The Effect of Maternal Labor Supply on Children: Evidence from Bunching

Carolina Caetano, Gregorio Caetano,
Eric Nielsen and Viviane Sanfelice

UGA/Fed Board/Temple Univ.

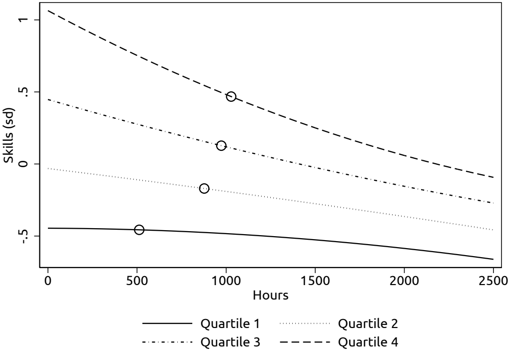
Motivation

- Maternal labor supply has increased in recent decades (Eckstein and Lifschitz 2011, Fogli and Veldkamp 2011).
- Quality parent-child interactions known to be important for child development (e.g. Todd and Wolpin 2007).
- How might maternal labor supply affect children in the short-run?
  1. time channel: more time at work $\Longrightarrow$ less time at home
  2. income channel: more time at work $\Longrightarrow$ more income
- Many policies (family leave, paid childcare, child tax credits, other tax changes, etc.) might affect maternal labor supply.

This Paper

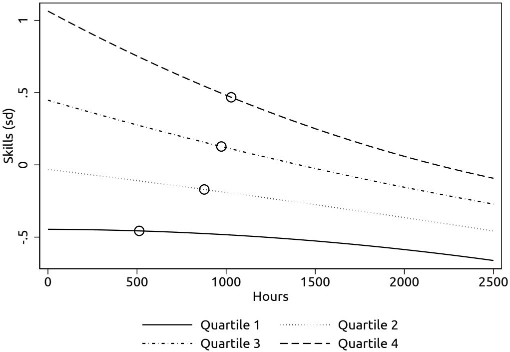
- We estimate the effect of maternal work hours during first 3 years of life on children's skills around age 6.
- We use a new approach leveraging bunching at zero to deal with endogeneity (Caetano, Caetano and Nielsen 2021).
- We focus on heterogeneous effects:
  - by the skill of the mother
  - by the quantity of her labor supply
- Why mothers and not all parents?
  - data limitations, prior literature
  - greater variation in maternal labor supply

Preview of Results: by Quartile of Maternal AFQT



- Negative effects on children's cognitive skills in the short-run.
- Less negative for less skilled mothers, except those who work long hours.

Preview of Results: by Quartile of Maternal AFQT



- Why is higher skilled maternal labor supply so detrimental for children's skills in the short-run?
  - ~~Last hour is more costly for those working longer hours.~~
  - Money insufficient in the short-run to compensate for high-skill mother-child interactions.

Additional Heterogeneity: by Pre-Birth Maternal Wage

- Think of two similarly skilled mothers with different earnings (e.g. through college major or occupation).
- Try to vary income holding quality of home interaction constant. The skills that separate them are well-valued in the job market, but not necessarily in interactions with a young child.
- Money helps, but is not enough: Even highly-skilled, highly-paid mothers mothers cannot fully compensate for their absence with money.
- Flexible schedules and work-from-home may be a better policies than giving financial incentives to work.

Literature

- NLSY79/CNLSY. Effect of working hours while child is 0-3 on child's early outcomes.
  - Negative effects:

    Ruhm 2004, Bernal 2008, Desai et al. 1989, Baydar and Brooks-Gunn 1991, Hill and O'Neill 1994, Baum 2003
  - Zero effects:

    James-Burdumy 2005, Parcel and Menaghan 1994, Blau et al. 1992, Waldfogel et al. 2002
  - Positive effects:

    Vandell and Ramanan 1992 (low-income families)
- More recent literature: focus on time-money trade-off
  - Negative time effects not fully offset by income effects:

    Agostinelli and Sorrenti (2021): NLSY79/CNLSY. Effect of working hours while child is 4-16 on contemporaneous child outcome.
  - Income effects fully offset negative time effects:

    Nicoletti, Salvanes and Tominey (2022): Norwegian data. Effect of working hours while child is 0-5 on outcome when child is 11-15.

Data

NLSY79 (Mothers)

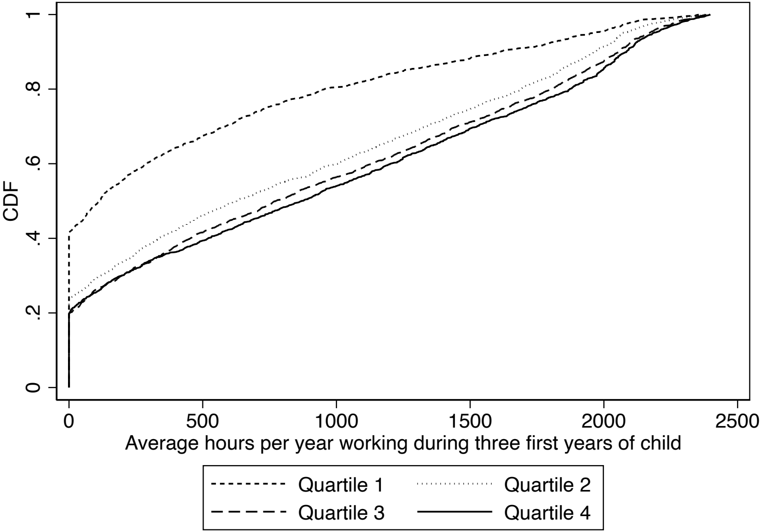
- AFQT, education, marital status, age at birth, hh income
- average hours worked over three years following birth
- exclude three months immediately after birth (maternity leave)
- family structure

CNLSY (Children)

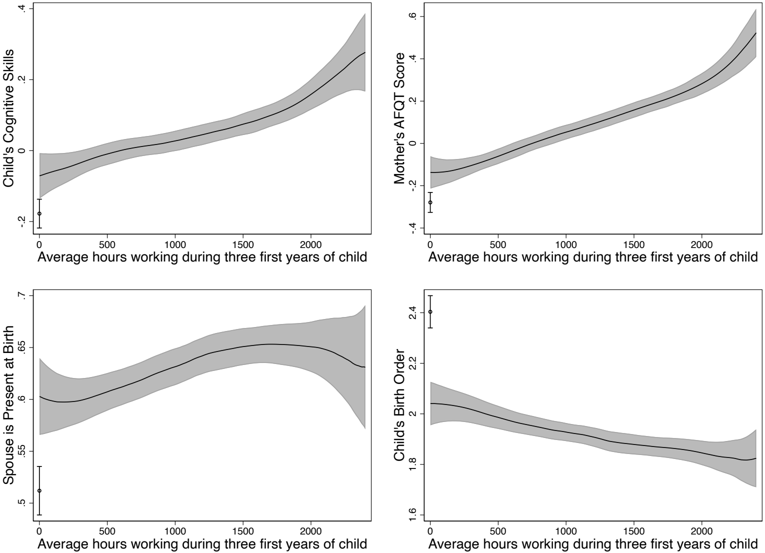
- race/ethnicity, sex
- cognitive skills – iterated principal factor analysis applied to PIAT math and reading scores, administered around age 6

Final sample consists of 6,924 mother-child pairs.

Bunching of Labor Supply



Selective Bunching



Intuition (No Controls)

Given skills S, labor supply L, and *desired* labor supply L^*,

$$S = \alpha + \beta L + \delta L^* + \epsilon.$$

$L^* = L$ when $L^* \geq 0$. At $L = 0$, L^* indexes possibly confounding unobservables.

What is the effect of increasing hours from $L = l_0$ to $L = l_1$?

$$\underbrace{\mathbb{E}[S|L = l_1, L^* = l_1] - \mathbb{E}[S|L = l_0, L^* = l_0]}_{\text{what we observe}} =$$

$$\underbrace{\mathbb{E}[S|L = l_1, L^* = l_1] - \mathbb{E}[S|L = l_0, L^* = l_1]}_{\text{marginal treatment effect}} +$$

$$\underbrace{\mathbb{E}[S|L = l_0, L^* = l_1] - \mathbb{E}[S|L = l_0, L^* = l_0]}_{\text{selection bias}}$$

Intuition (No Controls) RDD

Usual approach turns off selection term. We turn off treatment effect term. At $L = 0$:

- treatment does not vary
- remaining variation in S informative about effect of unobservables (indexed by L^*) on skills
- if this "selection" effect can be identified, so can the treatment effect of interest

Like an "upside-down" RDD:

- "running variable" L varies continuously with L (trivially)
- unobservables L^* discontinuous at $L = 0$
- outcome discontinuities at $L = 0$ due to L^*
- dist. assumptions provide "first stage" of upside-down RDD

A Model of Constrained Choice

$$S = \beta L + g(X) + \epsilon, \quad \mathbb{E}[\epsilon | L, X] \neq 0$$

A Model of Constrained Choice

$$S = \beta L + g(X) + \overbrace{\delta\eta + \varepsilon}^{\epsilon}, \quad \mathbb{E}[\varepsilon | L, X, \eta] = 0$$

A Model of Constrained Choice

$$S = \beta L + g(X) + \overbrace{\delta\eta + \varepsilon}^{\epsilon}, \quad \mathbb{E}[\varepsilon | L, X, \eta] = 0$$

$$L^* = h(X) + \eta$$

$$L = \max\{0, L^*\}$$

A Model of Constrained Choice

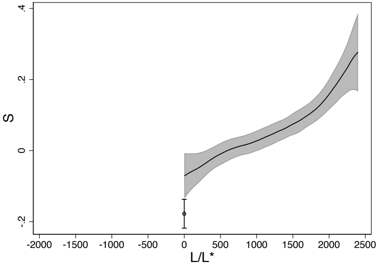
$$S = \beta L + g(X) + \overbrace{\delta \eta + \varepsilon}^{\epsilon}, \quad \mathbb{E}[\varepsilon | L, X, \eta] = 0$$

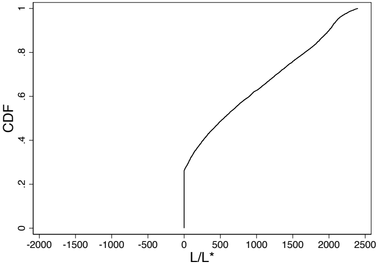
$$L^* = h(X) + \eta$$

$$L = \max\{0, L^*\}$$

$$\mathbb{E}[S|L,X] = \beta L + \underbrace{g(X) - \delta h(X)}_{m(X)} + \delta \underbrace{\left[L + \mathbb{E}[L^*|L=0,X]1(L=0)\right]}_{\text{new regressor}}$$

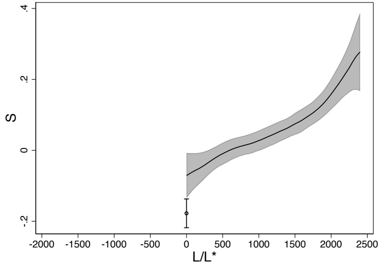
Identification with Distributional Assumptions

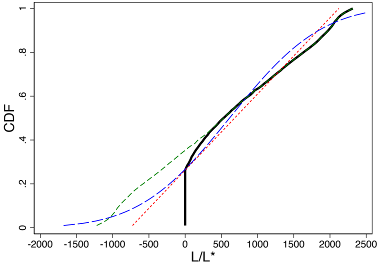




$$
\begin{aligned}
S =& \beta L + g(X) + \delta\eta + \varepsilon, \quad \mathbb{E}[\varepsilon | L, X, \eta] = 0 \\
L^* =& h(X) + \eta \\
\mathbb{E}[S|L, X] =& \beta L + m(X) + \delta \underbrace{\left[L + \mathbb{E}[L^* | L = 0, X] 1(L = 0)\right]}_{\text{new regressor}}
\end{aligned}
$$

Identification with Distributional Assumptions



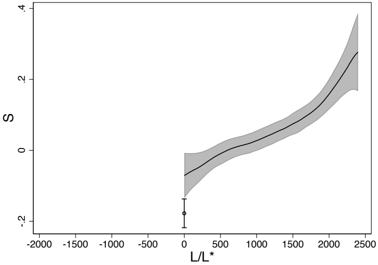


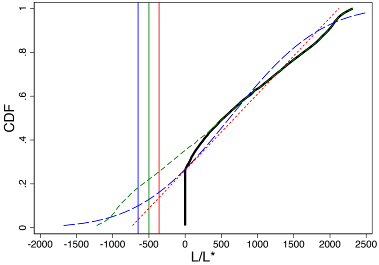
$$S = \beta L + g(X) + \delta\eta + \varepsilon, \quad \mathbb{E}[\varepsilon | L, X, \eta] = 0$$

$$L^* = h(X) + \eta, \qquad \eta \sim \text{Normal, Uniform, or Symmetric}$$

$$\mathbb{E}[S|L,X] = \beta L + m(X) + \delta \underbrace{[L + \mathbb{E}[L^*|L=0,X]1(L=0)]}_{\text{new regressor}}$$

Identification with Distributional Assumptions





$$S = \beta L + g(X) + \delta\eta + \varepsilon, \quad \mathbb{E}[\varepsilon|L, X, \eta] = 0$$

$$L^* = h(X) + \eta, \qquad \eta \sim \text{Normal}, \text{Uniform}, \text{ or Symmetric}$$

$$\mathbb{E}[S|L,X] = \beta L + m(X) + \delta \underbrace{[L + \mathbb{E}[L^*|L=0,X]1(L=0)]}_{\text{new regressor}}$$

$$S = \beta L + g(X) + \delta\eta + \varepsilon$$

	(i) No Controls	(ii) Controls	(iii) Het. Tobit	(iv) Het. Uniform	(v) Het. Symmetric
β	0.014**	0.000	-0.016**	-0.019**	-0.019**
	(0.001)	(0.001)	(0.005)	(0.006)	(0.005)
δ			0.014**	0.017**	0.017**
			(0.004)	(0.005)	(0.005)

empirical details

- +10 hrs/wk (over 3 years) lowers cog skills by 0.1 s.d.
- Context – 0.1 sd similar to
  - effect of a 1 sd improvement in teacher quality
  - effect of a 35% class size reduction
  - maternal labor effects in prior literature
- Non-cognitive effects directionally similar but not significant

Robustness

Identifying assumptions are (partially) testable.

- $\mathbb{E}[\epsilon|L, X] = 0$ linearity
- $\mathbb{E}[\varepsilon|L, X, \eta] = 0$ truncation
- KS tests almost never reject full symmetry

We also assess the robustness/plausibility of our estimates.

- alternative samples diff. samples
- plausible degree of selection Oster (2019)
- violations in the distributional assumption dist. robustness
- alternative ways of controlling for observables clusters
- more flexible models for δ. $\delta(X)$

Heterogeneity by Maternal Skill and Hours

$$S = f(L, X; \theta) + g(X) + \delta(X)\eta + \varepsilon$$

$A =$ AFQT score of the mother

$$f(L, X; \theta) = (\beta + \beta_A A + \beta_L L + \beta_{AL} A L)L$$
$$\delta(X) = \delta + \delta_A A$$

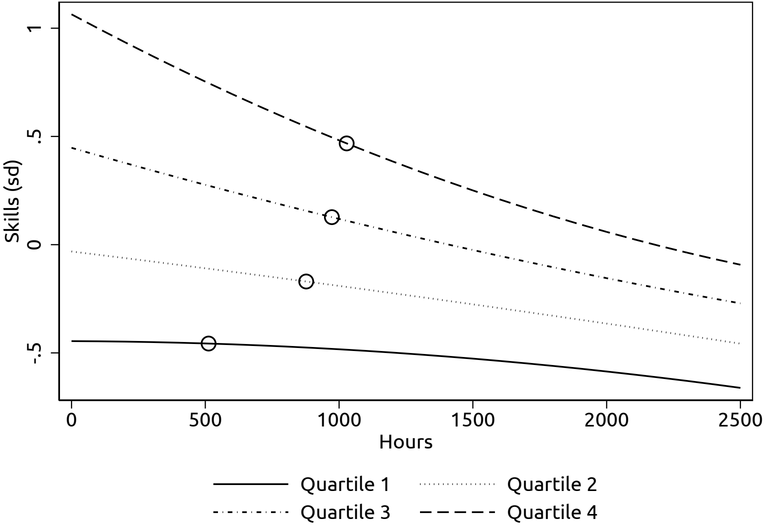
- $(\beta, \beta_A, \beta_L, \beta_{AL}, \delta, \delta_A)$ are identified (Caetano et al. 2021)
- higher order terms identified but not significant

Heterogeneity by Maternal Skill and Hours

	(i) No Controls	(ii) Controls	(iii) Het. Tobit	(iv) Het. Uniform	(v) Het. Symmetric
β	0.018**	0.003	-0.023**	-0.026**	-0.030**
	(0.004)	(0.003)	(0.008)	(0.010)	(0.009)
β_A	0.047**	-0.010**	-0.017**	-0.021**	-0.026**
	(0.003)	(0.004)	(0.008)	(0.009)	(0.009)
β_L ($\times$1000)	-0.004**	-0.001	0.002	0.001	0.002
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
β_{AL} ($\times$1000)	-0.015**	0.003**	0.004**	0.004**	0.004**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
δ			0.016**	0.022**	0.023**
			(0.005)	(0.007)	(0.006)
δ_A			0.005	0.009	0.013**
			(0.005)	(0.006)	(0.006)

- $\hat{\beta}_A < 0$ skill-intensity of skill production
- $\hat{\beta}_L > 0, \ \hat{\beta}_{AL} > 0$ positive income effects
- $\hat{\delta}_A > 0$ more positive selection for more skilled mothers

Visualizing the Heterogeneous Effects



Further Heterogeneity: by Pre-Birth Wage

$$S = f(L, X; \theta) + g(X) + \delta(X)\eta + \varepsilon$$

$A =$ AFQT score, $W =$ Pre-birth wage

$$f(L, X; \theta) = (\beta + \beta_A A + \beta_W W + \beta_L L + \beta_{AL} AL + \beta_{WL} WL)L$$
$$\delta(X) = \delta + \delta_A A + \delta_W W$$

- $(\beta, \beta_A, \beta_W, \beta_L, \beta_{AL}, \beta_{WL}\delta, \delta_A, \delta_W)$ are identified
- For two mothers with the same skills, is the effect less negative for the higher-wage mother?

Heterogeneity by Maternal Skill, Wages, and Hours

	(i) No Controls	(ii) Controls	(iii) Het. Tobit	(iv) Het. Uniform	(v) Het. Symmetric
β	-0.004	-0.002	-0.009	-0.013	-0.017
	(0.005)	(0.004)	(0.014)	(0.026)	(0.017)
β_A	0.039**	-0.016**	-0.027**	-0.047**	-0.035**
	(0.004)	(0.004)	(0.014)	(0.023)	(0.017)
β_W	0.007*	0.001	0.014	0.029	0.018
	(0.004)	(0.005)	(0.016)	(0.026)	(0.019)
β_L ($\times$1000)	0.002	0.001	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
β_{AL} ($\times$1000)	-0.012**	0.005**	0.006**	0.007**	0.007**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
β_{WL} ($\times$1000)	-0.002	0.001	-0.000	-0.001	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
δ			0.006	0.010	0.013
			(0.010)	(0.022)	(0.013)
δ_A			0.008	0.027	0.015
			(0.009)	(0.020)	(0.013)
δ_W			-0.009	-0.024	-0.014
			(0.010)	(0.021)	(0.014)

Conclusion – Results

- Effect of maternal hours worked in a child's first 3 years on cognitive skills around age 6: negative on average, very negative for high-skilled mothers.
  - Last hour of work is more costly the longer the mother works?

    No – effects are close to linear and slightly convex for high-skilled mothers.
  - Incremental earnings insufficient to offset direct time effect?

    Yes, even for high-skilled, high-wage mothers. Estimates are noisy, though.
- Control for endogeneity using novel approach leveraging bunching of mothers at zero labor supply.
  - does not require instruments or special data structures
  - complementary to other identification methods
  - broadly applicable in empirical settings

Conclusion – Discussion

- What work-promoting policies would avoid negative effects on children's skills?
  - For low-skilled mothers: no negative effects unless mother works close to full time.
  - Increasing the financial rewards to working would likely be ineffective for higher-skilled mothers.
  - Flexible work arrangements may be particularly helpful for higher-skilled mothers.

    maintain working hours and income while increasing time spent with children
  - Flexible work for partners may be complementary.

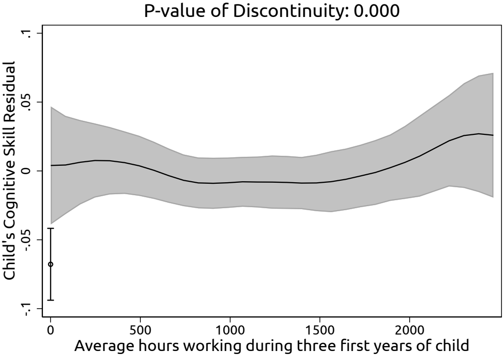
Thank you!

Estimation Details

Back

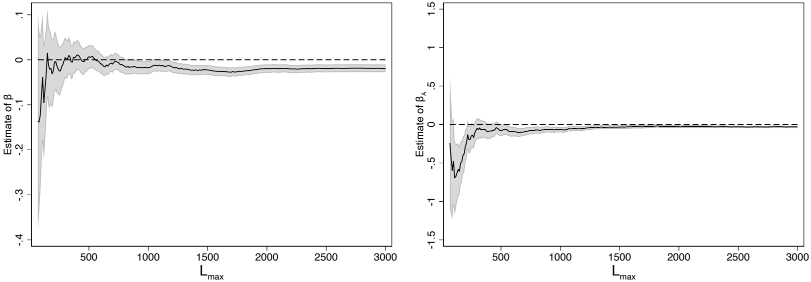
- It is a good idea to discretize X first, then use cluster indicators to approximate $m(X)$, $\delta(X)$ and $\mathbb{E}[L^*|L^* \leq 0, X]$.
- We use hierarchical cluster with 50 clusters in all reported results.
- Intuition: as the number of clusters increase, these approximations improve.
- See Caetano, Caetano and Nielsen (2021) for details.

Testing Assumption $\mathbb{E}[\epsilon|L, X] = 0$ Back



$$S = \beta L + X'\tau + \sum_{k=1}^{K} \alpha_k \mathbf{1}(X \in \hat{\mathcal{C}}_k) + \delta[L + \hat{\mathbb{E}}[L^*|L = 0, \hat{C}_K]\mathbf{1}(L = 0)]$$

Testing Assumption $\mathbb{E}[\varepsilon|L, X, \eta] = 0$ Back



Restrict the sample to $L \leq L_{\text{max}}$ for different values of L_{max}.

Alternative Samples/Models ▸ Back

		(i) Uncorrected No Controls	(ii) Uncorrected Controls	(iii) Het. Tobit	(iv) Het. Uniform	(v) Het. Symmetric
Not Single	β	0.014**	0.000	-0.016**	-0.019**	-0.019**
		(0.001)	(0.001)	(0.005)	(0.006)	(0.005)
	δ			0.014**	0.017**	0.017**
				(0.004)	(0.005)	(0.005)
Not College	β	0.012**	0.001	-0.015**	-0.019**	-0.018**
		(0.001)	(0.001)	(0.005)	(0.006)	(0.006)
	δ			0.014**	0.018**	0.017**
				(0.004)	(0.005)	(0.005)
Other Income	β	0.013**	-0.000	-0.016**	-0.020**	-0.019**
		(0.001)	(0.001)	(0.006)	(0.007)	(0.007)
	δ			0.013**	0.017**	0.017**
				(0.005)	(0.006)	(0.006)
$\delta(X)$ by Cluster	β	0.014**	0.000	-0.014**	-0.019**	-0.017**
		(0.001)	(0.001)	(0.005)	(0.007)	(0.006)

Independent Evidence of Selection: Oster (2019)

Back

Consider a regression of S on L, X, and *all confounders*, W.

- R_{max} is the R^2 of this hypothetical regression

Fix the true value of β:

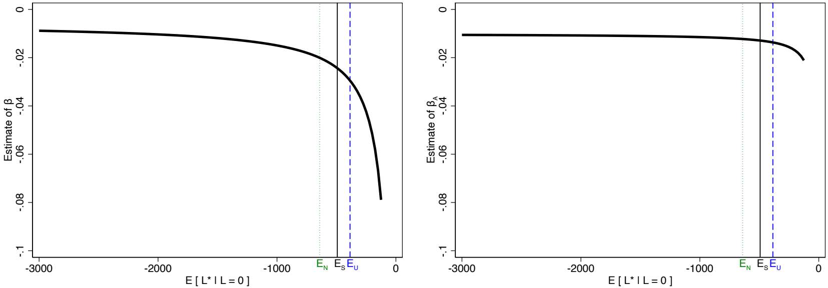
- let $\hat{\beta}$ be the OLS estimate of L in the uncorrected regression
- What value of $\delta_{\text{Oster}} = \left(\frac{\sigma_{L,W}}{\sigma_W^2}\right) / \left(\frac{\sigma_{L,X}}{\sigma_X^2}\right)$ has plim $\hat{\beta} = 0.000$?

	True $\beta =$ -0.019					
R_{max}	0.50	0.60	0.70	0.80	0.90	1.00
δ_{Oster}	1.12	0.82	0.65	0.54	0.46	0.40

Observing an uncorrected estimate equal to what we find typically requires less selection on unobservables than observables.

Robustness to violations in distributional assumptions

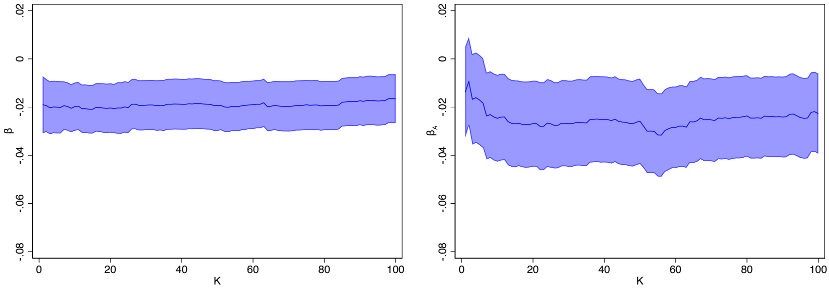
Back



Black line shows β and β_A estimates for a wide range of assumed values of $\mathbb{E}[L^*|L=0,X]$.

Changing the number of clusters

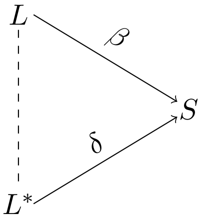
Back



Relation to RDD

Back

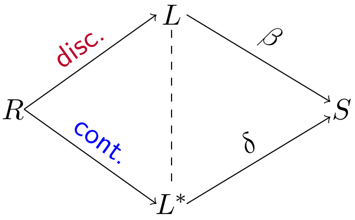
$$S = \alpha + \beta L + \delta L^* + \epsilon$$



Relation to RDD

Back

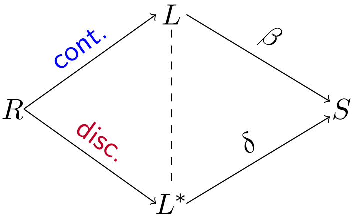
$$S = \alpha + \beta L + \delta L^* + \epsilon$$



Relation to RDD

Back

$$S = \alpha + \beta L + \delta L^* + \epsilon$$

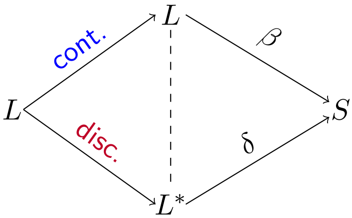


Relation to RDD

Back

$$S = \alpha + \beta L + \delta L^* + \epsilon$$

$$L = \max\{0, L^*\}$$



Relation to RDD

Back

$$S = \alpha + \beta L + \delta L^* + \epsilon$$
$$L = \max\{0, L^*\}$$

$$\begin{tikzcd}
L \arrow[d, "\text{disc.}"] \arrow[dr, "\beta"] & \\
 & S \\
L^* \arrow[ur, "\delta"'] &
\end{tikzcd}$$

Cont. Treatment: $\lim_{l \to 0^+} \mathbb{E}[\beta L | L = l] - \mathbb{E}[\beta L | L = 0] = 0$

Disc. Endogeneity:

$\lim_{l \to 0^+} \mathbb{E}[\delta L^* | L = l] - \mathbb{E}[\delta L^* | L = 0] = -\delta \mathbb{E}[L^* | L = 0]$