SAMPLE YEAR: 2005

SUMMARY STATISTICS

By worker

By worker status					
	All		With income information		Share of total with fiscal info
	Observations	Share of total	Observations	Share of total	
A	127,770	11.28%	74,504	7.70%	58.31%
N	369,511	32.62%	319,667	33.05%	86.51%
P	316,986	27.98%	288,934	29.88%	91.15%
T	242,987	21.45%	210,899	21.81%	86.79%
U	75,537	6.67%	73,093	7.56%	96.76%
Total	1,132,791	100%	967,097	100%	85.37%

Legend: A – Self-employed; N – Out of the labour force (pensioner); P – Employed with a Permanent contract; T – Employed with a Temporary contract; U – unemployed (registered)

	mean	sd	min	max	median	90th percentile	N
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

Notes: *education* is coded in levels from 0 (illiterate) to 6 (postgraduate). See appendix for a detailed explanation. Wages and profits in euros, per year

By firm

	mean	sd	min	max	median	90th percentile	N
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			

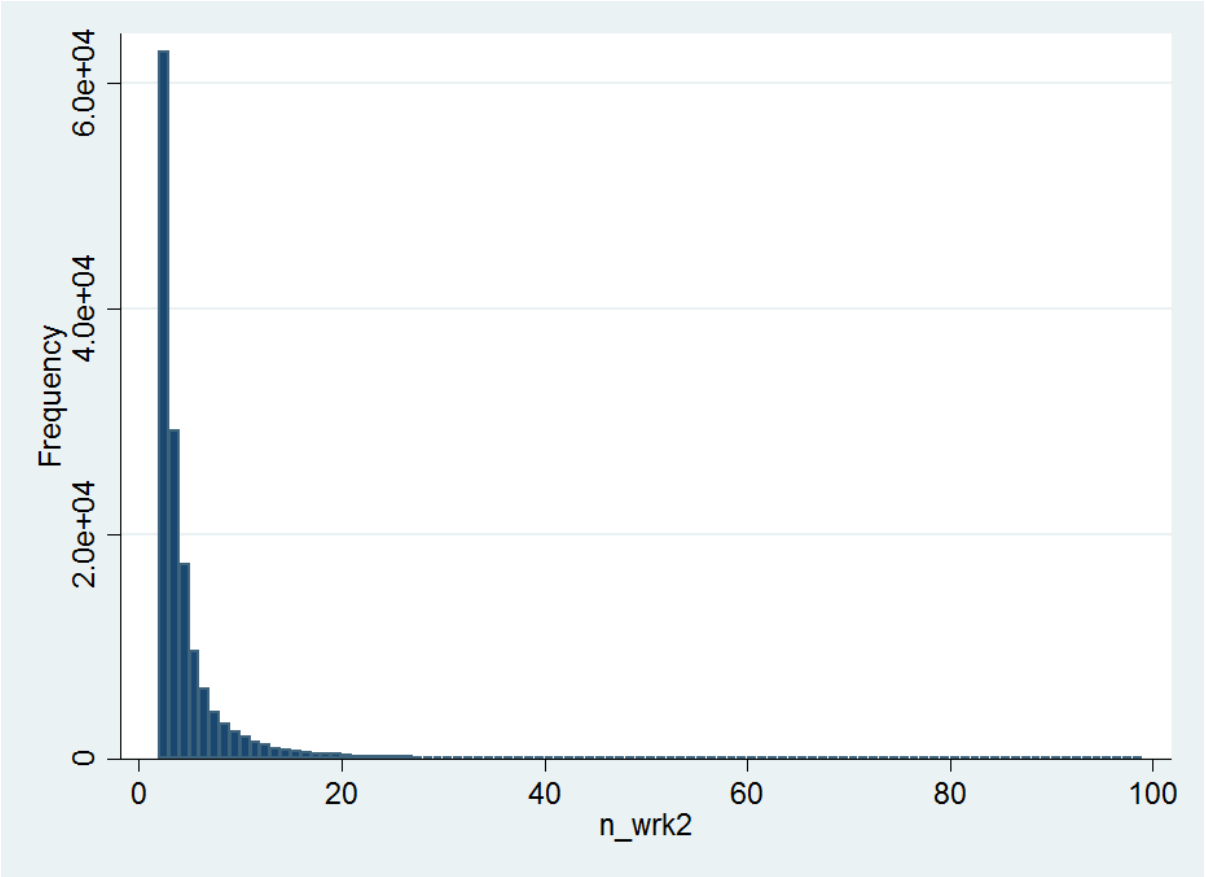
Notes: *Firm age* is in years, *Av profit* and *Av wage* in euros per year, *Av worker education* is coded as education above. *Av wage (daily)* is the average of daily wages of workers employed in the sample by each firm.

Av wage may include the wages of temporary workers, so Av wage (daily) is probably a more accurate indicator. There are less observations than per worker, because some firms include more than one worker.

Distributions and Histograms

1. Number of observations per firm
2. Average wage per firm
3. Average wage per worker
4. Firm size distribution
5. Average Profit by firm
6. Average Profit by entrepreneurs/business owners

1. Number of observations per firm (censored so n.workers>1 and n.workers<500, with is 1 worker)



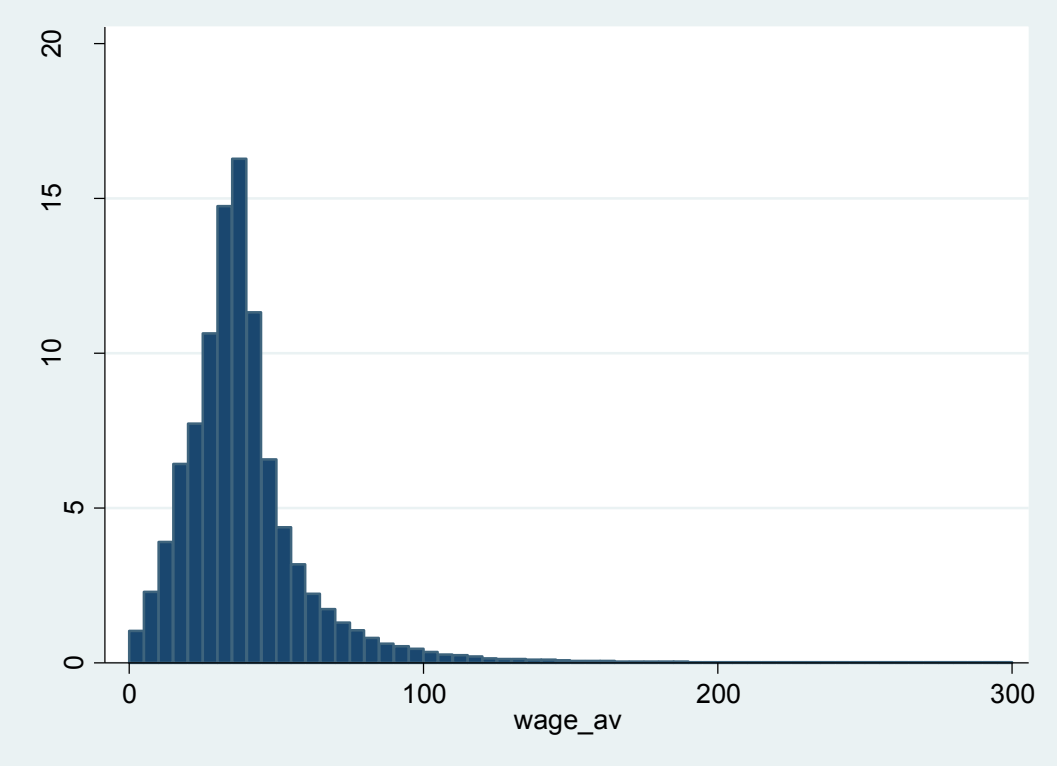
	mean	sd	min	max	count	p25	p50	p75
n_wrk2	6.22	17.39	2	497	150788	2	3	5

Notes: p25/p50/p75 – 25/50/75% percentile (50% percentile corresponds to the median). Count stands for number of observations.

- Very skewed to the left, the median is 3 observations per firm.
- Approximately 38% of all firms in the sample have more than one observation (150,788 / 403,372).

It could be seen as a problem that we have so few observations per firm. But the good news is that sample size is pretty big so even with two or three observations means are likely to be unbiased. Take as an example the over 75,000 of firms that have 2 or 3 workers.

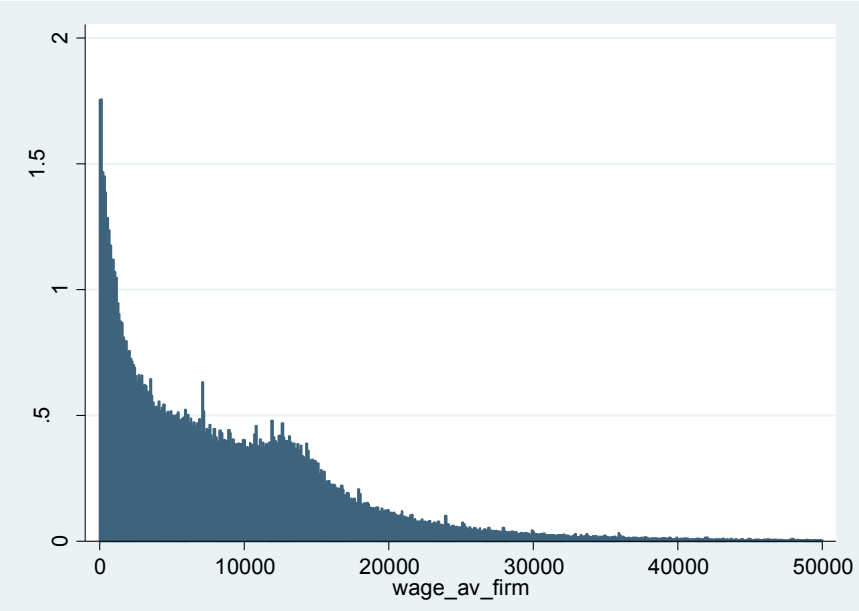
2. Average daily wage per firm (column with is 5€, censored at 300)



	mean	sd	min	max	median	90th perctl	N
Average wage	41.57382	92.41179	.0001111	26461.27	35.97176	63.53001	250,349
Monthly equ.	1247.215	2772.354	.0033333	793838.2	1079.153	1905.9	250,349

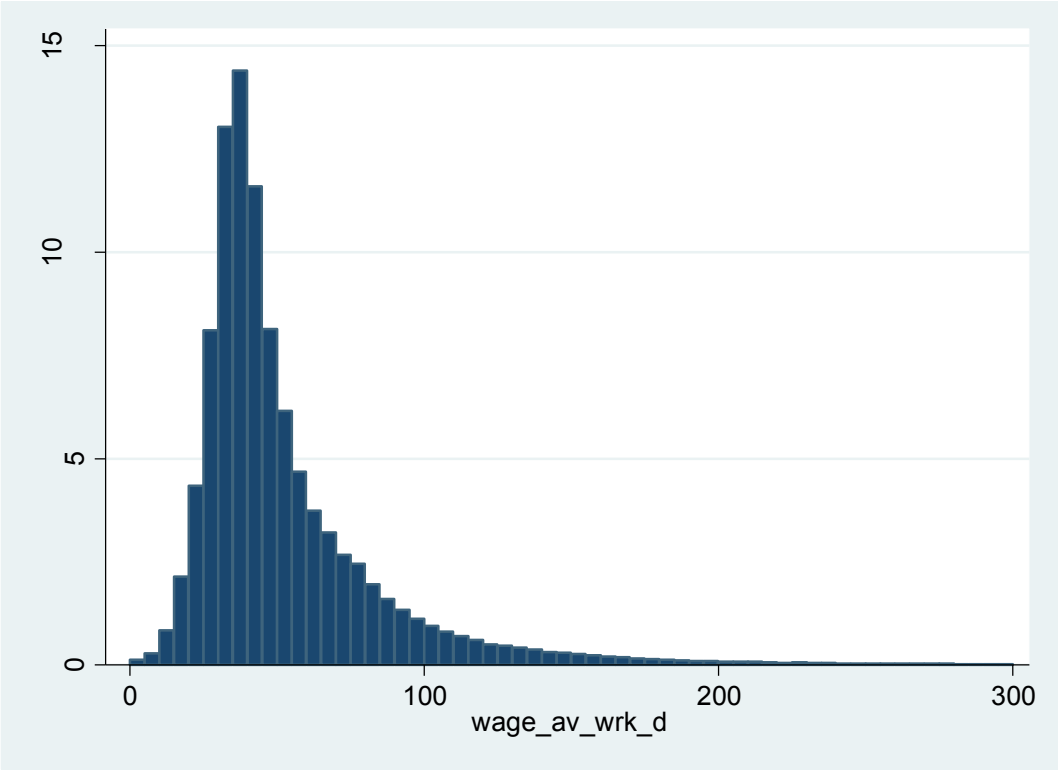
Notes: *Monthly equ.* = *Average wage**30, monthly equivalent average wage.

2.b Average yearly wage per firm (column with is 100€, censored at 50,000)



Notice the bump? It doesn't appear when we normalize the wage by days of service, meaning that everything from the bump to the left is probably the wages of temporary guys.

3. Average daily wage per worker (column width is 5€, censored at 300)



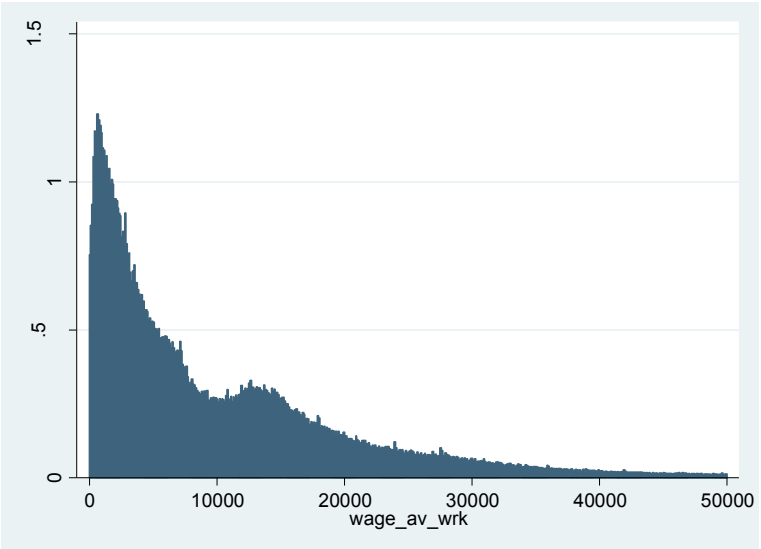
	mean	sd	min	max	p50	p90	count
Average wage	59.42004	241.2422	.0000274	68373.84	42.83677	94.32518	566884
Monthly equ.	1782.601	7237.265	.0008219	2051215	1285.103	2829.755	566884

Notes: *Monthly equ.* = *Average wage**30, monthly equivalent average wage.

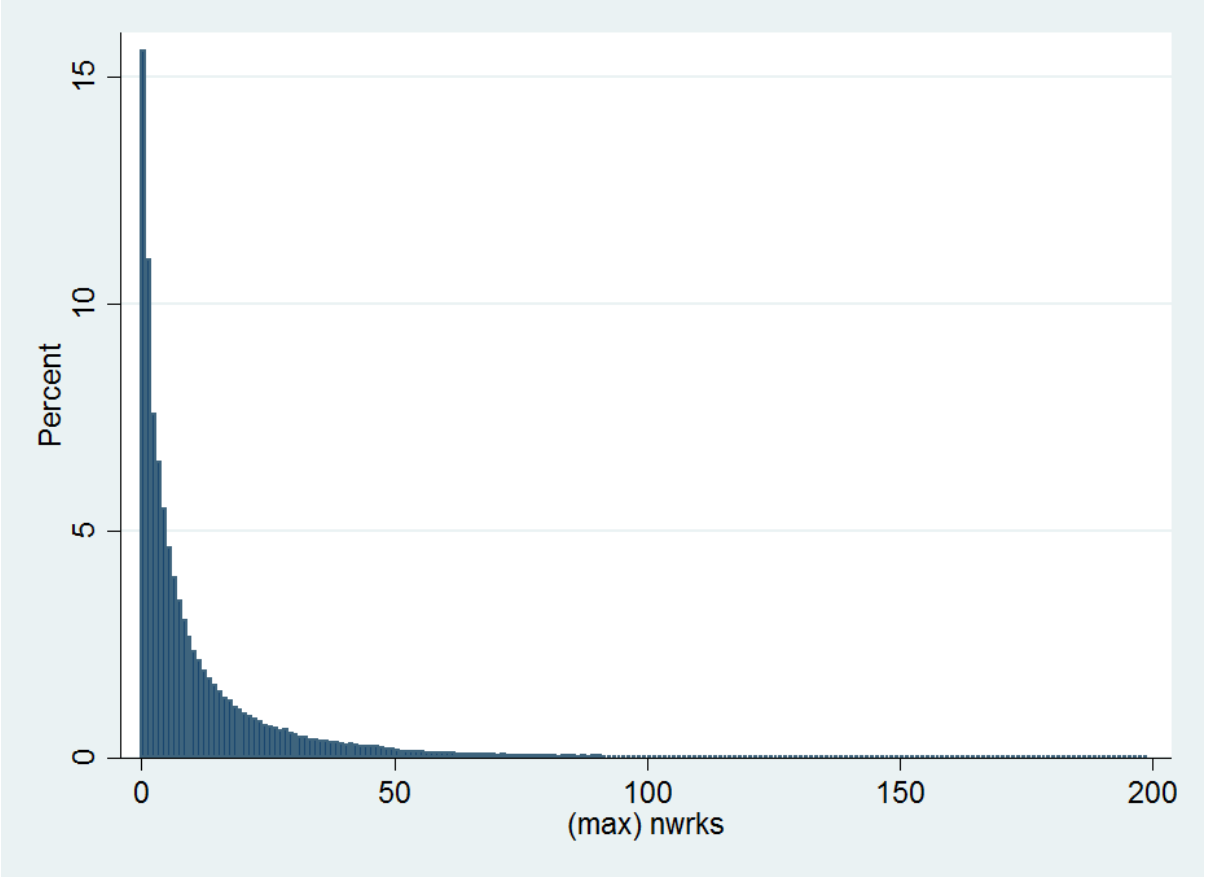
This histogram is similar, but not equal to the one before. Individual wages are more widely spread, compare the 90th percentile here (94.33) to the one for the firms (63.53). This makes perfect sense, as when calculating average wages per firm we average across different workers.

The measure of wages used here is gross wages – not unemployment subsidies, pensions or others.

3.b Average yearly wage per worker (column width is 100€, censored at 50,000)



4. Firm size distribution (bar width is 1 worker, censored at 200)



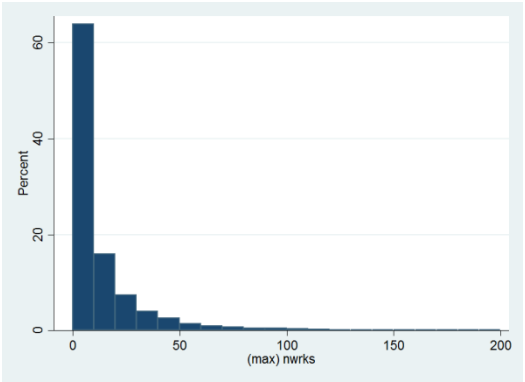
	mean	sd	min	max	count	p25	p50	p75
nwrks	20.12	268.03	0	96402	326454	1	5	16

Notes: p25/p50/p75 – 25/50/75% percentile (50% percentile corresponds to the median). Count stands for number of observations.

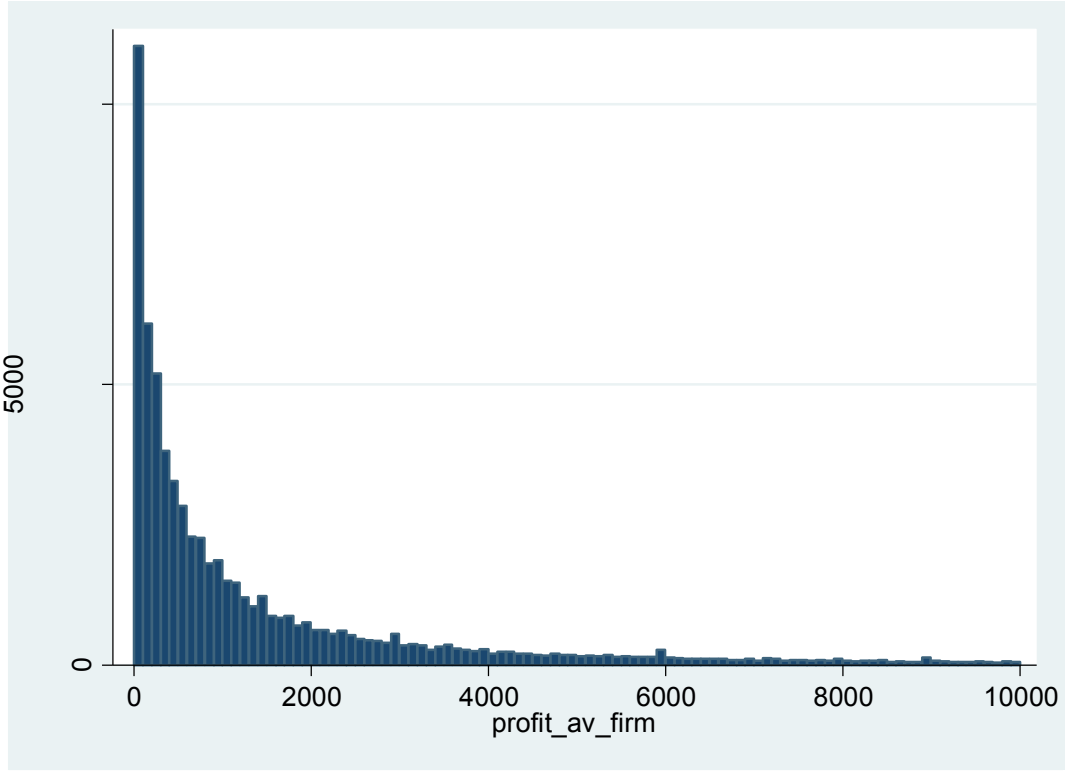
Of 326,454 individual firms, most of them very tiny: more than 60% of firms have less than 10 workers! This is a well stablished fact about Spanish firm size. It goes beyond small to the micro realm.

Notice how well this maps to the firms graph, were I censored the number of observations per firm to be strictly greater than one. And this is good indeed for interpreting the income of self-employed in individual firms as profits. As for bigger firms, we still have a big sample of workers.

Detail (bar width = 10)



5. “Profit” distribution by firm – Declared profits in euros (censored at 10,000, bar width of 100)



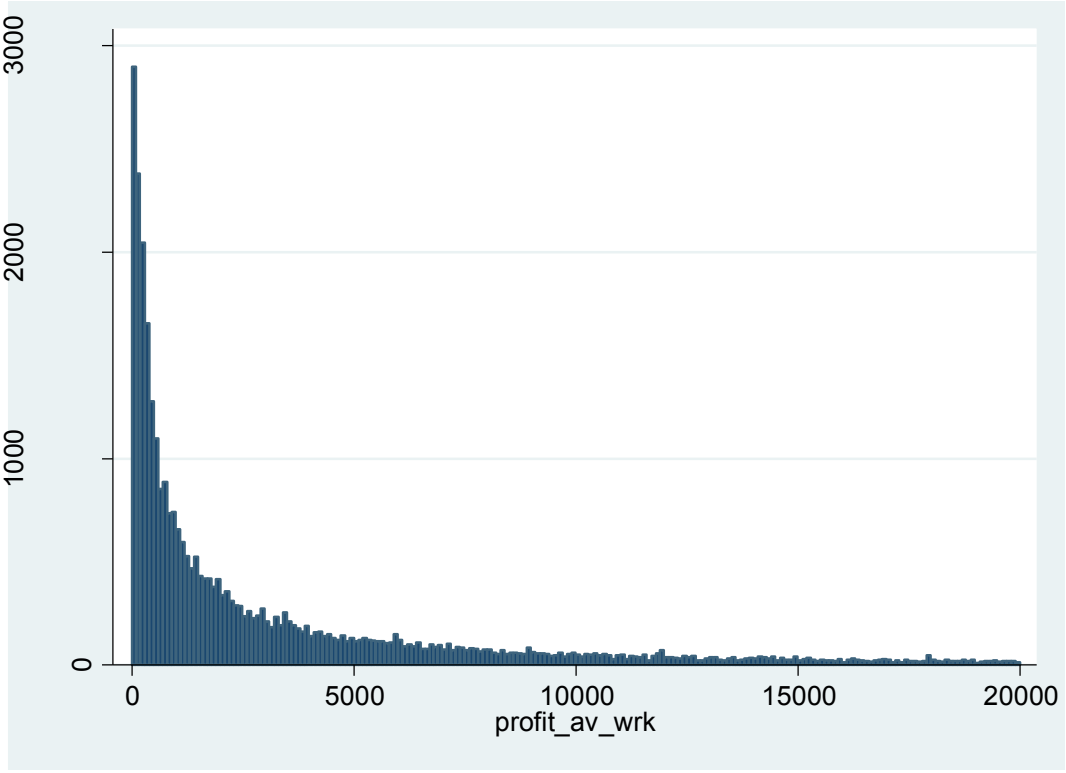
Variable	Obs	Mean	Std. Dev.	Min	Max
profit_av_~m	73315	3730.275	16469.81	.01	3005061

- “Only” 73,315 observations of individual firms, 18% of a total of 403,372 firms that we have information on.

I’m still cleaning the data, so more things are likely to count as profits as I progress. For example, there is an income category called “Administrators and CEOs” that is different from this one “Profits from economic activities”. I haven’t taken these into account, but it may as well reflect the productivity (measure in profitability) of the firm.

It’s difficult to assess how representative of the firm pool in Spain this sample is. It is definitely representative of the population contained in it, so the next graph (average profits per individual) may be a better picture on the whole. But considering how many micro firms there exist in Spain, and noting that all of those profits are declared by their owners as income tax, it is quite representative.

6. “Profit” distribution by individual – Declared profits in euros (censored at 20,000)



Variable	Obs	Mean	Std. Dev.	Min	Max
profit_av_~k	35486	6524.15	24845.03	.01	3005061

We have 35,486 unique observations. That means that many of the entrepreneurs have more than one firm – or they receive profits from more than one firm.

Of those 35,486, a total of 12,361 (35.06%) belong to people whose principal occupation is owning a firm. The rest is made of workers that have some income coming from profits – entrepreneurs in the sense that own and manage the firm.

Job category of those receiving profits		
	N	%
Self-employed	12,361	35.05772
OLF	4,542	12.88182
Permanent	8,223	23.32171
Temporary	8,747	24.80785
Unemployed	1,386	3.930911
Total	35,259	100

To be more precise, 35% of those receiving profits are classified as entrepreneurs in the Social Security. Note that workers may have more than one entry, and the one reflected here is the one that prevails at the end of the year. It won’t be far-fetched to consider all of them entrepreneurs – and interpret their rents as profits.

Some Toy Regressions

1. Profits on number of workers, average wage, and average education

```
. 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

APPENDIX

Simplified classification of educational level

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

It is difficult to transform levels into years of education, because people brought up in different educational regimes (and there have been at least 3 or 4 since 1960) have spent different years to achieve the same level. For example, Primary education used to last 8 years, but since the 90s it has been cut to 6. Conversely, Secondary education has been expanded, and so has pre-University education (*BUP* was one year, *Bachillerato* is two). There is still a difference (it used to be bigger) between short university degrees and long ones. I guess other countries may have similar problems.

And by the way, I should warn you that the education variable is only updated every ten years, as it comes from the census. This biases downwards the education variable for young people. But given that are not that many in the sample, the size of this bias is not big overall. If you look at people in their 20s it has probably a big effect. The sample from 2012 should be up to date.