# Better Bunching, Nicer Notching

Marinho Bertanha
*University of Notre Dame and University of Chicago*

Andrew H. McCallum
*Board of Governors of the Federal Reserve System*

Nathan Seegert
*University of Utah*

August 28, 2019

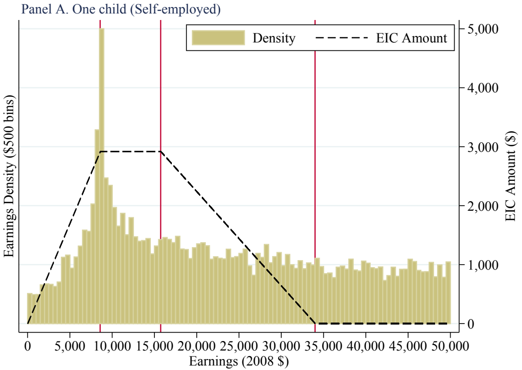
EEA-ESEM, Manchester

Disclaimer: Any opinions and conclusions expressed herein are solely the responsibility of the authors and do not necessarily represent the views of the Board of Governors of the Federal Reserve System or any other person associated with the Federal Reserve System.

# Goal: Estimate the Tax Elasticity of Reported Income.

Saez [2010] and Kleven and Waseem [2013]'s insight is that the elasticity relates to amount of bunchers.

More bunchers = larger elasticity; *ceteris paribus*.



## The Bunching Estimator

- Ideally, you would like to observe the same distribution of individuals facing different tax changes
  - There are not that many quasi experiments.

- The bunching estimator is an appealing identification strategy because it only takes one population of individuals and one (piece-wise linear) budget set to estimate the elasticity.

- Many examples of applications with piece-wise linear budget sets: Prescription drug insurance [Einav et al., 2017], pensions systems [Brown and Laschever, 2012], welfare programs [Camacho and Conover, 2011], education policy [Dee et al., 2011], labor regulations [Garicano et al., 2016], minimum wages [Dube et al., 2017], fuel economy policy [Sallee and Slemrod, 2012], real estate taxes [Kopczuk and Munroe, 2015], and pricing schemes in electricity [Ito, 2014], cellular service [Huang, 2008], and water markets [Olmstead et al., 2007].

- Our focus is on one population facing one budget set.
  - There are other estimators using variation from piece-wise linear budget constraints e.g., Blomquist and Newey [2002].

# What Do We Learn From Bunching?

- This paper clarifies that non-parametric identification of the elasticity is impossible in the case of kinks.

- We propose identifying conditions that are weaker than those used before.

Solution 1 (Bounds)
- non-parametric shape restriction: partial identification

Solution 2 (Truncated Tobit)
- covariates and semi-parametric distribution restriction: point identification

# What Do We Learn From Bunching?

- This paper clarifies that non-parametric identification of the elasticity is impossible in the case of kinks.
- We propose identifying conditions that are weaker than those used before.

weakest assumptions

Solution 1 (Bounds)
- non-parametric shape restriction: partial identification

Solution 2 (Truncated Tobit)
- covariates and semi-parametric distribution restriction: point identification

strongest assumptions

# Utility Maximization Problem

- Agent type $N^*$ maximizes utility $U(C, Y; N^*)$ choosing consumption $C$ and labor income $Y$

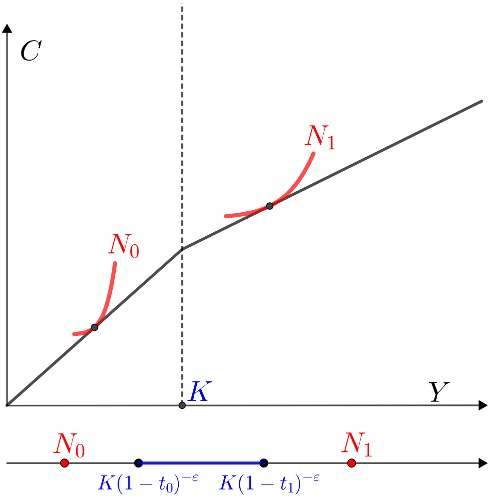
$$\max_{C,Y} \quad C - \frac{N^*}{1+1/\varepsilon}\left(\frac{Y}{N^*}\right)^{1+\frac{1}{\varepsilon}}$$

$$s.t. \quad C = \mathbb{I}\{Y \leq K\}[I_0 + (1-t_0)Y] + \mathbb{I}\{Y > K\}\left[I_1 + (1-t_1)(Y-K)\right]$$

- Piece-wise linear budget with intercept $I_j$ and slope $1 - t_j$

- There is a tax rate change $t_0 < t_1$ at $Y = K$ (kink)

# Solution to Utility Maximization Problem



# Goal: Invert Solution Mapping to Get $\varepsilon$

The solution in logs is:

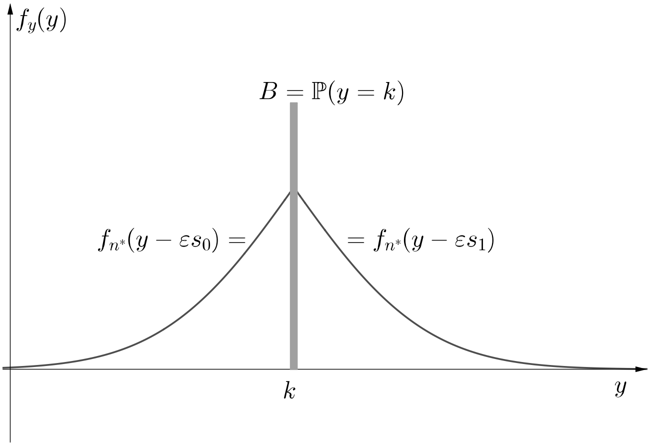
$$y = \begin{cases} n^* + \varepsilon s_0 & \text{, if } 0 < n^* < k - \varepsilon s_0 \\ k & \text{, if } k - \varepsilon s_0 \leq n^* \leq k - \varepsilon s_1 \\ n^* + \varepsilon s_1 & \text{, if } k - \varepsilon s_1 < n^* < \infty \end{cases} \tag{1}$$

where:
$y = \log(Y)$ , $n^* = \log(N^*)$, $k = \log(K)$, $s_j = \log(1 - t_j)$,

- It is helpful to think in terms of the counterfactual scenario of no tax change: $t_0 = t_1$
- Counterfactual income : $y_0 = n^* + \varepsilon s_0$

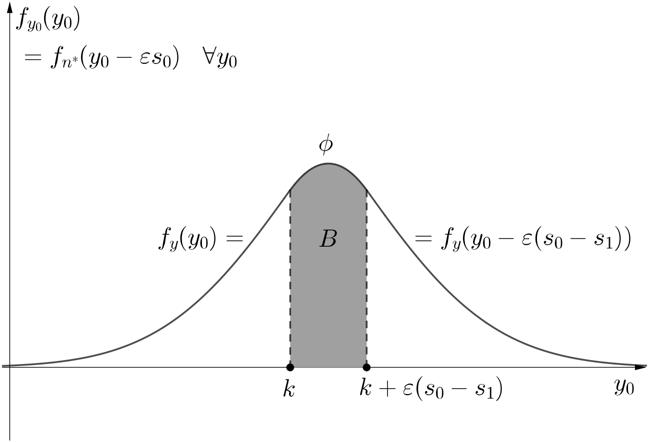
# Distribution of Income

## Observed Data



# Distribution of Income

## Unobserved Counterfactual



# How Does the Literature Identify the Elasticity?

- The 'bunching estimator' of Saez [2010] uses the definition of $B$

$$B = \int_{k}^{k+\varepsilon(s_0 - s_1)} f_{y_0}(y) \; dy,$$

  - makes a trapezoidal approximation,

$$\cong 0.5 \left[ f_{y_0}\left(k + \varepsilon\left(s_0 - s_1\right)\right) + f_{y_0}\left(k\right) \right] \varepsilon\left(s_0 - s_1\right),$$

  - replaces $f_{y_0}$ by $f_y$,

$$= 0.5 \left[ f_y(k^+) + f_y(k^-) \right] \varepsilon\left(s_0 - s_1\right),$$

  - and solves for the elasticity

$$\varepsilon \cong B / \left\{ \left(s_0 - s_1\right) 0.5 \left[ f_y(k^+) + f_y(k^-) \right] \right\}$$

- This restricts the PDF $f_{n^*}$ to be "linear" for $n^* \in [k - \varepsilon s_0, k - \varepsilon s_1]$.

# Literature Relies on Strong Functional Form Assumptions

- "*But isn't linearity approximately true if the bunching interval is small?*"

- There is no way to know because the length of the interval depends on the unknown $\varepsilon$.

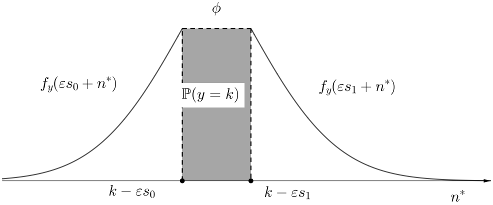
- Chetty et al. [2011] make a stronger functional form assumption:
  - $n^*$ has a flat PDF inside the bunching interval.

- These are not non-parametric identification strategies.

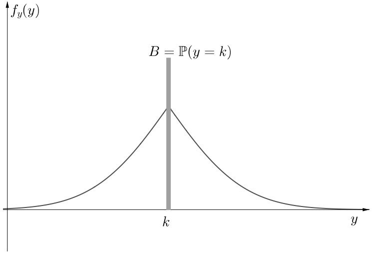
- We find $\varepsilon$ is very sensitive to the shape of $f_{n^*}$.

# Identification without Restrictions on $f_{n^*}$ is Impossible

Unobserved ability distribution $f_{n^*}$ and $\varepsilon$



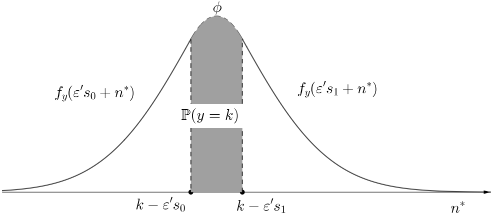

$\Downarrow$ $\Downarrow$ Solution in Equation (1) $\Downarrow$ $\Downarrow$



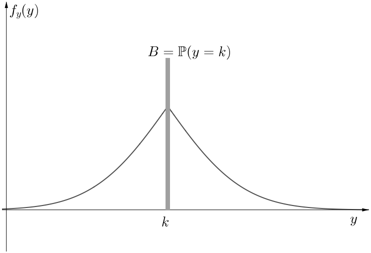

Observed income distribution $f_y$

# Identification without Restrictions on $f_{n^*}$ is Impossible

Unobserved ability distribution $f_{n^*}$ and $\varepsilon'$




⇓ ⇓ Solution in Equation (1) ⇓ ⇓




Observed income distribution $f_y$

# Data Justifies Any Elasticity

**Lemma:** Assume $f_{n^*}$ is an unobserved continuous PDF with support $(-\infty, +\infty)$, and that we observe $f_y$.

Then, for every elasticity $\varepsilon > 0$, there exists a $f_{n^*}$ such that Equation (1) maps the distribution of $n^*$ into the distribution of $y$.

Therefore, it is impossible to identify the elasticity.

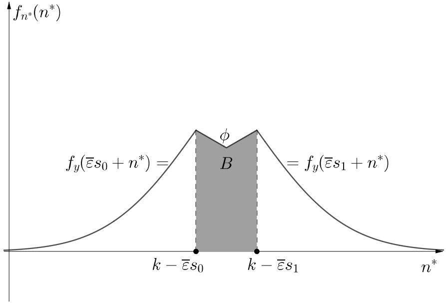
- There is an old literature on impossible inference.
  - For a review see Bertanha and Moreira (2019).
  - Here we have a worse problem: not even identification.

- This result also appears in Blomquist and Newey [2017].

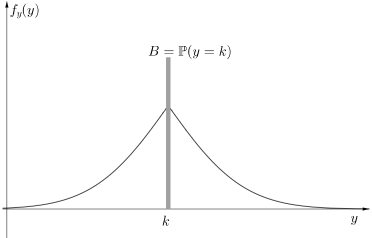
# Shape Restriction Yields Partial Identification

Assume $f_{n^*}$ is Lipschitz with constant $M$

Unobserved ability distribution, $f_{n^*}$



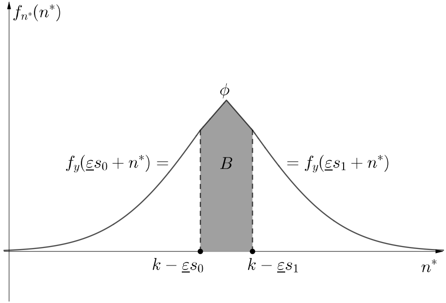

⇓ ⇓ Solution in Equation (1) ⇓ ⇓



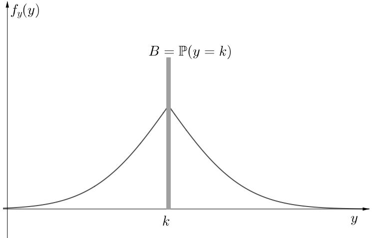
# Shape Restriction Yields Partial Identification

Assume $f_{n^*}$ is Lipschitz with constant $M$

Unobserved ability distribution, $f_{n^*}$




⇓ ⇓ Solution in Equation (1) ⇓ ⇓



# Partial Identification of $\varepsilon$

**Theorem:** Assume $f_{n^*}$ is Lipschitz with known constant $M$.

Then,

$$\varepsilon \in \Upsilon = \begin{cases} \emptyset & \text{, if } B < \frac{\left|f_y(k^+) - f_y(k^-)\right| \left[f_y(k^+) + f_y(k^-)\right]}{2M} \\ [\underline{\varepsilon}, \overline{\varepsilon}] & \text{, if } \frac{\left|f_y(k^+) - f_y(k^-)\right| \left[f_y(k^+) + f_y(k^-)\right]}{2M} \leq B < \frac{f_y(k^+)^2 + f_y(k^-)^2}{2M} \\ [\underline{\varepsilon}, \infty) & \text{, if } \frac{f_y(k^+)^2 + f_y(k^-)^2}{2M} \leq B \end{cases}$$

where

$$\underline{\varepsilon} = \frac{2\left[f_y(k^+)^2/2 + f_y(k^-)^2/2 + M\ B\right]^{1/2} - \left(f_y(k^+) + f_y(k^-)\right)}{M(s_0 - s_1)}$$

$$\overline{\varepsilon} = \frac{-2\left[f_y(k^+)^2/2 + f_y(k^-)^2/2 - M\ B\right]^{1/2} + \left(f_y(k^+) + f_y(k^-)\right)}{M(s_0 - s_1)}$$

**Solution 1 (Bounds) :** We recommend a sensitivity analysis for various choices of $M$.

# Bunching is a Censored Regression Model

- **Classic Tobit**

  Suppose $n^* \sim N(\mu, \sigma^2)$.

  Left censored at $k$: $y = \max\{k; n^*\}$, or

  right censored at $k$: $y = \min\{k; n^*\}$.

- **Mid-censored Model**: Equation (1) rewrites as

  $$y = \min\{\varepsilon s_0 + n^* \ , \ \max\{k \ , \ \varepsilon s_1 + n^*\}\}.$$

  - The elasticity is the difference of intercepts divided by $(s_1 - s_0)$

# Use Covariates to Predict Ability with a Tobit

- **Mid-censored Tobit**
  -There is a vector of covariates $X$

- **Idea:** instead of polynomial or linear extrapolations, use the relationship between $n^*$ and $X$ for non-bunching people to predict the distribution of $n^*$ for bunching people.

Assumption: $n^* \mid X \sim N(X\beta; \sigma^2)$

- **Question:** what if conditional normality is incorrect?

  - MLE is inconsistent ...

# Tobit is Robust to Non-normality

Assumption: there exists an unique $(\beta, \sigma)$ such that $F_{n^*}(n) = \mathbb{E}\left[\Phi\left(\frac{n - X\beta}{\sigma}\right)\right]$, plus some regularity conditions.

- Semi-parametric class that does NOT require conditional normality.

**Lemma:** Let $\theta^* = (\beta^*, \sigma^*, \varepsilon^*)$ be the probability limit of the Tobit MLE, and suppose the Assumption above holds.

If the true distribution $F_y(y)$ matches the Tobit fitted distribution $G_y(y; \theta^*)$, then $\varepsilon^*$ is equal to the true elasticity.

**Solution 2 (Truncated Tobit)**:

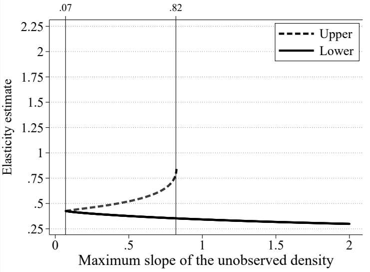
- truncate the sample of $y$ to a neighborhood of $k$;
- vary the size of the neighborhood, from full to smallest sample;
- examine how the fit and estimates vary.

# Application

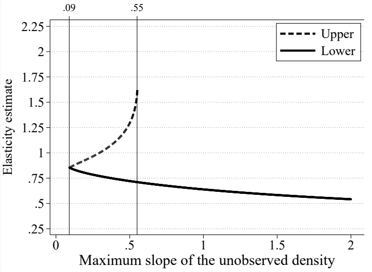
- Employ data originally studied by Saez [2010]
  - – Individual Public Use Tax Files constructed by the IRS
  - – Annual cross-sections 1995-2004
- Focus on $8,580 kink in the EITC schedule:
  - – tax rate changes from -34 % to 0%
- We constrast trapezoid-based estimates with
  - \- Solutions 1 (Bounds)
  - \- Solutions 2 (Truncated Tobit)
    - – dummy covariates: year effects, types of deductions, received social security benefits, used a tax prep software, etc.

# Partial Identification

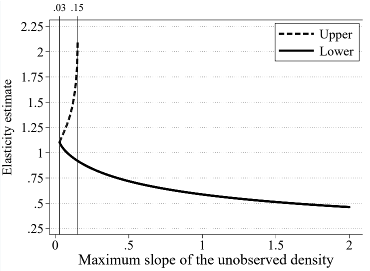
(a) All Filers



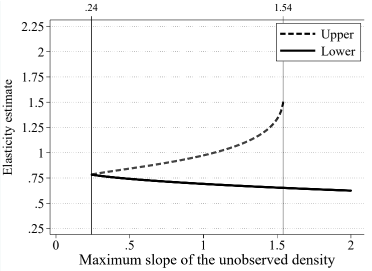

(b) Self-Employed Filers



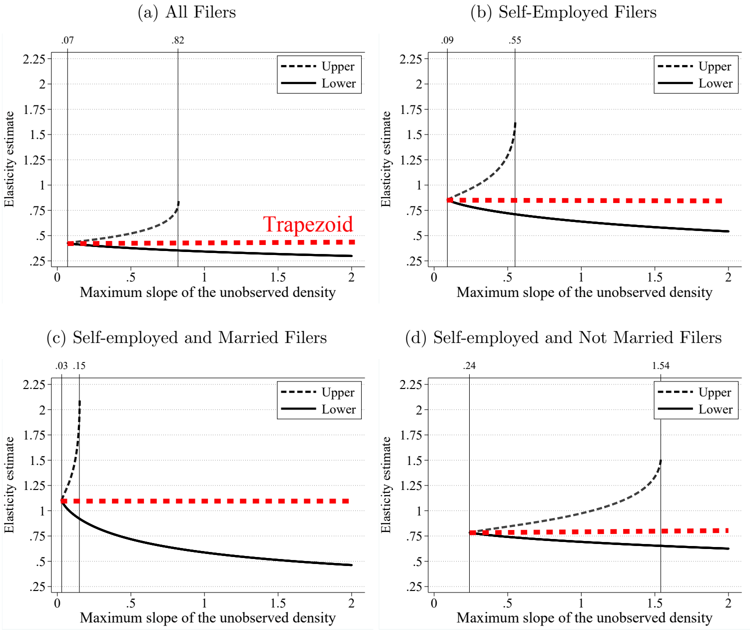

(c) Self-employed and Married Filers



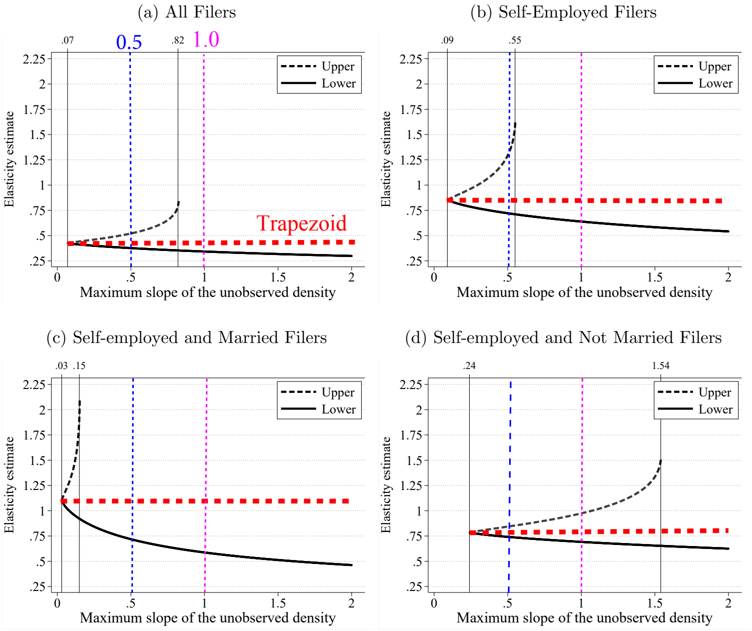

(d) Self-employed and Not Married Filers



# Partial Identification

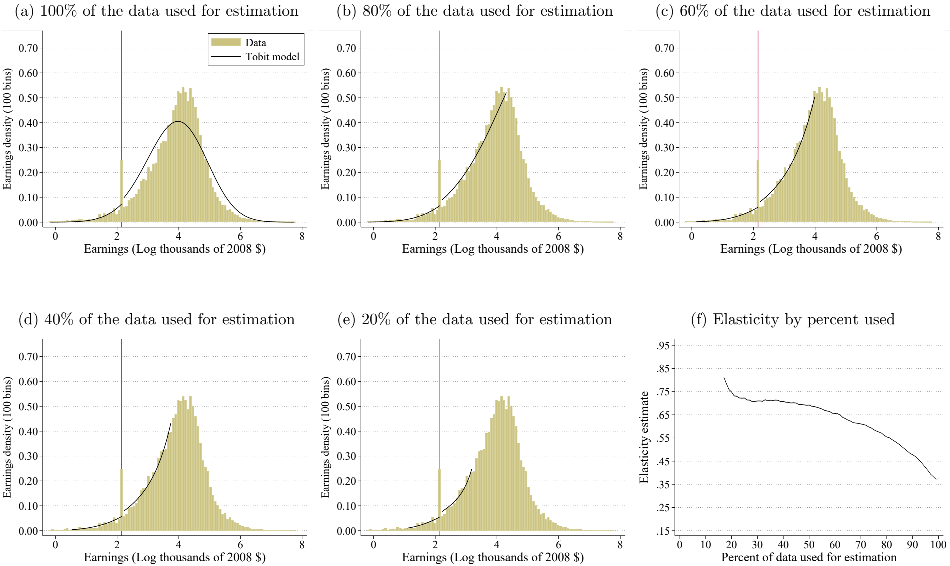


# Partial Identification



# Truncated Tobit

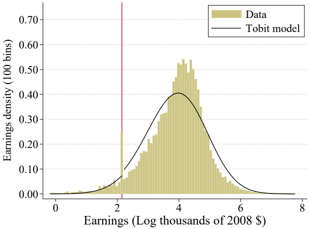
## Self-employed Married



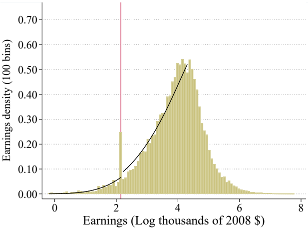
# Truncated Tobit

## Self-employed Married

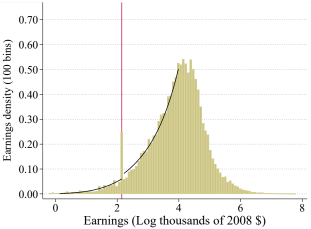
(a) 100% of the data used for estimation



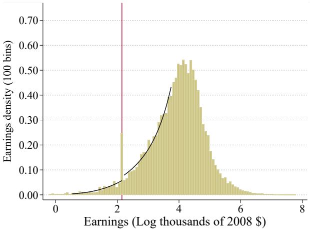

(b) 80% of the data used for estimation



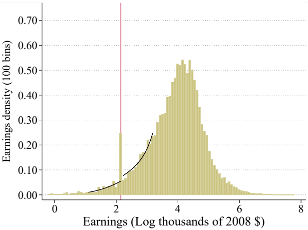

(c) 60% of the data used for estimation




(d) 40% of the data used for estimation



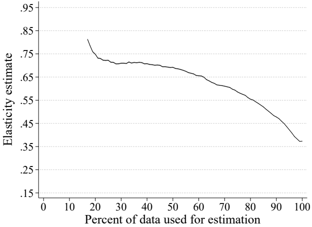

(e) 20% of the data used for estimation




Trapezoid

1.102

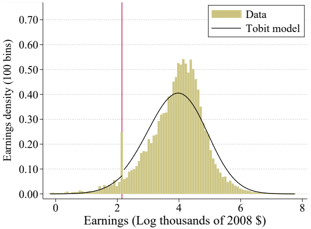
(f) Elasticity by percent used



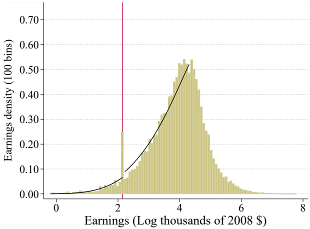
# Truncated Tobit

## Self-employed Married

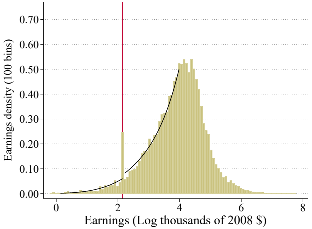
(a) 100% of the data used for estimation



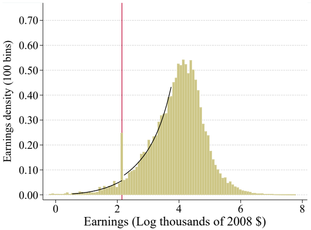

(b) 80% of the data used for estimation



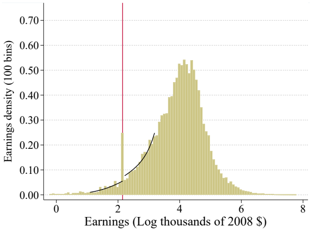

(c) 60% of the data used for estimation

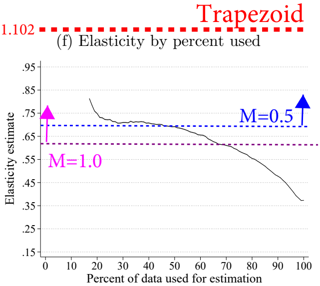



(d) 40% of the data used for estimation




(e) 20% of the data used for estimation




(f) Elasticity by percent used

# Conclusions

- Minimal restrictions partially identify the elasticity
- Connection between bunching and censored regressions allows for:
  - Covariates: more meaningful extrapolation
  - Semi-parametric restrictions on the distribution of $n^*$

**Future of Bunching:**

- examine sensitivity of estimates using Bounds and Truncated Tobit
- STATA package coming soon!

**More on the Paper, Skipped Today:**

- failure of the widely used "polynomial strategy"
- multiple kinks and notches
- non-parametric identification is possible in the case of notches
- point indentification with censored quantile regressions