# Estimating a Model of Assortative Matching with Large Firms

With German and Spanish data

# Part I: The Model & Estimation Strategy

# Model

- The economy consists of workers with different skill $x$, and firms with different managerial skill $y$. Both are characterised by cumulative density functions $H_w(x)$, $H_f(y)$.
- They both match and produce output according to a production technology
$$F(x, y, l_x, r_x)$$
- Which depends on their skills, the $l_x$ number of workers of type $x$ the firm type $y$ employs, and the managerial resources $r_x$ (time the manager can spend supervising her workers of skill $x$). $\int r_x \, dx = 1$ (resources of the firm are normalized to one)
- Total output of the firm is the sum of output across worker types:
$$\int F(x, y, l_x, r_x) \; dx$$

# Model

- The optimization problem of the firm is therefore

$$\max_{l_x, r_x} \int [F(x, y, l_x, r_x) - w(x)] \; dx$$

- Dividing by $r_x$ and denoting $\theta = l_x / r_x$, the problem becomes

$$\max_{x,\theta} f(x, y, \theta) - \theta w(x)$$

- Where $f(x, y, \theta) = F(x, y, l_x / r_x, 1)$ is the production function in intensive form.

This means that optimally firms only hire one type of worker, and they have to decide which type and how many to hire.

# Assortative Matching

- An equilibrium is therefore a feasible allocation $R(x, y, \theta)$ and a strictly positive wage $w(x)$ that solves the firms problem.
- We focus on assortative matching, that is, monotonic allocations that are monotonic in $x, y$. That is:
  - Positive Assortative Matching (PAM): High $x$ matches with high $y$.
  - Negative Assortative Matching (NAM): High $x$ matches with low $y$.
- Here is when the complementarities of the different inputs become important

# Input Complementarities

Denote $F_{ij}$ the cross partial derivate w.r.t inputs I and j:

- $F_{xy}$ is *type complementarity* – good firms do better with good workers.
- $F_{lr}$ is *quantities complementarity* – always positive with constant returns to scale.
- $F_{yl}$ is *span of control complementarity* - good firms do better with more workers.
- $F_{xr}$ is *managerial resource complementarity* – if positive (and large) more time spent with good workers is more productive than time spent with *bad* workers

# Input Complementarities

Why are the cross-derivatives important?

- A necessary condition for PAM is:

$$F_{xy}F_{lr} \geq F_{yl}F_{xr}$$

- The opposite inequality is necessary and sufficient condition for NAM.
- The interpretation of the input complementarities relates to questions of skill-bias technological change vs quantity-bias technological change (or increases in $F_{xy}$ vs increases in $F_{yl}$).

# Solving the Model

The solution to the model is given by solving a system of two differential equations:

- Under PAM: $\theta'(x) = \frac{H(x)F_{yl} - F_{xr}}{F_{lr}}$; $\mu'(x) = \frac{H(x)}{\theta(x)}$;
- Under NAM: $\theta'(x) = \frac{H(x)F_{yl} + F_{xr}}{F_{lr}}$; $\mu'(x) = -\frac{H(x)}{\theta(x)}$;

Where $\mu(x)$ is the map $y^*(x)$.

- This system can be solved using numerical methods.

# Choice of Function

A very nice special case arises when the production function F is multiplicative separable. In particular, the function we use is:

$$F(x, y, l, r) = A(x, y) * B(l, r)$$

Where:

$$A(x, y) = \left(\omega_A * x^{\frac{(\sigma-1)}{\sigma}} + (1 - \omega_A) * y^{\frac{(\sigma-1)}{\sigma}}\right)^{\frac{\sigma}{(\sigma-1)}}$$

$$B(l, r) = l^{\omega_B} * r^{(1-\omega_B)}$$

- Our goal is to estimate $\omega_B$, $\omega_B$ and $\sigma$ using real world data.

# Targets for estimation

- That means that we have 3 unknown parameters. What can we target to get them?
- The solver takes as inputs the distributions of $H_w(x)$ and $H_f(y)$.
- The solver delivers matched vectors of $(x, \mu(x), \theta(x, \mu(x), w(x))$, which in turn can be used to calculate moments of the distributions of $\theta$ and $w$.

# Estimation Strategy

- The idea is to estimate a distribution of worker skill and firm skill from the data, and get a distribution of firm size, wages and profits.
- The solver takes as inputs $\widehat{H}_w(x)$ and $\widehat{H}_f(y)$, calculates the moments and the distributions of $\hat{\theta}$ and $\widehat{w}(x)$, and then calculates the distance between the actual firm size and wage distribution and the one implied by the model.
- An optimization routine does this over and over until it finds the parameter combo that produces the result that best fits the data.

# Estimation Strategy

Some pending questions:

- Should it be best to target distribution moments of firm size and wages, or distance between distributions? $\rightarrow$ if using distance, we would need the distribution of profits as well: 3 parameters to estimate need 3 equations to solve.
- What is the best proxy for firm and worker skill? $\rightarrow$ This depends on the data…

( GERMAN DATA STARTS HERE)

(Chop this slide off)

# Part III: Spain

# Data Source

- Data comes from the *Muestra Contínua de Vidas Laborales* elaborated by the Spanish Social Security.
- It is an administrative dataset that comprises a panel of a representative sample of the Spanish workforce.
- It also includes matched records form Income Tax declarations of workers – so we can observe their wages and profits.
- Currently it covers the years 2005-2013.

# Data Source

- This is not an employer-employee matched data set by construction.
- But the unique identifiers for firms and workers allow us to build it that way – and add information about firm size, age, location, sector.
- This means that, unfortunately, we have to drop unmatched workers and firms.
- Good news is that the data set is already very big, so even when dropping unmatched data, it is still a big sample.

# Descriptive statistics – by worker

| | Mean | Standard Deviation | Min | Max | Median | 90th percentile | Obs |
|---|---|---|---|---|---|---|---|
| Education | 2.166434 | 1.017146 | 0 | 6 | 2 | 3 | 552,364 |
| Av wage | 15,519.59 | 16,691.64 | .01 | 1,733,613 | 12,988.59 | 30,602.05 | 374,717 |
| Av wage (daily) | 58.06863 | 249.641 | 0.00003 | 68,373.84 | 43.39396 | 91.57422 | 319,919 |
| Av profit | 7,187.25 | 27,099.97 | .01 | 3,005,061 | 1,829.724 | 17,724.17 | 27,225 |

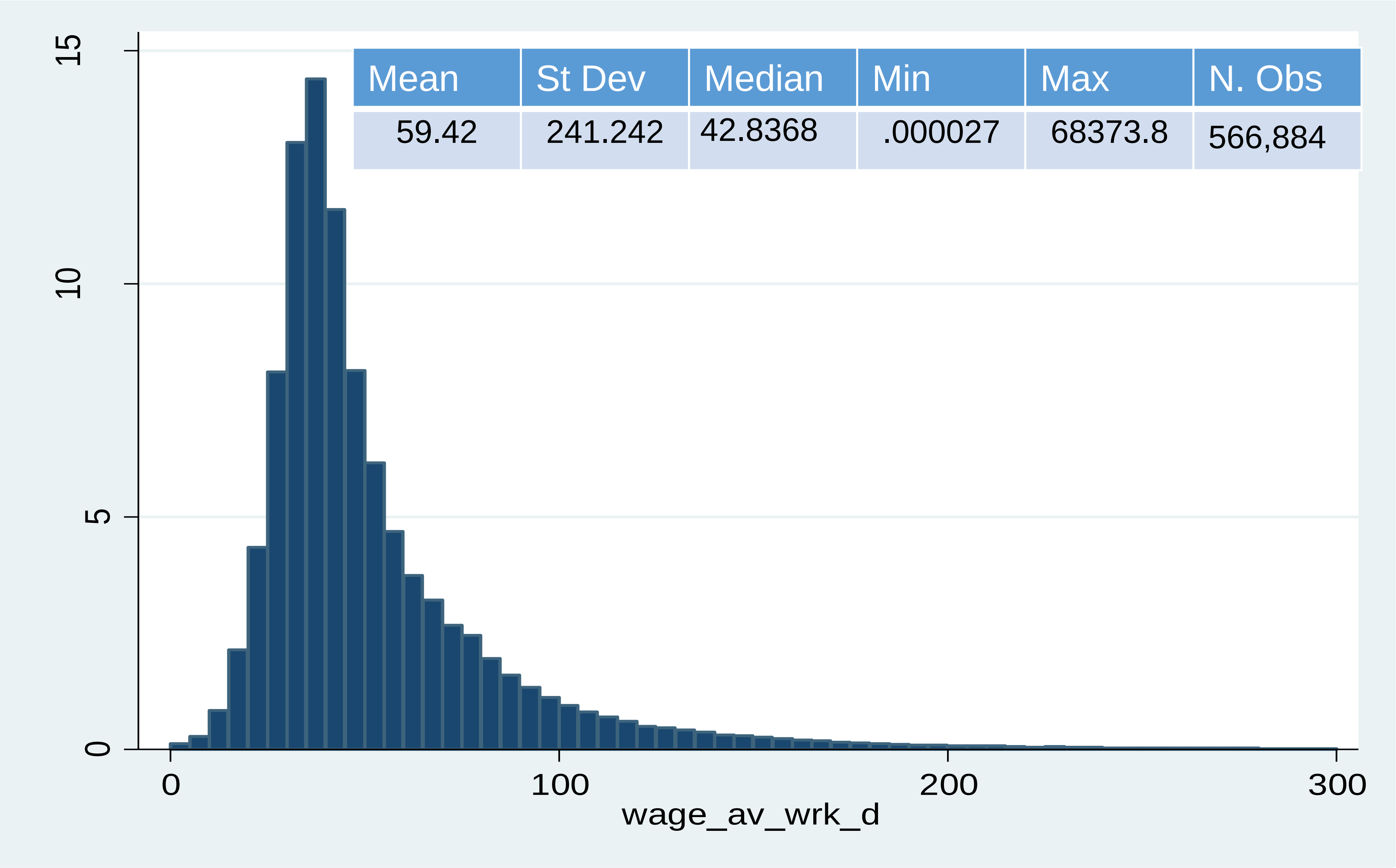
# Descriptive statistics – by firm

| | Mean | Standard Deviation | Min | Max | Median | 90th percentile | Obs |
|---|---|---|---|---|---|---|---|
| Number of workers | 19.10815 | 255.9593 | 0 | 96,402 | 5 | 37 | 376,891 |
| Firm age | 9.262611 | 9.61008 | 0 | 105.737 | 6.00274 | 21.51233 | 353,072 |
| Av profit | 3,601.598 | 16,809.16 | .01 | 3,005,061 | 761.25 | 8,235.29 | 64,817 |
| Av wage | 8,865.947 | 10,613.06 | .01 | 733,636.4 | 6,552.8 | 18,487.82 | 255,872 |
| Av wage (daily) | 41.5741 | 92.41269 | .0001111 | 26,461.27 | 35.97185 | 63.53 | 250,344 |
| Av worker education | 1.987123 | .87669 | 0 | 6 | 2 | 3 | 342,023 |
| Number of firms with some income information | | | | 403,372 | | | |

# Choice of variables

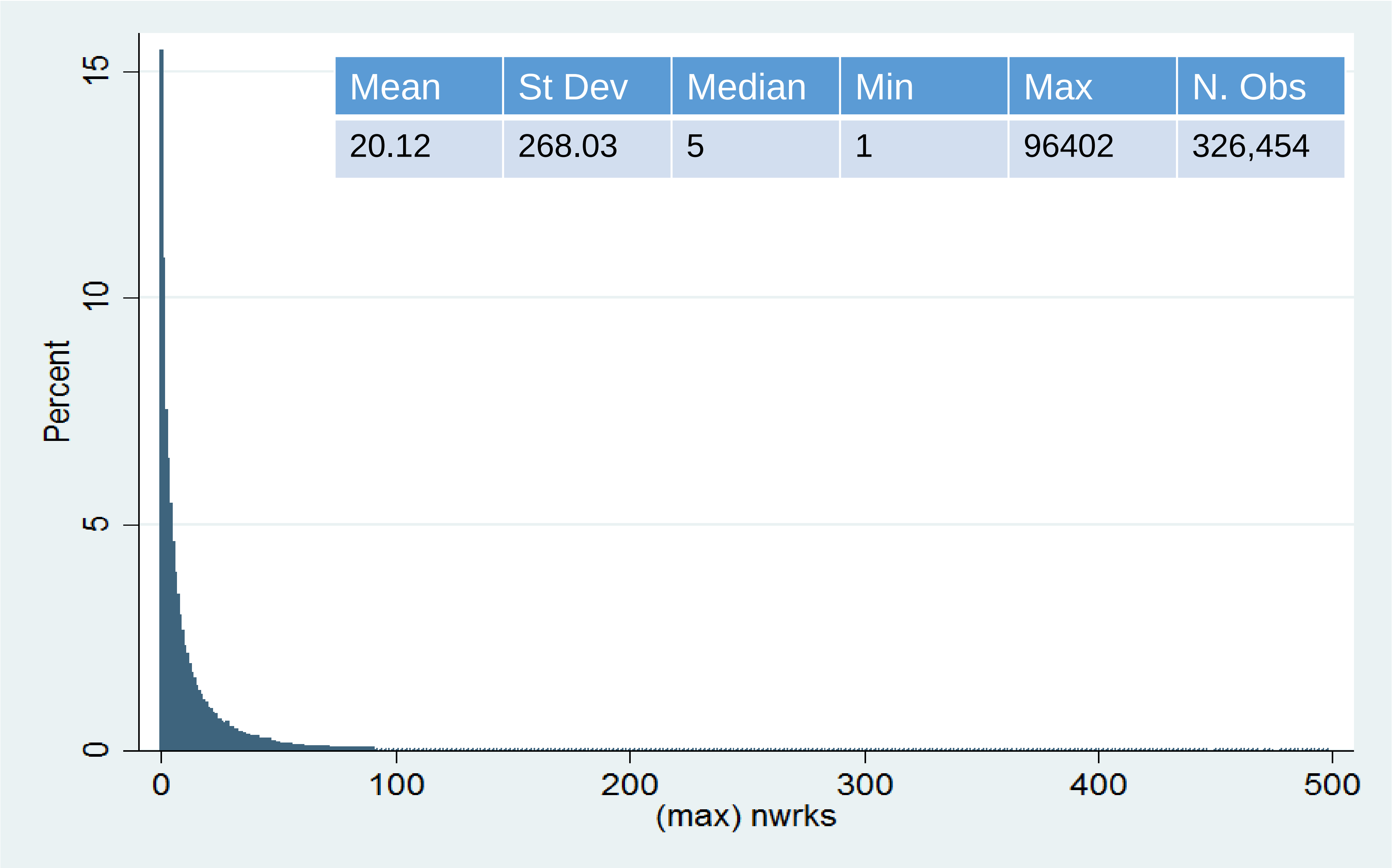
| Model variable | Data variable | Alternatives |
|---|---|---|
| Worker Skill | Education level (0 to 6)<br>(illiterate to PhD) | Log Wages + experience<br>Wage regression residual |
| Manager Skill | Log Average Reported Profits<br>(assumes same distribution as profits) | Profit regression residual<br>Profit per worker |
| Firm Size | Firm size (in workers) | |
| Wages (return to worker skill) | Average Daily Wages | |
| Profits (return to manager skill) | Average Reported Profits | Average Reported Profits per worker |

# Average Daily Wage distribution by firm



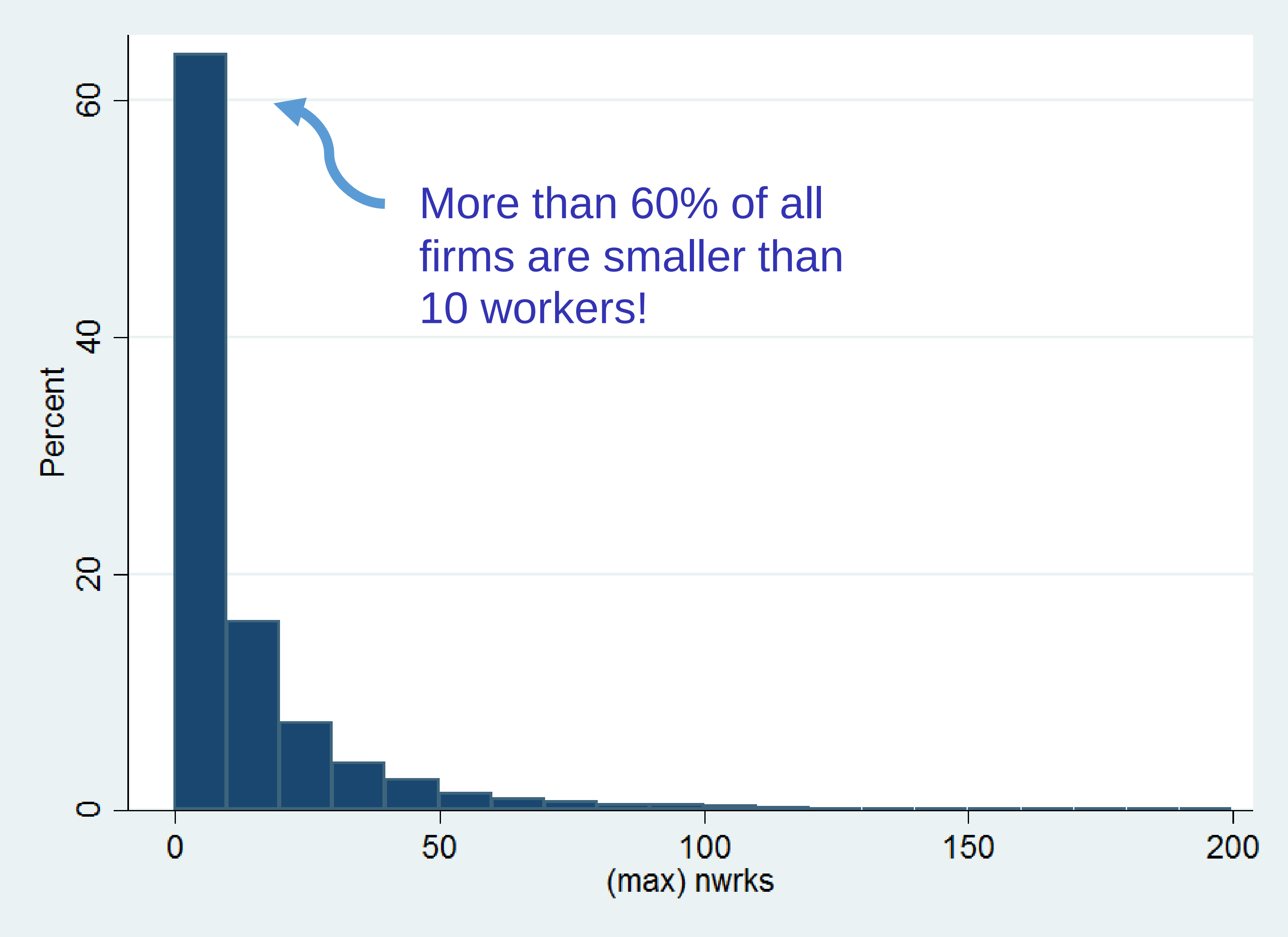

| Mean | St Dev | Median | Min | Max | N. Obs |
|---|---|---|---|---|---|
| 59.42 | 241.242 | 42.8368 | .000027 | 68373.8 | 566,884 |

# Firm Size Distribution

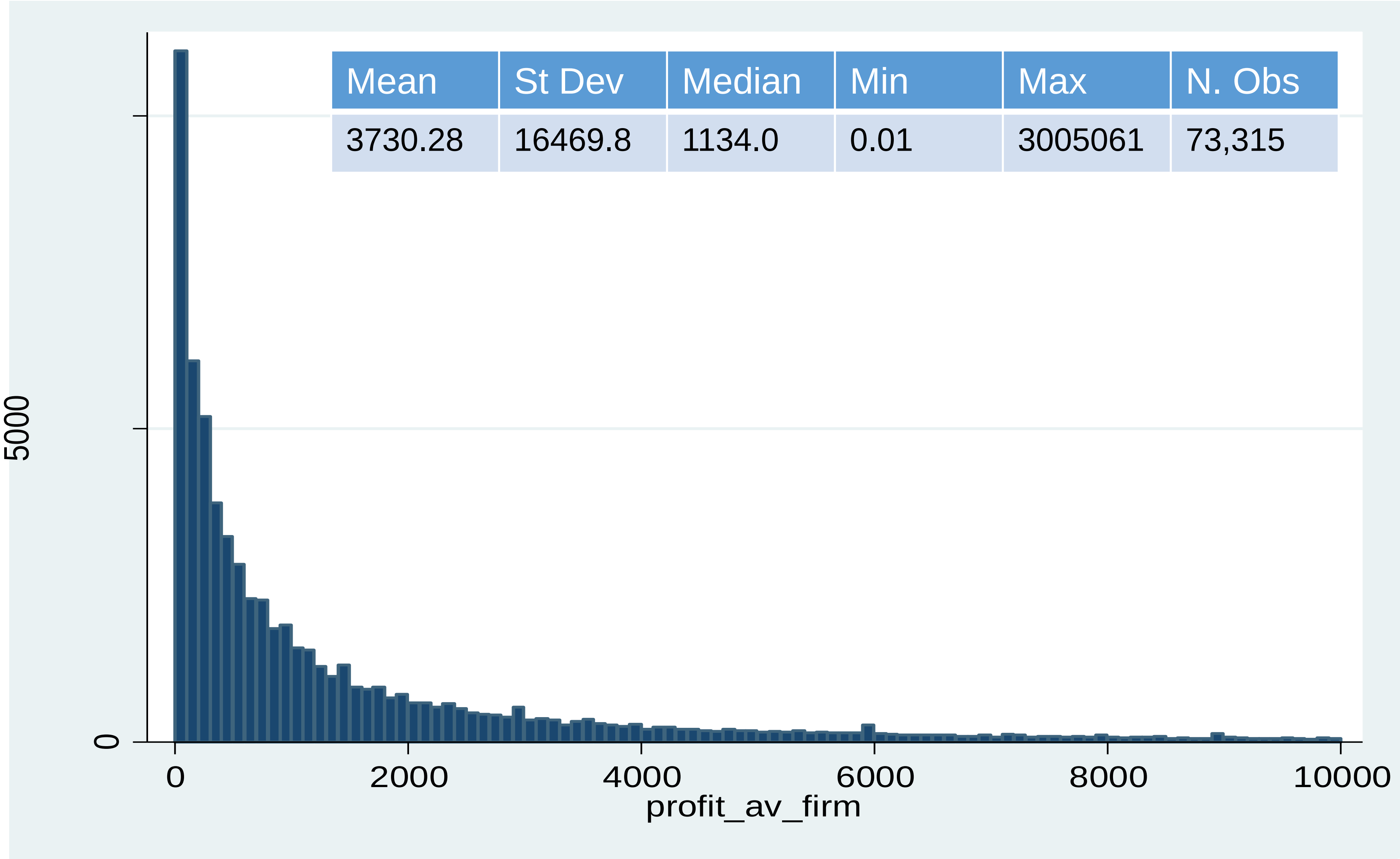



| Mean | St Dev | Median | Min | Max | N. Obs |
|---|---|---|---|---|---|
| 20.12 | 268.03 | 5 | 1 | 96402 | 326,454 |

# Firm Size Distribution (all)



# Profit distribution



# Matched data

- For the estimation, we can only take in observations that have everything:

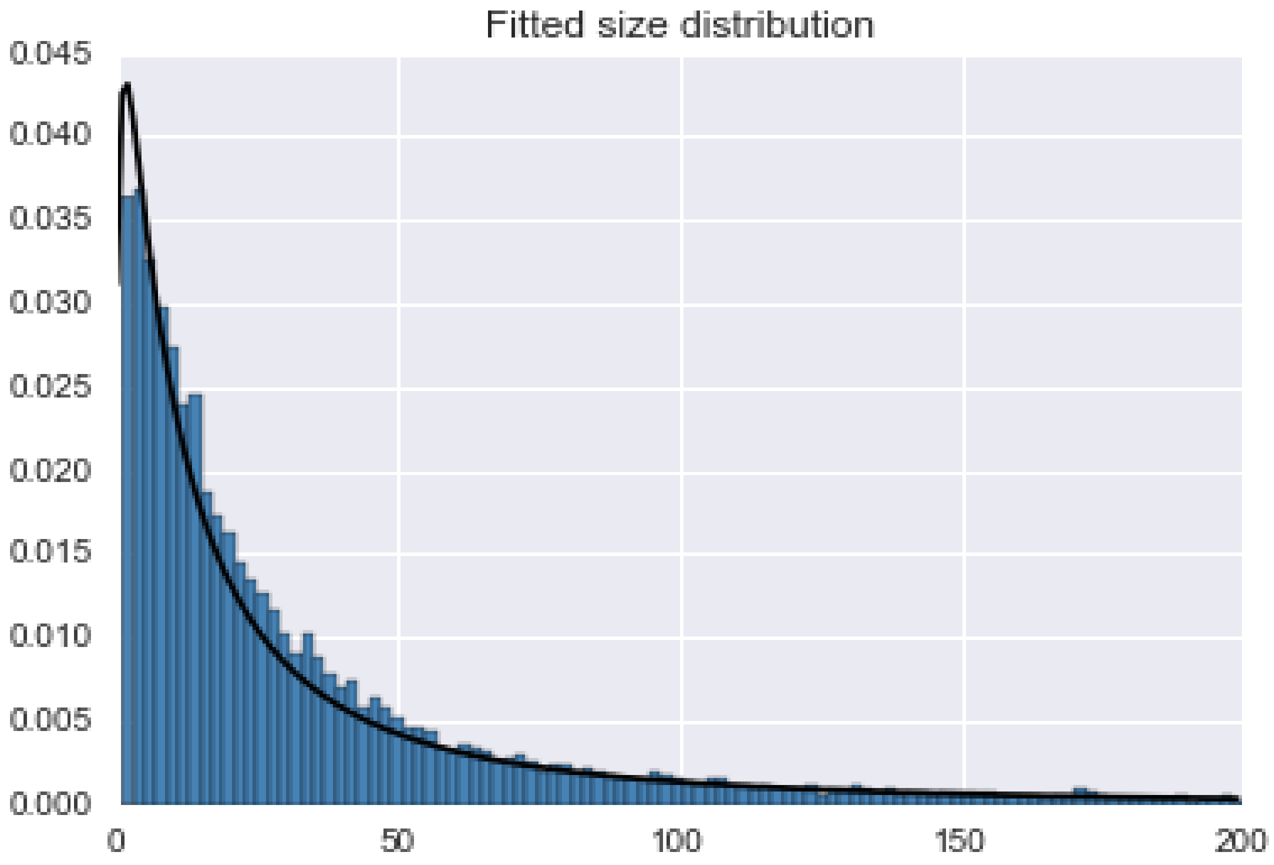
Firm size – av. Wage – av.profit – av education

- This leaves us with around 20,000 observations:

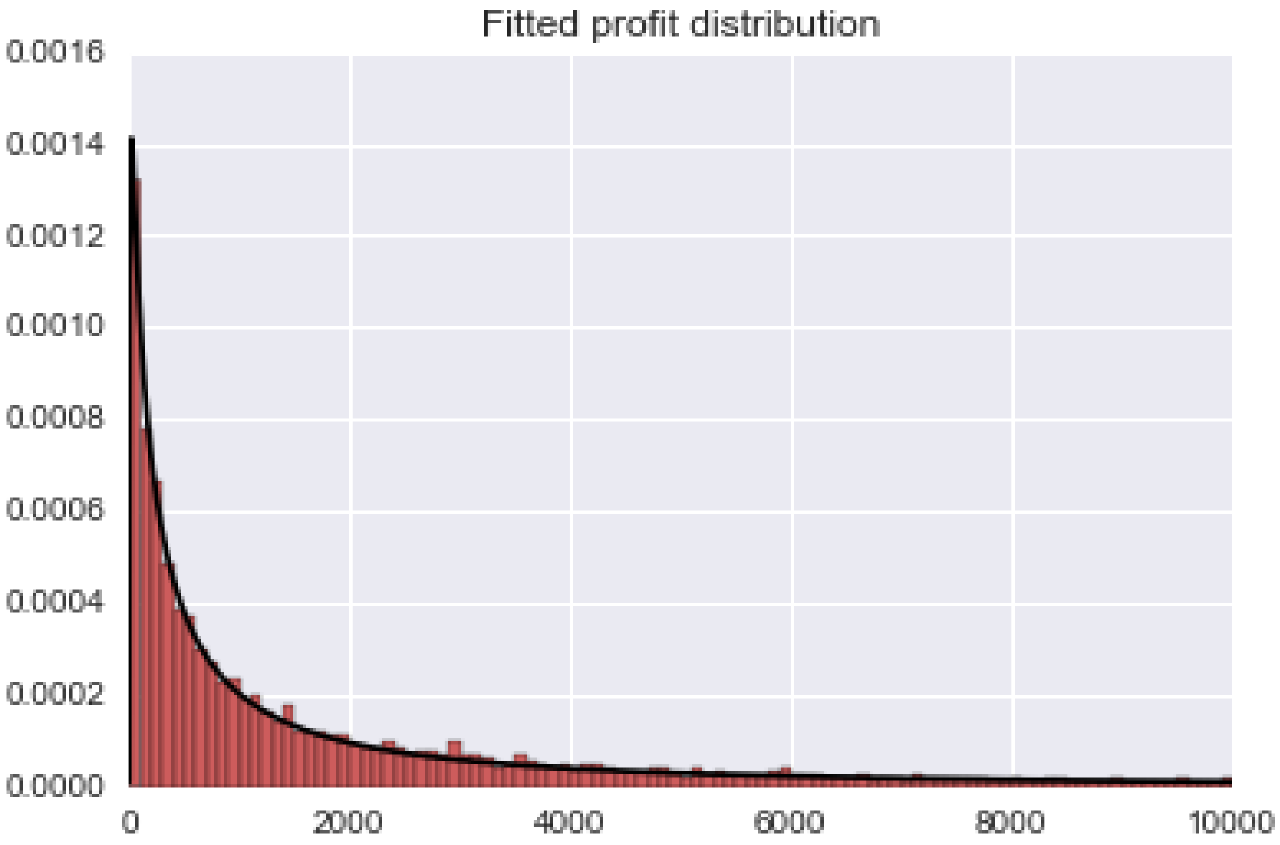
```
. reg profit_av_firm nwrks educ_av_firm wage_av_firm if firmID[_n]!=firmID[_n+1

      Source |       SS       df       MS              Number of obs =   21925
-------------+------------------------------           F(  3, 21921) =   40.79
       Model |  3.6716e+10     3  1.2239e+10           Prob > F      =  0.0000
    Residual |  6.5770e+12 21921   300034139           R-squared     =  0.0056
-------------+------------------------------           Adj R-squared =  0.0054
       Total |  6.6138e+12 21924   301667757           Root MSE      =   17321

------------------------------------------------------------------------------
profit_av_~m |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       nwrks |   .7055322   .3788993     1.86   0.063    -.0371378    1.448202
educ_av_firm |   355.0792   136.6959     2.60   0.009     87.14528     623.013
wage_av_firm |   .0856381   .0093068     9.20   0.000     .0673961    .1038801
       _cons |   2976.042   329.3207     9.04   0.000     2330.549    3621.534
------------------------------------------------------------------------------
```

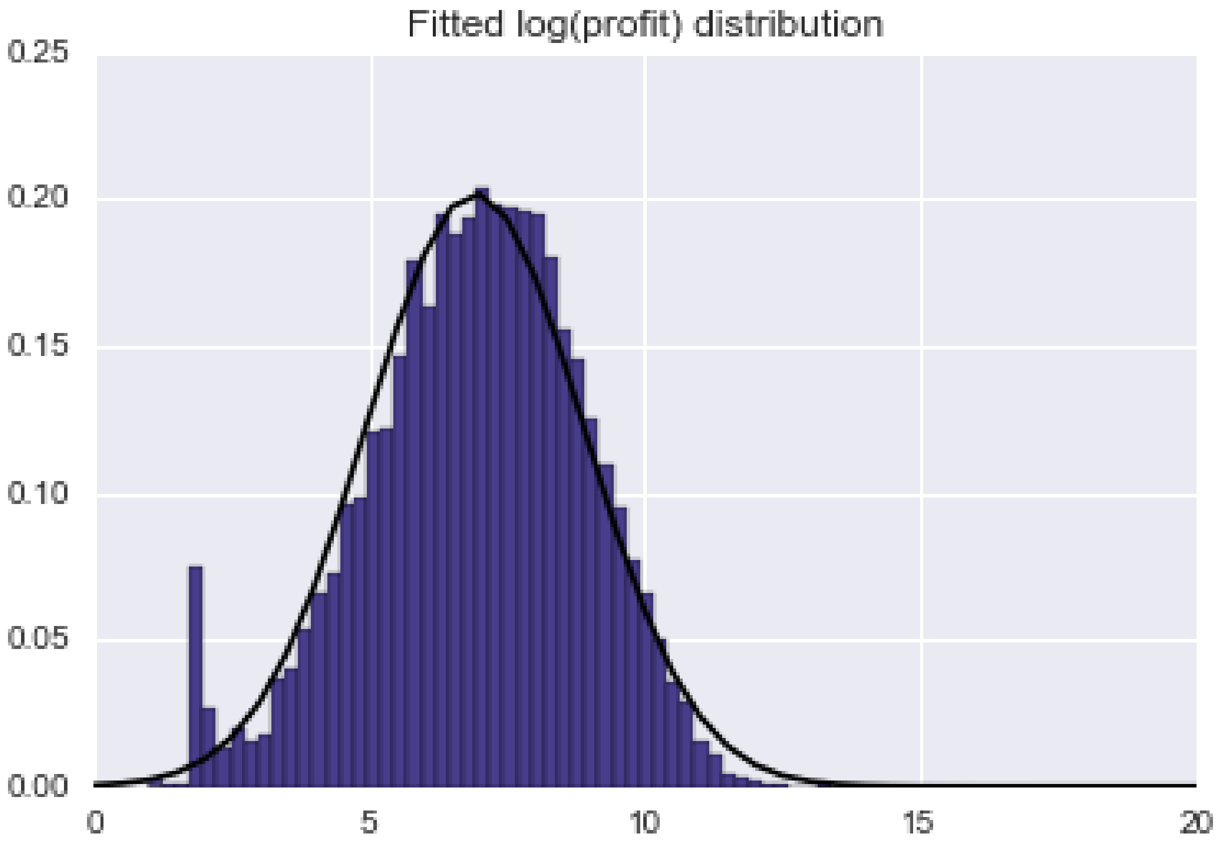
# Fitted distributions: Average Wage by firm

# Fitted distributions: Firm Size

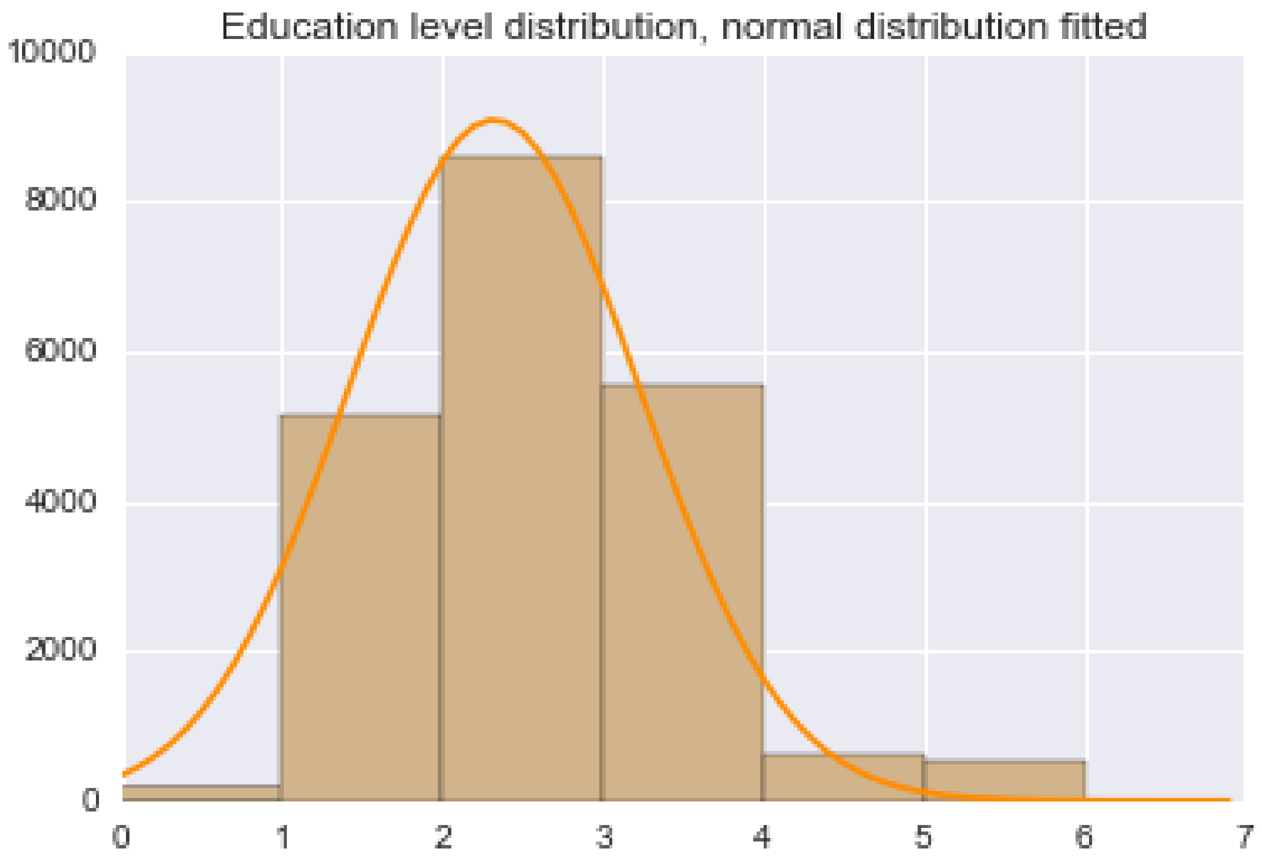


# Fitted distributions: Profits

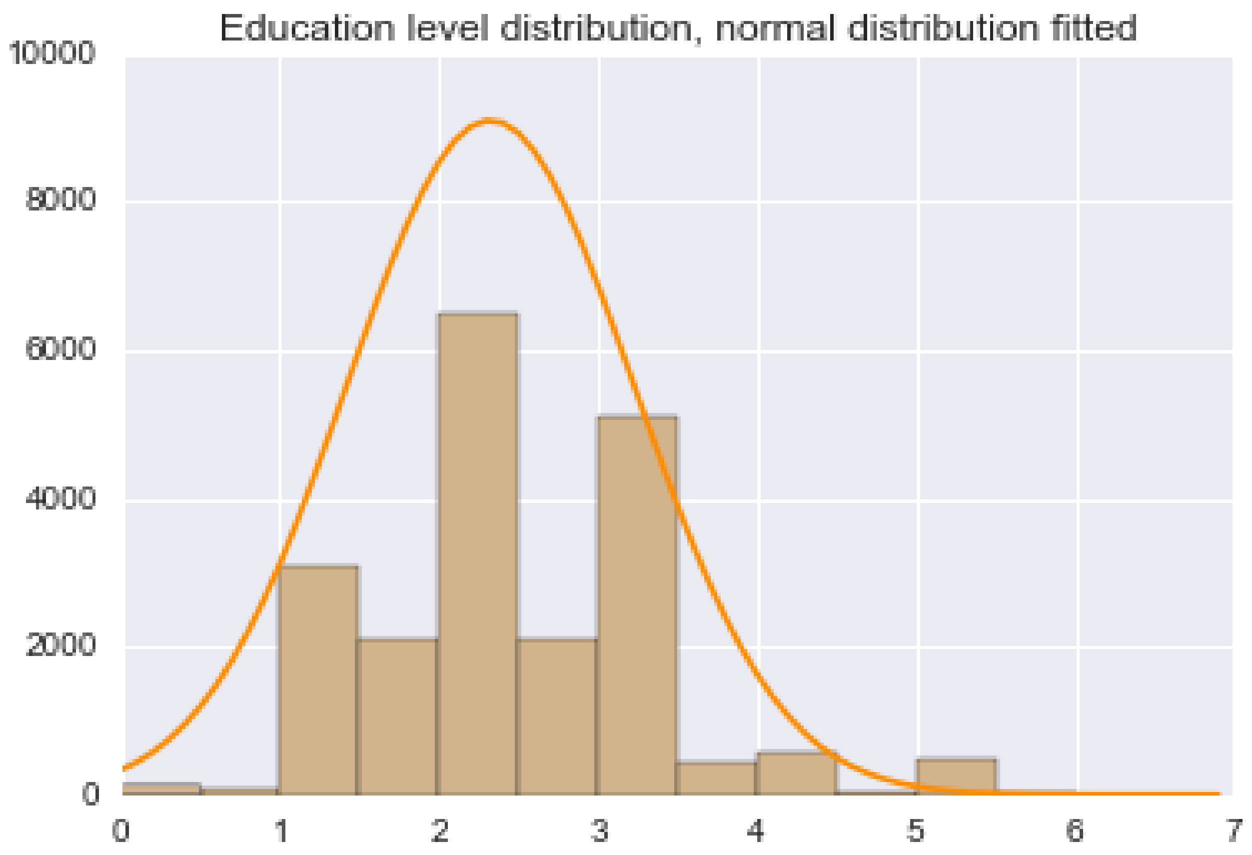


# Fitted distributions: Average Worker Education by firm

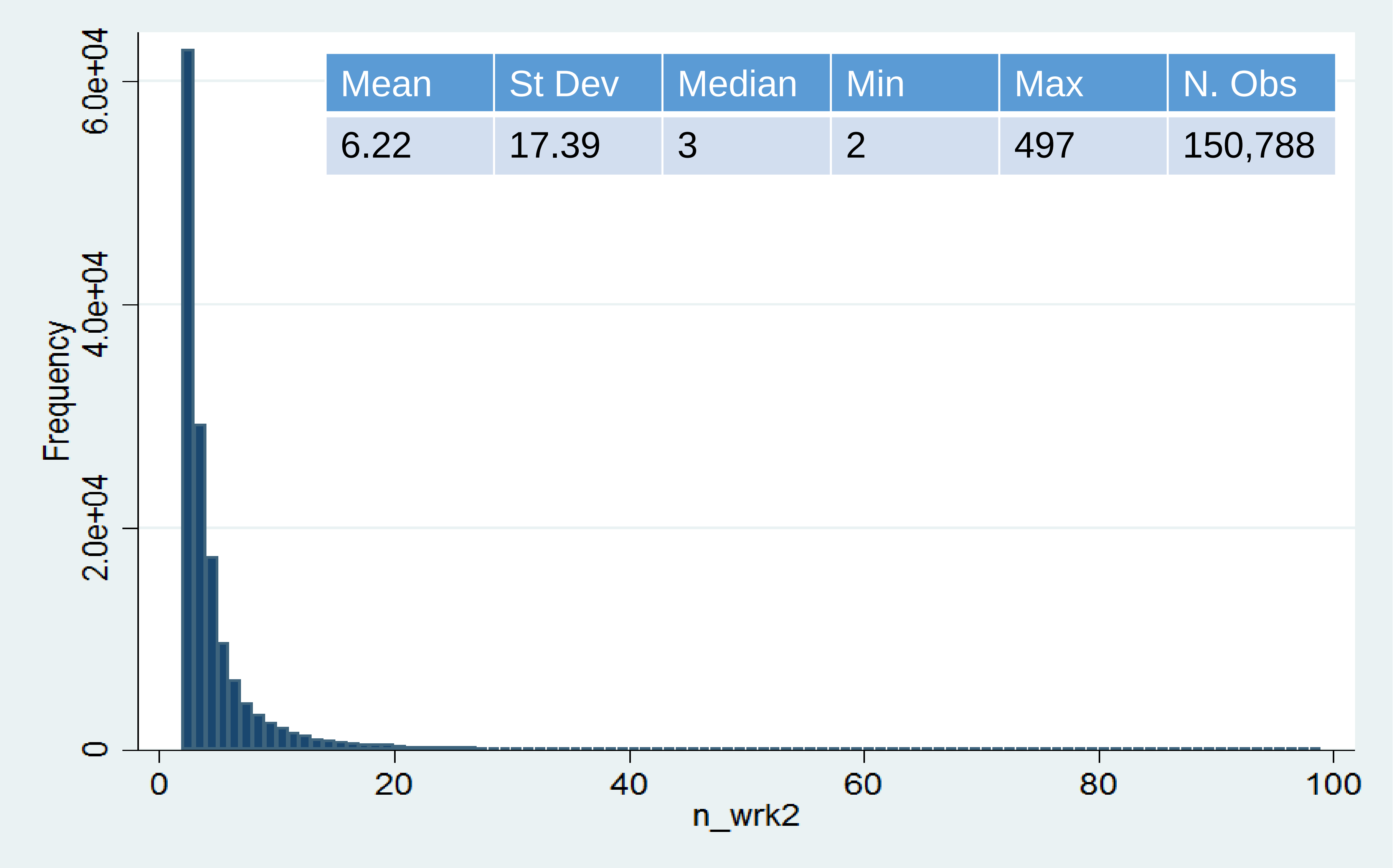


# Fitted distributions: Average Worker Education by firm



# Appendix: More Graphs

# Number of observations per firm (>1)




| Mean | St Dev | Median | Min | Max | N. Obs |
|---|---|---|---|---|---|
| 6.22 | 17.39 | 3 | 2 | 497 | 150,788 |

# Education Coding

| | |
|---|---|
| 0 | Illiterate |
| 1 | Primary Education Completed |
| 2 | Secondary Education Completed |
| 3 | Pre-university Education (Bachillerato and equivalent) Completed |
| 4 | Short University Diploma (Diplomatura and Technical School equivalent) |
| 5 | Graduate (Licenciado) |
| 6 | Postgraduate |