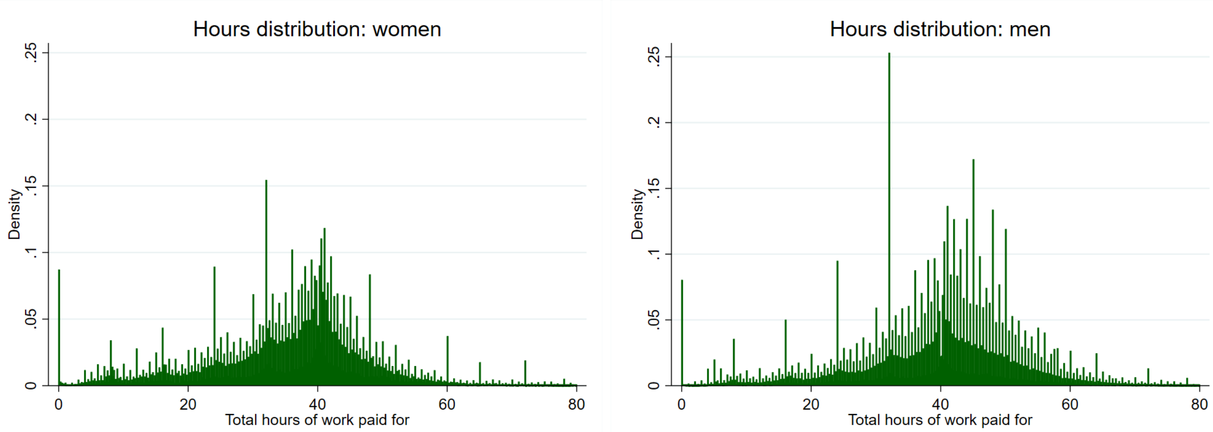

Table 4: Hours distribution by gender, conditional on different than 40 for visibility (bunching can be read from Fig 5).

	p=0	p from PTO	p=0	p from PTO
Net bunching:	0.090	0.011	0.124	0.031
	[0.083, 0.098]	[0.009, 0.012]	[0.119, 0.129]	[0.028, 0.034]
Buncher ATE	[1.507, 1.709]	[0.187, 0.190]	[3.074, 3.635]	[0.828, 0.868]
	[1.387, 1.855]	[0.150, 0.227]	[2.777, 3.991]	[0.717, 0.986]
Buncher ATE as elasticity	[0.093, 0.105]	[0.012, 0.012]	[0.190, 0.224]	[0.051, 0.053]
	[0.086, 0.114]	[0.009, 0.014]	[0.171, 0.246]	[0.044, 0.061]
Average effect of kink on hours	[-0.633, -0.489]	[-0.078, -0.054]	[-1.867, -1.271]	[-0.482, -0.311]
	[-0.688, -0.446]	[-0.094, -0.043]	[-2.060, -1.149]	[-0.549, -0.269]
Num observations	147953	147953	482264	482264
Num clusters	352	352	524	524

Table 5: Results of the bunching estimator among women (left) vs. men (right).

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

Table 6: Treatment effects in levels with comparison to alternative shape constraints. Rows "Linear density" and "Monotonic density" assume homogenous treatment effects.

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

Table 7: Estimates of policy effects (replicating Table 4) ignoring the potential effects of changes to straight wages.

H Interdependencies among hours within the firm

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

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

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

4 are $h_{0i} = \mathtt{h}_{0i}(\mathbf{h_{-i}})$ and $h_{1i} = \mathtt{h}_{1i}(\mathbf{h_{-i}})$. As in Section 4.4 let $(h_i^*, \mathbf{h^*}_{-i})$ denote the hours profile that would occur absent the FLSA, so that the average ex-post effect of the FLSA is $\mathbb{E}[h_i - h_i^*]$.

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

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

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

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

Thus both substitution and complementarity in hours can plausibly coexist within a firm, and it is difficult to sign theoretically the overall contribution of interdependencies on our parameter of interest θ (c.f. Eq. (8)). Given that neither occupations nor tasks are observed in the data, it is also difficult to obtain direct evidence even with the aid of functional-form assumptions like Eq. (1). I therefore turn to an indirect empirical test of whether these effects are likely to play a significant role in θ.

An ideal test of interdependencies between hours within a firm would leverage random individual-level shocks to a worker's hours, and look for a response in the hours of their colleagues. A worker taking sick-pay—thus reducing their hours of work—represents a compelling candidate as its timing may be uncorrelated with that of firm-level shocks (after controlling for seasonality). Figure 2 uses an event study design to show that in weeks when a worker receives a positive number of sick-pay hours, their individual hours worked for that week decline by about 8 hours on average. Yet I fail to find evidence of a corresponding change in the hours of others in the same firm. This suggests that short term variation in the hours of a worker's colleagues does not tend to translate into contemporaneous changes in their own (for example, if the firm were dividing a fixed number of hours across workers). Figure 3 produces similar results when repacing the two-wage-fixed specification of Figure 2 with an "imputation"-based approach similar to Borusyak et al. (2021) and Gardner (2021).

Table 8 shows another piece of evidence: that my overall effect estimates are similar between small, medium, and large firms. If firms were to compensate for overtime hours reductions by "giving" some hours to similar workers who would otherwise be working less than 40, for instance, then we would expect this to play a larger role in firms where there are a large number of substitutable workers–causing a bias that increases with firm size.

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

Table 8: Estimates of the ex-post effect of the kink by firm size. "Small" firms have between 1 and 25 workers, "Medium" have 26 to 50, and "Large" have more than 50. Note that the estimated net bunching caused by the FLSA is similar across firm sizes (right), despite the raw bunching observed in the data differing by firm size category.

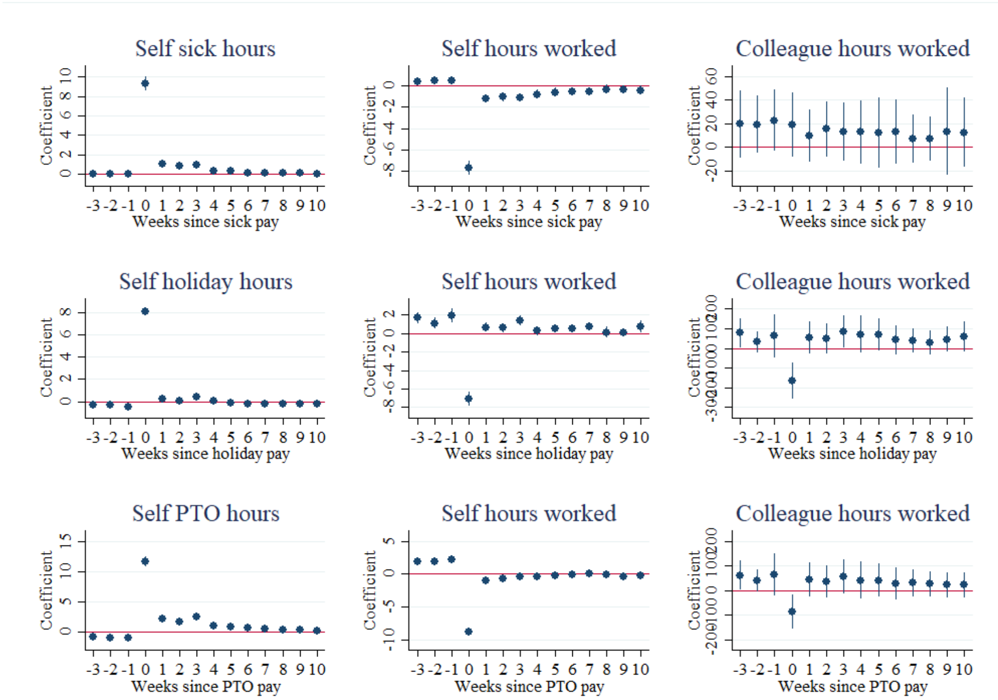

Figure 2: Event study coefficients β_j and 95% confidence intervals across an instance of a worker receiving pay for non-work hours (either sick pay, holiday pay, or paid time off–'PTO'). Confidence intervals are constructed by non-parametric bootstrap clustered by firm. Estimating equation is $y_{it} = \mu_t + \lambda_i + \sum_{j=-3}^{10} \beta_j D_{it,j} + u_{it}$, where $D_{it,j} = 1$ if worker i in week t has a positive number of a given type of non-work hours j weeks ago (after a period of at least three weeks in which they did not), λ_i are worker fixed effects, and μ_t are calendar week effects. Rows correspond to choices of the non-work pay type: either sick, holiday, PTO. Columns indicate choices of the outcome y_{it}. "Colleague hours worked" sums the hours of work in t across all workers other than i in i's firm. The timing of both holiday and PTO hours appears to be correlated across workers, leading to a decrease in the working hours of i's colleagues in weeks in which i takes either holiday or PTO pay (center-right and bottom-right graphs). However I cannot reject that colleague work hours are unrelated to an instance of sick pay: before, during and after it occurs (top-right). Meanwhile i's hours of work reduce by about 8 hours on average during an instance of sick pay (top-center). This suggests that there is no contemporaneous reallocation of i's forgone work hours to their colleagues.

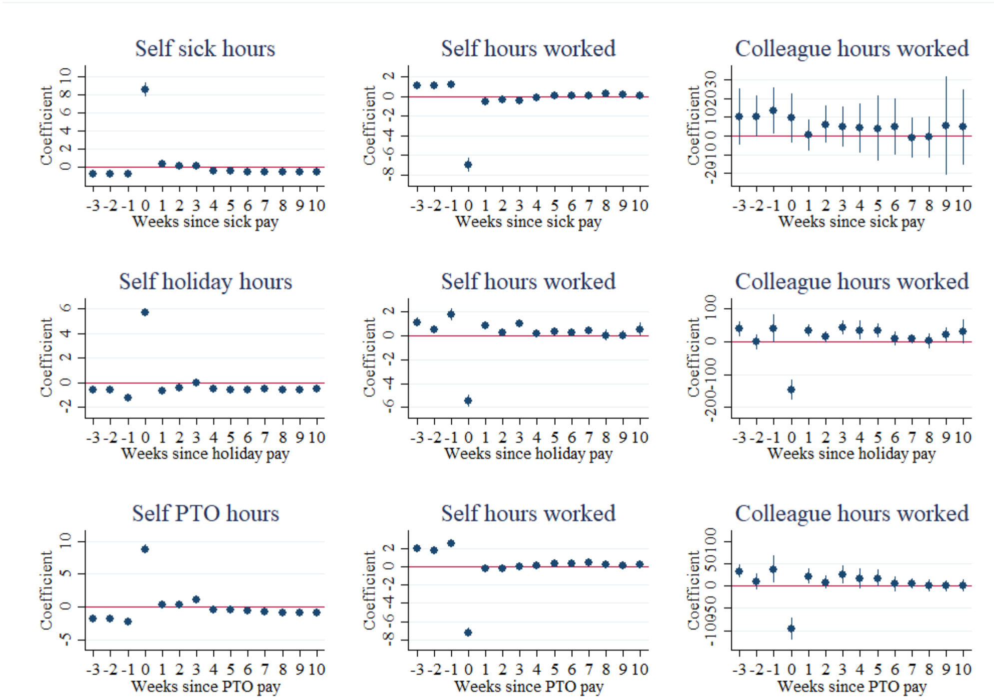

Figure 3: This figure replaces the two-way-fixed-effects estimator used in Figure 2 with an "imputation" approach similar to Borusyak et al. (2021) and Gardner (2021). Results are very similar to those in Figure 2. Specifically, I call all observations that are not between 3 weeks before and 10 weeks after a spell of non-work hours "clean controls", and estimate a first regression $y_{it} = \mu_t + \lambda_i + \epsilon_{it}$ using these observations only. This regression includes all paychecks for workers that never have the corresponding type of non-work hour (sick pay, holiday pay, or PTO), but also a subset of paychecks for nearly all workers who do have a spell of non-work hours at some point (allowing me to estimate their fixed effect λ_i). Given the $\hat{\mu}_t$ and $\hat{\lambda}_i$, I compute $\tilde{y}_{it} = y_{it} - \hat{\mu}_t - \hat{\lambda}_i$ among units that are not clean controls (i.e. those between -3 and 10 weeks after the start of a spell), and estimate a second regression $\tilde{y}_{it} = \sum_{j=-3}^{10} \beta_j D_{it,j} + e_{it}$ on these units only (dropping a small number of workers i for whom there were no clean-control observations). 95% confidence intervals are constructed by non-parametric bootstrap clustered by firm.

I Modeling the determination of wages and "typical" hours

I.1 A simple model with exogenous labor supply

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

$$\min_{z,h,K,N} N \cdot (z + \psi) + rK \text{ s.t. } F(Ne(h), K) \geq Q \text{ and } N \leq N(z, h) \tag{2}$$

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

To understand the properties of the solution to Equation (2), let us examine two illustrative special cases.

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

Much of the literature on hours determination has taken the hourly wage as a fixed input to the choice of hours, and assumed that at that wage the firm can hire any number of workers, regardless of hours. This can be motivated as a special case of Equation (2) in which there is perfect competition on the straight-time wage, i.e. $N(z,h) = \bar{N}\mathbb{1}(w_s(z,h) \geq w)$ for some large number $\bar{N}$ and wage w exogenous to the firm, where the function $w_s(\cdot)$ is defined in Equation (1). Then Equation (2) reduces to:

$$\min_{N,h,K} N \cdot (hw + \mathbb{1}(h > 40)(w/2)(h-40) + \psi) + rK \text{ s.t. } F(Ne(h), K) \geq Q \tag{3}$$

By limiting the scope of labor supply effects in the firm's decision, Equation (3) is well-suited to illustrating the competing forces that shape hours choice on the production side: namely the fixed costs ψ on the one hand and the concavity of $e(h)$ on the other. Were ψ equal to zero with $e(h)$ strictly concave globally, a firm solving Equation (3) would always find it cheaper to produce a given level of output with more workers working less hours each. On the other hand, were ψ positive and e weakly convex, it would always be cheapest to hire a single worker to work all of the firm's hours. In general, fixed costs and declining hours productivity introduce a tradeoff that leads to an interior solution for hours.[2]

Equation (3) introduces a kink into the firm's costs as a function of hours, much as short-run wage rigidity does in my dynamic analysis. However, the assumption that the firm can demand any number of hours at a set straight-time wage rate is harder to defend when thinking about firms long-run expectations, a point emphasized by Lewis (1969). Equilibrium considerations will also tend to run against the independence of hourly wages and hours - a mechanism explored in Appendix I.2.

[2] In the fixed-wage special case, these two forces along with the wage are in fact sufficient to pin down hours, which do not depend on the production function F or the chosen output level Q. See e.g. Cahuc and Zylberberg (2014) for the case in which $e(h)$ is iso-elastic.

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

By placing some functional form restrictions on Equation (2), we can obtain a closed-form expression for (z^*, h^*). In particular, when both labor supply and $e(h)$ are iso-elastic, production is separable between capital and labor and linear in the latter, and firms set the output target Q to maximize profits, Proposition I.1 characterizes the firm's choice of earnings and hours:

Proposition I.1. *When i) $e(h) = e_0 h^\eta$ and $N(z,h) = N_0 z^{\beta_z} h^{\beta_h}$; ii)$F(L,K) = L + \phi(K)$ for some function ϕ; and iii) Q is chosen to maximize profits, the (z^*, h^*) that solve Equation (2) are:*

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

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

Proof. See Appendix F. □

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

I.2 Endogenizing labor supply in an equilibrium search model

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

In this section, I show that the prediction of the fixed-job model that the FLSA has litte to no effect on h^* or z^* is robust to embedding Equation (2) into an extension of the Burdett and Mortensen (1998) model of equilibrium with on-the-job search.[3] In the context of the search model, the only effect of the overtime rule on the distribution of h^* is mediated

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

through the minimum wage, which rules out some of the (z^*, h^*) pairs that would occur in the unregulated equilibrium. In a numerical calibration, this effect is quite small, suggesting that equilibrium effects play only a minor role in how the FLSA overtime rule impacts anticipated hours or straight-time wages. This motivates the strategy in Section 4.4, in which z^* and h^* are treated as fixed when considering the impact of the FLSA on straight-wages.

I.2.1 The model

I focus on a minimal extension of Burdett and Mortensen (1998) that takes firms to be homogeneous in their technology and workers to be homogeneous in their tastes over the tradeoff between income and working hours.[4] Let there be a large number N_w of workers and large number N_f of firms, and define $m = N_w/N_f$.[5] Formally, we model this as a continuum of workers with mass m, and continuum of firms with unit mass. Firms choose a value of pay z and hours h to apply to all of their workers. Each period, there is an exogenous probability λ that any given worker receives a job offer, drawn uniformly from the set of all firms. Employed workers accept a job offer when they receive an earnings-hours package that they prefer to the one they currently hold, where preferences are captured by a utility function $u(z, h)$ taken to be homogeneous across workers and strictly quasiconcave, where $u_z > 0$ and $u_h < 0$. If a worker is not currently employed, they leave unemployment for a job offer if $u(z, h) \geq u(b, 0)$, where b represents a reservation earnings level required to incent a worker to enter employment. Workers leave the labor market with probability δ each period, and an equal number enters the non-employed labor force.

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

$$F(z, h) := P_{ZH}(u(Z_j, H_j) \leq u(z, h)) \tag{4}$$

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

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

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

of new workers to a given firm each period, we define the related quantity $G(z,h)$ – the fraction of employed workers that are at inferior firms to (z,h). In a steady state, note that $G(z,h)$ 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

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

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

$$N(z,h) = R(z,h)/s(z,h) = \frac{m\delta\lambda}{(\delta+\lambda(1-F(z,h))^2} \tag{5}$$

Eq. (5) is analogous to the baseline Burdett and Mortensen model without hours, with the quantity $F(z,h)$ playing the role of the firm-level CDF of wages from the baseline model.

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

$$\pi(z,h) = N(z,h)(p(h)-z) = m\delta\lambda\cdot\frac{p(h)-z}{(\delta+\lambda(1-F(z,h))^2} \tag{6}$$

where the function $p(h)$ corresponds to net revenue per worker $e(h)-\psi$, with $e(h)$ being some weakly concave and increasing function with $e(0)=0$, and ψ 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^*:

$$F(z,h) = 1+\frac{\delta}{\lambda}-\sqrt{\frac{m\delta}{\lambda}\cdot\frac{p(h)-z}{\pi^*}} \tag{7}$$

It follows from Eqs. (7) and (5) that we can rank firms in equilibrium by $F(z,h)$ and therefore by size according to the quantity $z-p(h)$. Since Eq. (5) 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).[6]

To fully characterize the equilibrium, first note that P_{ZH} can put a positive density on at most one point along each isoquant of $z - p(h)$, given that utility is strictly quasiconcave but $z - p(h)$ is weakly convex. Offers in the support of P_{ZH} thus lie along a one dimensional path through $\mathbb{R}^2$, and we can parametrize them by a scalar $t \in [0, 1]$, such that $\text{supp}(P_{ZH}) = \{(z(t), h(t))\}_{t\in[0,1]}$ and $t = F(z(t), h(t))$. Observe that each $(z(t), h(t))$ must pick out the point along its respective isoquant of $z - p(h)$ which delivers the highest utility to workers, i.e.:

$$(z(t), h(t)) = \text{argmax}_{z,h} u(z, h) \text{ s.t. } z - p(h) = \eta(t) \tag{8}$$

where $\eta(t) = \frac{\pi^* \lambda}{m\delta}(1 - \frac{t}{1+\delta/\lambda})^2$ is the value of $p(h(t)) - z(t)$ such that $F(z(t), h(t)) = t$ according to Eq.(7), viewed as a function of t.[7] The slope of the path $(z(t), h(t))$ can be derived from the first order condition for the above problem and the implicit function theorem:

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

If preferences were quasilinear in income, then the curve AB shown in Figure 4 would be a vertical line rising from point A and there would be no hours dispersion in equilibrium (as in Hwang et al., 1998). Figure 4 instead depicts the path $\{(z(t), h(t))\}_{t\in[0,1]}$ for a generic case in which preferences are neither homothetic nor quasilinear. If preferences were homothetic AB would be a straight line.

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 4.[8] I assume that the marginal rate of substitution between compensation and hours is less than $p'(0)$ at $(z, h) = (b, 0)$ (such that there are gains from trade) and increases continuously with h, eventually passing $p'(h)$ at some point h^*. This point is unique given strict quasiconcavity of $u(\cdot)$.

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.

[6] 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. (6).

[7] If instead we had $u(z(t), h(t)) < \max_{(z,h):z-p(h)=F^{-1}(t)} u(z, h)$, then any firm located at $(z(t), h(t))$ could profitably deviate to the argmax while keeping profits per worker constant but increasing their labor supply by attracting workers from $(z(t), h(t))$.

[8] 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.

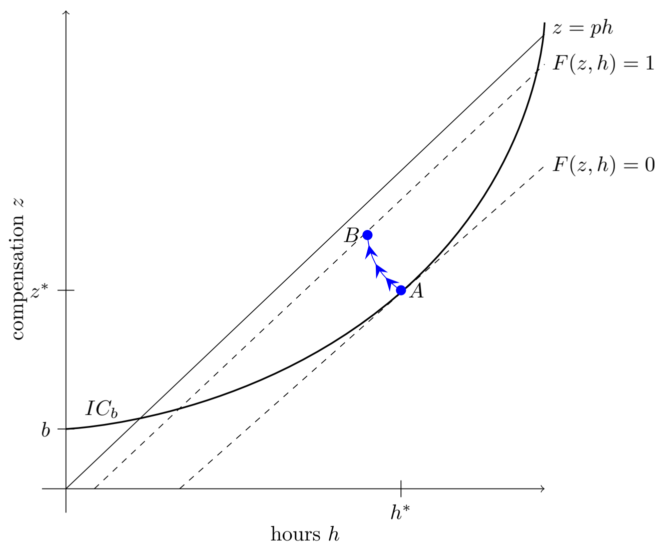

Figure 4: The support of the equilibrium distribution of compensation-hours offers (z, h) lies along the arrowed (blue) curve AB. Figure shows the case of perfect hours substitutability $p(h) = ph$. Plain curve IC_b is the indifference curve passing through the unemployment point $(b, 0)$. The least desirable firm in the economy lies at the pair (z^*, h^*), labeled by A, where IC_b has a slope of p. The other points chosen by firms are found by starting at point A and moving in the direction of higher utility, while maintaining a marginal rate of substitution of p between hours and earnings. This path intersects the line of solutions to $F(z, h) = 1$ given Eq. (7) at point B. Note that this line still lies below the zero profit line $z = ph$, as firms make positive profit.

Using that $F(z^*, h^*) = 0$ and $\pi^* = \pi(z^*, h^*)$, we can rewrite Equation (7) as:

$$F(z,h) = \left(1 + \frac{\delta}{\lambda}\right)\left[1 - \sqrt{\frac{p(h) - z}{p(h^*) - z^*}}\right] \tag{9}$$

The firms at point B in Figure 4 thus solve $z - p(h) = \left(\frac{\delta}{\delta+\lambda}\right)^2 (z^* - p(h^*))$, and 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 $z = p(h)$.[9]

[9]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.

I.2.2 Effects of FLSA policies

Now consider the introduction of a minimum wage, which introduces a floor on the hourly wage $w := z/h$. I assume that the point (z^*, h^*) does not satisfy the minimum wage, so that the minimum wage binds and rules out part of the unregulated support of P_{ZH}. The left panel of Figure 5 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 4, which now follows a subset of the original path AB, reflecting a reduction in hours and an increase in total compensation.

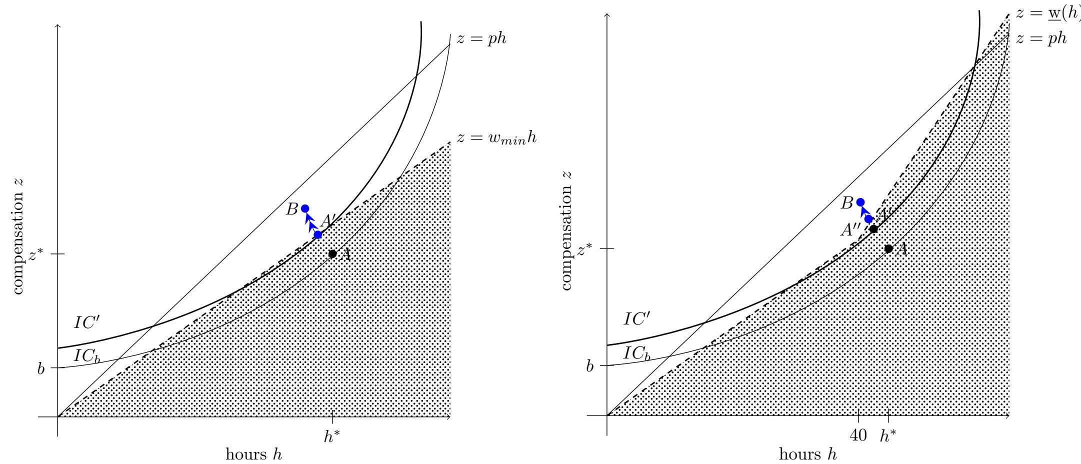

Figure 5: Left panel shows the support of the equilibrium distribution of compensation-hours offers (z, h) under a binding minimum wage. The compensation hours pairs that do not meet $\underline{w}$ are indicated by the shaded region. The lowest-wage offer in the economy moves from point A in the unregulated equilibrium to the point A' on the minimum wage line $z = \underline{w}h$ at which the marginal rate of substitution between compensation and hours equals p. Right panel shows how this effect is augmented when overtime premium pay for hours in excess of 40 is required, and A' lies at greater than 40 hours. In this case the support of P_{ZH} begins at point A'', which lies on the kinked minimum wage function $\underline{w}(h)$.

The right panel of Figure 5 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 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.

I.2.3 Calibration

To allow wealth effects in worker utility while facilitating calibration based on existing empirical studies, I take worker utility to follow the Stone-Geary functional form:

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

This simple specification allows a closed form solution to the path $(z(t), h(t))$, given by the following Proposition, which follows from the optimization problem (8) and working out the initial point $(z(0), h(0))$ in each policy regime.

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

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

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

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

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

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

I calibrate the model focusing on a lower-wage labor market where productivity is a constant $p = \$15$. I allow non-wage costs of $\psi = \$100$ a week, with the value based on estimates of benefit costs in the low-wage labor market.[10] I take $b = \$250$ corresponding to unemployment benefits that can be accrued at zero weekly hours of work.[11]. I calibrate

[10]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 $102.425 \approx 100$.

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

the factor λ/δ using estimates from Manning (2003) using the proportion of recruits from unemployment. Using Manning's estimates from the US in 1990 of about 55% of recruits coming from unemployment, this implies a value of $\lambda/\delta \approx 3$ in the baseline Burdett and Mortensen (1998) model.

To calibrate the preference parameters, I first pin down β from estimates of the marginal propensity to reduce earnings after random lottery wins (Imbens et al. 2001; Cesarini et al. 2017). Both of these studies report a value in the neighborhood of $\beta = 0.85$. I take a value of $\gamma_z = \$200$ as the level of consumption at which the marginal willingness to work is infinite, and take $\gamma_h = 50$ hours of work per week. I choose this value according to a rule-of-thumb as the average hours among full-time workers in the US (42.5), divided by β.[12]

Given these values, we can compute moments of functions of the joint distribution of compensation and hours using the Proposition and numerical evaluation of the integrals. Table 9 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 across workers is much closer to the $t = 1$ value than the $t = 0$ value (the largest firm is about 16 times as large as the smallest).

	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 9: 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 9 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

[12]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.

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.

J Additional identification results for the bunching design

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

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

J.1 A generalized notion of homogeneous treatment effects

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

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

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

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

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

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

$$\mathcal{B} = P(Y_{0i} \in [k, k+\Delta]) \text{ where } \Delta = G^{-1}\left(G(k)+\delta\right) - k \tag{10}$$

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

J.2 Parametric approaches with homogeneous treatment effects

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

Proposition J.1 (identification by linear interpolation, à la Saez 2010). *If $\delta_i^G = \delta$ for some G, $F_1(y)$ and $F_0(y)$ are continuously differentiable, and $f_0(y)$ is linear on the interval $[k, k+\Delta]$, then with CONVEX, CHOICE:*

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

Proof. See Section F. □

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

$$\mathcal{B} = \frac{\Delta}{2}\left\{\lim_{y\uparrow k} f(y) + \frac{k}{k+\Delta}\lim_{y\downarrow k} f(y)\right\} = \frac{k}{2}\left(\left(\frac{\rho_0}{\rho_1}\right)^{\epsilon} - 1\right)\left(\lim_{y\uparrow k} f(y) + \left(\frac{\rho_0}{\rho_1}\right)^{-\epsilon}\lim_{y\downarrow k} f(y)\right) \tag{11}$$

which serves as the main estimating equation from Saez (2010) (and can be solved for ϵ by the quadratic formula). The empirical approach of Saez (2010) can thus be seen as applying a result justified in a much more general model than the iso-elastic utility function assumed therein, provided that the researcher is willing to assume homogeneous treatment effects (possibly after some known transformation G, and/or conditional on observables).[13] Note that the linearity assumption of Proposition J.1 could be falsified by visual inspection: it implies that right and left limits of the derivative of the density of Y_i at the kink are equal.

A more popular approach, following Chetty et al. (2011), is to use a global polynomial approximation to $f_0(y)$, which interpolates $f_0(y)$ inwards from both directions across the missing region of unknown width Δ. This technique has the added advantage of accommodating diffuse bunching, for which the relevant $\mathcal{B}$ is the total "excess-mass" around k rather than a perfect point mass at k. I focus here on the simplest case in which bunching is exact, as in the overtime setting. The polynomial approach can be seen as a special case of the following result:

Proposition J.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 $\delta_i^G = \delta$ for some G and CONVEX and CHOICE hold. Then, if:*

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

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

$$f(y) = \begin{cases} g(y;\theta_0) & y < k \\ g(y+\Delta(\theta_0);\theta_0) & y > k \end{cases} \tag{12}$$

Proof. See Section F. □

The standard approach of fitting a high-order polynomial to $f_0(y)$ can satisfy the assumptions of Proposition J.2, since polynomial functions are analytic everywhere. Proposition J.2 yields an identification result that can justify an estimation approach similar to one often made

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

in the literature, based on Chetty et al. (2011).[14] 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 J.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.

J.3 Nonparametric approaches with homogeneous treatment effects

The additional assumptions from the preceding section have allowed for point-identification of causal effects under an assumption of homogenous treatment effects. These assumptions have taken the form of parametric restrictions on the density of counterfactual choices Y_{0i} in the missing region $[k, k+\Delta]$: that this density is linear or more generally fits a parametric family of analytic functions. As has been argued in Blomquist and Newey (2017), these parametric assumptions drive all of the identification, an undesirable feature from the standpoint of robustness to departures from them. I now explore some non-parametric assumptions about $f_0(y)$ that yield bounds on Δ in a model with homogeneous treatment effects.

For example, monotonicity of $f_0(y)$ has been suggested by Blomquist and Newey (2017) as an alternative assumption in the context of the iso-elastic model. A result based on monotonicity that allows heterogeneous treatment effects is presented in Section J.4. However, monotonicity may be restrictive if the density of Y_0 has a mode near the kink point. In this case, local log-concavity of $f_0(y)$ may be a more attractive assumption (concavity or convexity would be another). [15] Note that log-concavity is a stronger version of the bi-log-concavity assumption used in the main text, but still nests many common parametric distributions such as the uniform, normal, exponential extreme value and logistic. For simplicity, this result assumes homogeneous treatment effects in levels (rather than after applying a function G).

Proposition J.3 (bounds from log-concavity). *Suppose that $\Delta_i = \Delta$ and that $f_0(y)$ is log-concave in the interval $y \in [k, k+\Delta]$ and continuously differentiable at k and $k+\Delta$.*

[14] The estimation technique proposed by Chetty et al. (2011) ignores the shift term $\Delta(\theta)$ in Equation (12), a limitation discussed by Kleven (2016). This is perhaps less problematic in typical settings where the bunching is somewhat diffuse around the kink, in contrast to the overtime setting in which bunching is exact, and the slope of the density is far from zero near 40. A more robust estimation procedure for parametric bunching designs could be based on iterating on Equation (12) after updating $\Delta(\theta)$, until convergence. This presents an interesting topic for future research.

[15] 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.

Then, under CONVEX and CHOICE:

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

where

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

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

Proof. See Figure 6. Derivation of expressions available by request. □

Intuition for Proposition J.3 is provided in Figure 6. If $f_0(y)$ is log convex rather than log-concave in the missing region, then the bounds Δ^L and Δ^U can simply be swapped. Or, if we suppose that f_0 is *either* log-concave or log-convex locally: $\Delta \in [\min\{\Delta^U, \Delta^L\}, \max\{\Delta^U, \Delta^L\}]$.

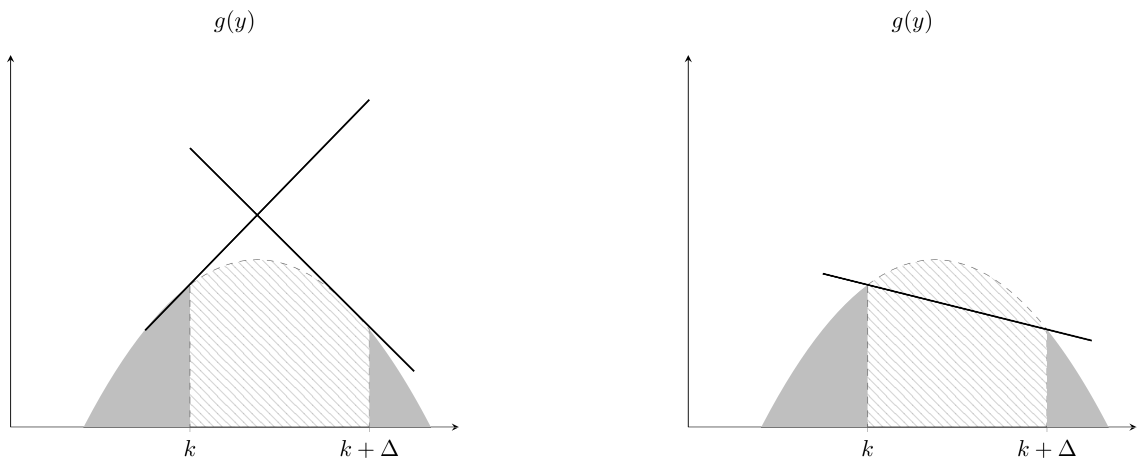

Figure 6: The left and right panels of this figure depict intuition for the lower and upper bounds on Δ in Proposition J.3. In both panels, the hatched region is the missing region $[k, k+\Delta]$ which contains known mass $\mathcal{B}$. The function plotted is $g(y)$, the logarithm of $f_0(y)$. Outside 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, yielding the analytic bounds in Proposition J.3.

J.4 Alternative identification strategies with heterogeneous treatment effects

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

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

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

Proposition J.4 (identification of an ATE under uniform density approximation). *Let Δ_i and Y_{0i} admit a joint density $f(\Delta, y)$ that is continuous in y at $y = k$. For each Δ assume that $f(\Delta, Y_0) = f(\Delta, k)$ for all Y_0 in the region $[k, k + \Delta]$. Under Assumptions WARP and CHOICE*

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

with equality under CONVEX.

Proof. Note that

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

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

Lemma SMALL in Appendix A formalizes the idea that the uniform density approximation from Proposition J.4 becomes exact in the limit of a "small" kink.

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

Proposition J.5 (monotonicity with heterogeneous treatment effects). *Assume CONVEX and CHOICE, and suppose the joint density $f_0(\Delta, y)$ of Δ_i and Y_{0i} and the joint density $f_1(\Delta, y)$ of Δ_i both exist. Suppose first that $f_0(\Delta, y)$ is weakly increasing on the interval $y \in [k, k + \Delta]$ for all Δ in the support of Δ_i. Then*

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

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

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

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

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

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

□

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

Other approaches to identification with heterogeneous treatment effects are possible when the researcher observes covariates X_i that are unaffected by a counterfactual shift between B_1 and B_0. For example, assuming that $\mathbb{E}[X_i|Y_{0i} = y]$ or $\mathbb{E}[X_i|Y_{1i} = y]$ 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 $\mathbb{E}[X_i|Y_i = y]$ at $y = k$. Another strategy for using covariates would be to model the potential outcomes Y_{0i} and Y_{1i} as functions of them. If we are willing to suppose that

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

with U_{1i} and U_{0i} each statistically independent of X_i, then the censoring of the distributions of Y_{0i} and Y_{1i} in Lemma B.1 can be "undone", following the results of Lewbel and Linton (2002).[16]. This would allow estimation of the unconditional average treatment effect as well as quantile treatment effects at all levels. However, the assumption that U_0 and U_1 are independent of X is quite strong.

J.5 Two bunching design settings from the literature

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

Example 1: Labor supply with taxation

Individuals have preferences $\tilde{u}_i(c, y)$ over consumption c, and labor earnings y, where ϵ_i is a vector of parameters capturing heterogeneity over the disutility of labor, labor productivity, etc. The agent's budget constraint is $c \le y - B(y)$ where $B(y)$ is income tax as a function of pre-tax earnings y. $\tilde{u}_i(c, y)$ is taken to be strictly quasi-concave in (c, y) for each i as the opportunity cost of leisure rises with greater earnings, and monotonically increasing in consumption. Define $z = y - c$ to be tax liability, and let $u_i(z, y) = \tilde{u}_i(y - z, y)$ which is monotonically decreasing in tax. Individuals now choose a value of y (e.g. by adjusting

[16]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 $F_1(y) = P(Y_{1i} \le y)$ and $F_0(y) = P(Y_{0i} \le y)$ by deconvolution, and then estimate causal effects.

hours of work, number of jobs, or misreporting) given a progressive tax schedule $B_k(y) = \tau_0 y + 1(y \geq k)(\tau_1 - \tau_0)(y - k)$, with the kink arising from an increase in marginal tax rates from τ_0 to $\tau_1 > \tau_0$ at $y = k$. The budget functions are $B_0(y) = \tau_0 y$, $B_1(y) = \tau_1 y - (\tau_1 - \tau_0)k$, and the kinked budget constraint can bewritten $z \geq B_k(y) = \max\{B_0(y), B_1(y)\}$.

Example 2: Minimum tax schemes

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

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

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

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

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

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

$$u_i(z, r, \hat{w}) = r - w_i(r) - g_i(\hat{w} - w_i(r)) - z \tag{13}$$

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

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

$$\begin{aligned}
\Delta &\approx \left.\frac{dy(r,\hat{w})}{\hat{w}}\right|_{(r_0,\hat{w}_0)} \Delta_{\hat{w}} + \left.\frac{dy(r,\hat{w})}{r}\right|_{(r_0,\hat{w}_0)} \Delta_r \\
&= \frac{\hat{w}_0}{r_0^2}\Delta_r - \frac{1}{r_0}\left(\Delta_{w(r)} + \Delta_{(\hat{w}-w(r))}\right) \\
&\approx \frac{\hat{w}_0}{r_0^2}\Delta_r - \frac{1}{r_0}\left(w'(r_0)\Delta_{ri} + \Delta_{(\hat{w}-w(r))}\right) \\
&= \frac{1}{r_0}\left\{(1 - Y_0 - w'(r_0))\Delta_r \Delta_{\hat{w}}\right\} \approx \frac{1}{r}\left\{-k\Delta_r - \Delta_{(\hat{w}-w)}\right\} \\
&\approx \frac{1}{r_0}\left\{-\frac{\tau_r}{\tau_\pi}\cdot r\epsilon^r \frac{d(1-\tau_E)}{\tau_E} - \Delta_{\hat{w}i}\right\} = \frac{\tau_r^2}{\tau_\pi}\epsilon^r - \frac{\Delta_{(\hat{w}-w)}}{r_0}
\end{aligned} \tag{14}$$

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 (14) 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 B.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.

K Further proofs

K.1 Proof of Proposition C.1

Note: this proof follows the notation of Y_i from Appendix B, 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 B.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 B.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:

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

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

Second claim:

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

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

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

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

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

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

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

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

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

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 $\lim_{y\uparrow k} f(y) = P(W_{it} = 0) \lim_{y\uparrow k} f_{0|W=0}(y)$.

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

K.2 Proof of Theorem 1*

Note: this proof follows the notation of Y_i from Appendix B, 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:

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

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

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

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

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

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

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

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

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

Let

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

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

- $P(Y_i < k) = P(Y_{0i} < k \text{ and } W_i = 0) + A$

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

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

K.3 Proof of Proposition I.1

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

K.4 Proof of Proposition J.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.

K.5 Proof of Proposition J.2

We seek a Δ such that for some θ_0:

$$\mathcal{B} = \int_{\tilde{k}}^{k+\Delta} g(y;\theta_0)dy \tag{15}$$

and

$$f(y) = \begin{cases} g(y;\theta_0) & y < k \\ g(y+\Delta;\theta_0) & y > k \end{cases} \tag{16}$$

and

$$g(y;\theta_0) > 0 \text{ for all } y \in [k, k+\Delta] \tag{17}$$

Recall from Equation (10) 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 (15), we would need $\Delta' = \Delta(\theta')$ and $\Delta = \Delta(\theta)$ for $\theta, \theta' \in \Theta$, where $\Delta(\theta)$ is the Δ that satsifies Eq. (15) for a given θ, (which is unique for each permissible value of θ since $g(y;\theta_0) > 0$. To satisfy (16), we would also need

$$g(y;\theta) = \begin{cases} f(y) & y < k \\ f(y-\Delta(\theta)) & y > k+\Delta(\theta) \end{cases} \qquad g(y;\theta') = \begin{cases} f(y) & y < k \\ f(y-\Delta(\theta')) & y > k+\Delta(\theta') \end{cases} \tag{18}$$

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 (18), 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 $f_0(y)$.

K.6 Details of calculations for policy estimates

K.6.1 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:

$$\begin{aligned}\mathbb{E}\left[\ln Y_{0i} - \ln Y_{1i}|Y_i = k, K_i^* = 0\right] &= \mathbb{E}\left[\ln Y_{0i} - \ln Y_{1i}|Y_{0i} \in [k, Q_{0|K^*=0}(F_{1|K^*=0})], K_i^* = 0\right] \\ &= \int_{\rho_0}^{\rho_1} d\rho \cdot \mathbb{E}\left[\frac{dY_i(\rho)}{d\rho}\frac{1}{Y_i(\rho)}\middle| Y_{0i} \in [k, k+\Delta_0^*], K_i^* = 0\right] \\ &= \int_{\rho_0}^{\rho_1} d\ln\rho \cdot \frac{1}{\mathcal{B}^*}\int_k^{k+\Delta_0^*} dy \cdot f_0(y) \cdot \mathbb{E}\left[\frac{dY_i(\rho)}{d\rho}\frac{\rho}{Y_i(\rho)}\middle| Y_{0i} = y, K_i^* = 0\right] \quad (19) \\ &= \mathcal{E}\int_{\rho_0}^{\rho_1} d\ln\rho = \mathcal{E}\ln(\rho_1/\rho_0)\end{aligned}$$

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

K.6.2 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:

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

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

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

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

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

K.6.3 Effect of a change to the kink point on bunching

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

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

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

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

The lemma in the next section gives identified bounds on the counterfactual quantity that appears in the expression in each case.

K.6.4 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

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

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

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

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

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

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

and for any $y' < y_0$:

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

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

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

and

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

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

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':

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

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

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

Write the quantity of interest as:

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

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

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

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

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

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

Similarly, a lower bound is:

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

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

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

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

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

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

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

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

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

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

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.

K.7 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 7 below.

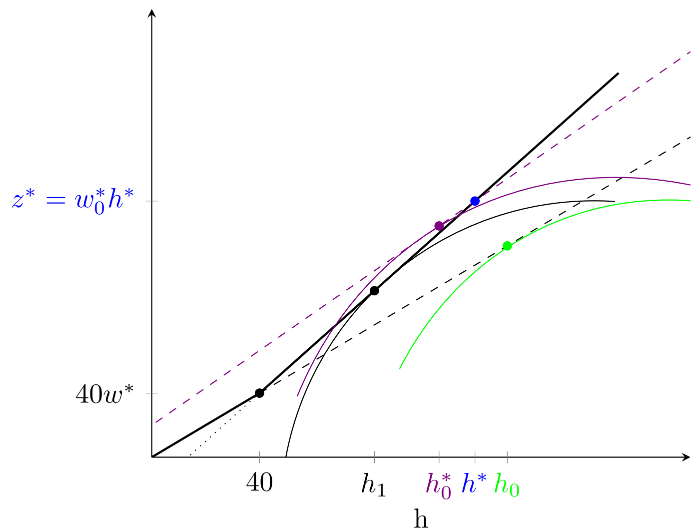

Figure 7: Depiction of h^*, h_0, h_0^* and h_1 for a single fixed worker that works overtime at h_1 hours this week. Their realized wage w^* has been set to yield earnings z^* based on anticipated hours h^* given the FLSA kink. In a world without the FLSA, the worker's wage would instead be $w_0^* = z^*/h^*$, 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 h^* 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: $w' h^* = w^*(k^* + 0.5(h^* - k^*))$ if h^* is between k and k'. The percentage change for these workers is thus

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

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

Additional 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.

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

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.

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.

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

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

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

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

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

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