Regional productivity differences in the UK and France: from the micro to the macro

Bridget Kauma (University of Sussex)
Giordano Mion (ESSEC and University of Nottingham)

Nottingham Lectures
March 2024

Motivation

Productivity quantifies how efficiently an economy uses its resources, i.e., transforms inputs into output.

- On the one hand, productivity is the basis for the living standards of an economy at both the national and regional levels.
- On the other hand, productivity gains are also the only sustainable source of long-term economic growth.

One key feature of productivity, and of economic activity in general, is that it is unevenly distributed across space and, in particular, across regions within a country.

Motivation

Consider for example the UK.

- Productivity, and thus wages and incomes, are much higher in the London area than anywhere else in the country.
- This gap is substantial – a factor of two or more for certain regions – and encompasses also other measures of economic and social wellbeing like employment and deprivation.

A similar picture arises for other countries and in particular for France.

Motivation

The literature on Regional and Urban Economics has long identified the importance of the density of economic activities as a key driver of local productivity

- This can be rationalized in a variety of ways (learning, matching, sharing, sorting)
- Rosenthal and Strange (2004) and Combes and Gobillon (2015) provide summaries of this literature and agree on a range for the key elasticity of productivity with respect to density of 0.02–0.10.
- These findings are robust to the endogeneity of current economic density and in particular to the use of long lags of historical density as instruments for current density (Ciccone and Hall, 1996; Ciccone, 2002).

Our contribution

- While most geographers would typically consider regions as the unit of analysis and directly work at this level of aggregation, economists are increasingly using firms–or even establishments–as the unit of analysis around which to reconstruct and attribute differences in economic performance across space.

- Crucially, the two approaches seem to provide (for manufacturing) different magnitudes regarding the elasticity of productivity with respect to local density (Jacob and Mion, 2020).

- We build upon two large-scale datasets (one for France and one for the UK) covering roughly the population of firms with one employee or more to systematically investigate the differences between the two approaches while going beyond manufacturing.

Our contribution

- In going from the micro (firm) to the macro (region) one key step is weighting: each firm needs to be weighted by its relative size within the region (typically employment).
- We find smaller values for the density elasticity when using unweighted firm-level regressions while getting larger values when considering revenue- or employment-weighted firm-level regressions.
- Keys to such differences are: 1) the correlation between productivity and firm size within a region; 2) the heterogeneity of the productivity return to density across the firm size distribution.

Our contribution

- While the correlation between firm productivity and firm size is positive in most regions it is systematically related to the density of the region.
- In particular, in denser areas the correlation is weaker meaning that more (less) productive firms are proportionally smaller (larger) if located in a denser region.
- However, the productivity return to density is strongly increasing in firm productivity (quantile regressions) and this effect dominates leading to an increase of the density coefficient when going from the micro (un-weighted regressions) to the macro (weighted regressions).

Data: UK

We use three panel data sets:

- **The Business Structure Database (BSD)** is an annual extract (the snapshot taking place at the end of a fiscal year) of the Inter-department Business Register (IDBR), a live database of business organisations in the UK.
- The BSD is administrated by the ONS and, while being one of the largest sources of data about business organisations in the UK, it contains only a limited number of variables.
- In our analysis, we borrow information about the number of employees, employment (number of employees plus owner(s)), industry affiliation and foreign ownership. A firm in the BSD is identified by a unique code named 'entref'.
- The BSD also provides information on the employment and location (up to the postcode level) of the different establishments belonging to a given firm that we also use in our analysis.

Data: UK

- **The Value Added Tax (VAT) panel** is an annual extract from VAT Returns providing information on organisations that are registered for VAT.
- The VAT panel database is administrated by HMRC and provide information on the value of purchases operated in a given (fiscal) year as well as the value of sales, which we use in our analysis
- A firm in the VAT panel database is identified by her unique VAT code that, once anonymised, is named 'vrn_anon'.

Data: UK

- **FAME** contains information on companies registered at Companies House in the UK. It covers company financials, corporate structures, shareholders and subsidiaries. The data are collected from various sources and are then compiled and organised by Bureau van Dijk (BvD).

- The coverage of variables like sales, intermediates purchases and employment in FAME is very patchy because only relatively large firms are required to report this information in their annual accounts.

- However, information on assets, an in particular on tangible fixed assets which we are going to use as our measure of the firm capital stock, is very well recorded. A firm in FAME is identified by her unique CHR number that, once anonymised in the HMRC Datalab, is named 'taxpayer_anon'.

Data: UK

- We match the 3 datasets over the period 2004-2017 building upon lookup Tables developed by HMRC and use an 'aggregate' definition of a firm encompassing the different identifiers.
- We focus on firms with at least one employee and filing a VAT declaration while eliminating firms involved in financial and insurance activities (SIC 2007 codes 64, 65 and 66).
- In the spatial analysis we use Travel to Work Areas (TTWA) as the geographical breakdown of the UK space (228 areas, 2011 version) and focus on single-TTWA firms (also provide results for all firms). Single-TTWA firms represents the vast majority of firms (around 97%) and account for about 43% of overall employment.

Data: France

For France the data job is much easier as there is a unique firm-identifier (SIREN). We use

- **FICUS** containing detailed accounting information (employment, sales, intermediates, capital, industry, etc.) for the population of French firms (SUSE scheme). Data is available up to 2007.
- **FARE** containing detailed accounting information (employment, sales, intermediates, capital, industry, etc.) for the population of French firms (ESANE scheme). Data is available from 2008 onwards.
- **Stock of establishments** allowing to link firms (SIREN codes) to establishments (SIRET codes) and providing the location (up to the municipality level) of the latter.

Data: France

- We match the 3 datasets over the period 2000-2017 using SIREN codes.
- We focus on firms with at least one employee while eliminating firms involved in financial and insurance activities (SIC 2007 codes 64, 65 and 66).
- In the spatial analysis we use Zones d'Emploi (ZE) as the geographical breakdown of France (297 areas for continental France, 2010 version) and focus on single-ZE firms (also provide results for all firms). Single-ZE firms represents the vast majority of firms (around 93%) and account for about 53% of overall employment.

TFP and Markups estimation

Denoting firms by i and time by t the production function we estimate is the following 3 inputs Cobb-Douglas:

$$R_{it} = L_{it}^{\alpha_L} M_{it}^{\alpha_M} K_{it}^{\alpha_K} A_{it},$$

where A_{it} is Total Factor Productivity (TFP) of firm i at time t, R_{it} is revenue, L_{it} is labour, M_{it} is intermediates, K_{it} is capital and α_L, α_M and α_K are the related output elasticities.

As for markups (Hall, 1986) cost-minimization of a variable input free of adjustment costs (intermediates) provides a simple rule to pin down markups μ_{it}:

$$\mu_{it} = \frac{\alpha_M}{s_{Mit}},$$

where s_{Mit} is the share of intermediates in revenue ($s_{Mit} \equiv \frac{W_{Mit} M_{it}}{R_{it}}$).

TFP and Markups estimation

We follow the estimation approach of Woodridge (2009) while assuming that: i) productivity follows an AR(1); ii) capital is predetermined, intermediates are fully-flexible and labour is semi-flexible.

Finally:

- We deflate revenue, intermediates and capital using corresponding indexes provided by the ONS/INSEE with the base year being 2017.
- We eliminate observations corresponding to a negative value added and apply a small trimming to the data (by industry) based on the ratios of: i) intermediates to sales; ii) capital to labour; iii) revenue to labour.
- We use a second-order polynomial in intermediates, capital and labour to smooth revenue and purge it from measurement error.
- We cluster standard errors at the firm-level.

Summary Stats 1: UK

Table: Key summary statistics across all years

	Mean	St.dev.	p5	p95	N. observ.
Revenue	4,305.85	219,289.69	31.93	5,287.83	9,954,131
Intermediates	3,159.46	171,339.37	7.38	3,655.35	9,954,131
Capital	2,424.68	245,007.26	1.20	664.60	9,954,131
Employment	21.95	622.40	1	38	9,954,131

Notes: Revenue, intermediates and capital are measured in thousand pounds. Values have been deflated using indexes provided by the ONS with the base year being 2017. Employment is number of employees count including the owner(s).

Summary Stats 1: France

Table: Key summary statistics across all years

	Mean	St.dev.	p5	p95	N. observ.
Revenue	2,968.16	92,725.74	64.62	5,751.42	17,641,530
Intermediates	2,110.03	70,285.80	23.92	3,897.24	17,641,530
Capital	1,806.16	174,892.63	6.27	1,717.29	17,641,530
Employment	12.74	372.00	1	32	17,641,530
Wage bill	581.54	19,062.60	15.20	1,344.75	17,641,530

Notes: Revenue, intermediates, capital and wage bill are measured in thousand euros. Values have been deflated using indexes provided by the INSEE with the base year being 2017. Employment is number of employees.

Summary Stats 2: UK

Table: Number of firms and total employment covered by year

Year	Number of firms	Total employment
2004	642,748	13,812,662
2005	681,104	14,198,956
2006	695,050	14,470,623
2007	717,933	14,851,475
2008	701,827	15,378,391
2009	684,485	15,307,760
2010	681,465	15,294,427
2011	700,898	15,544,064
2012	692,865	15,899,287
2013	716,939	16,263,075
2014	728,632	16,362,476
2015	740,365	16,609,343
2016	755,413	17,058,927
2017	814,407	17,441,714

Notes: Employment is number of employees count including the owner(s). Data are organised by fiscal year with, for example, the year 2017 corresponding to the fiscal year 2017-18.

Summary Stats 2: France

Table: Number of firms and total employment covered by year

Year	Number of firms	Total employment
2000	1,025,542	12,006,862
2001	1,012,852	12,294,591
2002	1,021,618	12,440,875
2003	1,044,963	12,073,664
2004	1,077,003	12,700,392
2005	1,046,706	12,570,017
2006	1,113,641	12,956,367
2007	1,144,423	13,018,617
2008	927,707	12,636,208
2009	927,597	12,294,506
2010	937,374	12,527,977
2011	936,053	12,659,021
2012	919,392	12,512,977
2013	871,200	12,328,195
2014	909,314	12,383,382
2015	885,391	12,543,022
2016	940,728	12,325,677
2017	900,026	12,406,277

Notes: Employment is number of employees.

Time evolution: UK

Table: Average (employment weighted) apparent labour productivity, labour productivity, OLS TFP, WLD TFP and markups by year

Year	Apparent Lab. Prod.	Lab. Prod.	OLS TFP	WLD TFP	Markups	N. of firms
2004	192,796	57,577	3.490	3.036	1.558	642,748
2005	202,646	57,349	3.514	3.055	1.565	681,104
2006	201,485	57,688	3.546	3.084	1.545	695,050
2007	209,504	56,681	3.543	3.079	1.561	717,933
2008	188,056	47,892	3.537	3.070	1.533	701,827
2009	179,307	47,832	3.528	3.062	1.534	684,485
2010	189,490	44,674	3.547	3.075	1.512	681,465
2011	191,634	43,756	3.548	3.074	1.513	700,898
2012	191,446	46,667	3.557	3.084	1.527	692,865
2013	190,029	47,480	3.594	3.123	1.532	716,939
2014	199,459	50,321	3.661	3.193	1.559	728,632
2015	197,796	54,829	3.706	3.237	1.570	740,365
2016	204,431	58,751	3.703	3.233	1.591	755,413
2017	206,930	59,777	3.736	3.268	1.620	814,407

Time evolution: UK

- While total factor productivity (both OLS TFP and WLD TFP) has only been both very lightly and very briefly affected by financial crisis, the same is not true for markups, apparent labour productivity and labour productivity, which is consistent with evidence provided in analyses bases on the smaller ARD/ABS datasets

- Inspection of markups reveals that they recovered their pre-financial crisis level around 2015 while for labour productivity the recovery year is 2016.
- Results are similar if we split the sample into single-TTWA firms (small firms) and multi-TTWA firms (large firms) with the recovery being stronger for multi-TTWA firms (large firms). For single-TTWA firms (small firms), labour productivity in 2017 is still below pre-financial crisis levels.

Time evolution: France

Table: Average (employment weighted) apparent labour productivity, labour productivity, OLS TFP, WLD TFP and markups by year

Year	Apparent Lab. Prod.	Lab. Prod.	OLS TFP	WLD TFP	Markups	N. of firms
2000	223,361	64,876	1.654	2.487	1.265	1,025,542
2001	224,933	66,008	1.659	2.498	1.256	1,012,852
2002	228,402	66,431	1.651	2.480	1.261	1,021,618
2003	231,482	66,605	1.628	2.332	1.257	1,044,963
2004	229,501	66,766	1.657	2.483	1.266	1,077,003
2005	232,417	67,696	1.663	2.490	1.264	1,046,706
2006	235,324	67,523	1.662	2.486	1.261	1,113,641
2007	239,435	67,855	1.666	2.489	1.262	1,144,423
2008	233,018	66,057	1.557	2.340	1.231	927,707
2009	222,918	67,669	1.614	2.482	1.246	927,597
2010	226,106	66,514	1.606	2.469	1.241	937,374
2011	239,428	66,912	1.617	2.482	1.232	936,053
2012	237,460	67,776	1.620	2.506	1.240	919,392
2013	237,251	68,404	1.626	2.523	1.244	871,200
2014	239,378	68,546	1.621	2.527	1.237	909,314
2015	232,609	67,769	1.622	2.503	1.246	885,391
2016	238,234	69,863	1.621	2.498	1.243	940,728
2017	242,910	69,482	1.620	2.493	1.244	900,026

Time evolution: France

- It is not entirely clear whether total factor productivity has by 2017 picked up its pre-financial crisis level (OLS vs WLD). On the other hand, apparent labour productivity and labour productivity have been little affected by the financial crisis.
- Inspection of markups reveals that they have not yet recovered their pre-financial crisis level suggesting that:
  - Firms seem to really struggle to achieve pre-financial crisis profit margins
- Results are similar if we split the sample into single-TTWA firms (small firms) and multi-TTWA firms (large firms) with the recovery being stronger for multi-TTWA firms (large firms). For multi-TTWA firms (large), total factor productivity has definitely picked up its pre-financial crisis level.

Spatial Analysis

We are interested in the variation of TFP across regions and how it is affected by aggregation/weighting. The baseline estimation equation is:

$$\bar{a}_{it} = \gamma density_{r(it)} + I_{r(it)} + I_t + \epsilon_{it},$$

where

- $\bar{a}_{it}$ is log TFP demeaned by the corresponding industry average (we net out composition effects)
- $density_{r(it)}$ is the log density of population in region r where firm i is observed at time t,
- $I_{r(it)}$ and I_t are region and year dummies
- ϵ_{it} is an error term.

Spatial Analysis

The **aggregate** productivity of a region is the **weighted** (typically by employment) productivity of the firms located in the region.

We now focus on single-region (ZE or TTWA) firms because we can uniquely attribute their productivity to a location. Yet we also provide robustness using all firms/establishments while attributing the same productivity to all of the establishments of a firm

We perform un-weighted and weighted (by employment) OLS estimations and cluster standard errors at the region-year (ZE or TTWA) level

We are interested in the estimates of γ and $I_{r(it)}$ and in particular by how much, if anything, those estimates get larger if we consider weighting, i.e., if we switch from the micro (firms) to the macro (regions).

Results: UK

Table: Spatial regressions

	Unweighted	Weighted
log density	0.0178***	0.0208***
	(0.0007)	(0.0009)
	Reference category is London	
East Midlands	-0.0676***	-0.1004***
	(0.0027)	(0.0072)
East of England	-0.0421***	-0.1056***
	(0.0031)	(0.0082)
North East	-0.0765***	-0.1021***
	(0.0030)	(0.0084)
North West	-0.0767***	-0.1252***
	(0.0029)	(0.0074)
Northern Ireland	-0.0211***	-0.0877***
	(0.0032)	(0.0073)
Scotland	-0.0128***	-0.0749***
	(0.0043)	(0.0077)
South East	-0.0293***	-0.0902***
	(0.0033)	(0.0074)
South West	-0.0718***	-0.1129***
	(0.0030)	(0.0069)
Wales	-0.0828***	-0.1373***
	(0.0030)	(0.0077)
West Midlands	-0.0772***	-0.1147***
	(0.0026)	(0.0070)
Yorkshire and The Humber	-0.0804***	-0.1137***
	(0.0028)	(0.0076)
Observations	9,663,658	9,663,658
R-squared	0.0094	0.0071
R-squared 'aggregate'		0.2613

Results: UK

The weighted γ is around 2% and in line with the literature.

All regions in the UK suffer, beyond density, a strong productivity gap with respect to London

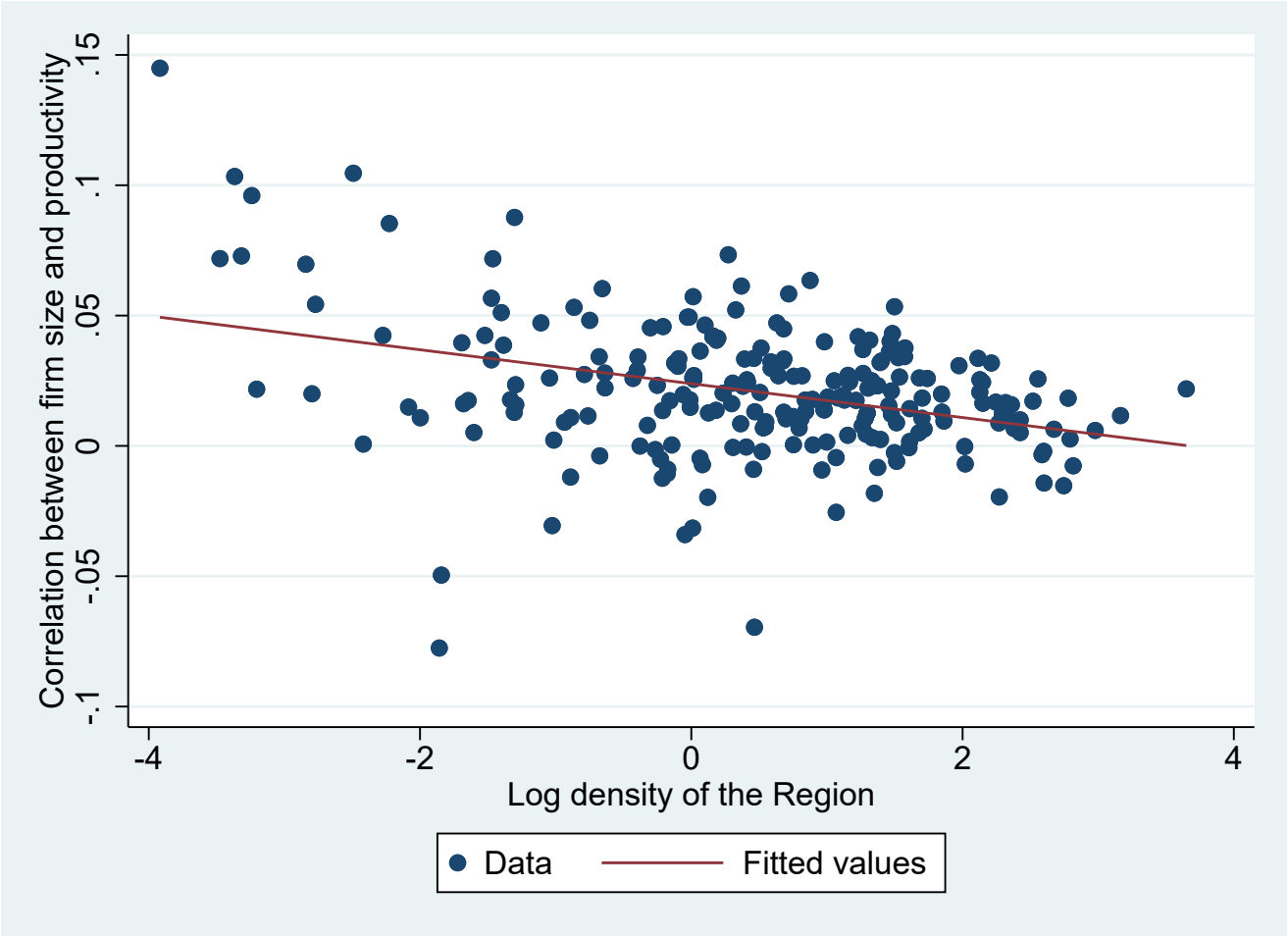
These estimates imply that the aggregate productivity difference between the median density region (Banbury, East Midlands) and London is 16.6% while the unweighted productivity difference between firms in the two regions is 12.4%, i.e., the latter accounts for about 75% of the aggregate difference.

The difference between the two sets of estimates could be driven by:

1. A correlation between firm size and productivity varying across regions and in particular increasing with density (NOPE! Actually slightly decreasing)
2. A productivity return on density being stronger for the most productive firms (YES!)

Results: UK

Figure: UK: Correlation between firm productivity and size within each region



Results: UK

Table: UK: Quantile regressions and the heterogeneous impact of density across the productivity distribution

VARIABLES	1st decile	2nd decile	3rd decile	4th decile	5th decile	6th decile	7th decile	8th decile	9th decile
log density	0.0115***	0.0133***	0.0124***	0.0115***	0.0122***	0.0144***	0.0180***	0.0238***	0.0355***
	(0.0004)	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0004)

Results: France

Table: Spatial regressions

	Unweighted	Weighted
log density	0.0041***	0.0196***
	(0.0010)	(0.0051)
	Reference category is Île-de-France	
Auvergne-Rhône-Alpes	-0.0098*	-0.0207
	(0.0057)	(0.0185)
Bourgogne-Franche-Comté	-0.0162***	-0.0189
	(0.0059)	(0.0181)
Bretagne	0.0096*	0.0033
	(0.0051)	(0.0221)
Centre-Val de Loire	-0.0118**	-0.0134
	(0.0059)	(0.0185)
Grand Est	-0.0181***	-0.0115
	(0.0051)	(0.0196)
Hauts-de-France	-0.0111**	-0.0178
	(0.0044)	(0.0215)
Normandie	-0.008	-0.0064
	(0.0052)	(0.0215)
Nouvelle-Aquitaine	-0.0229***	-0.0320*
	(0.0055)	(0.0181)
Occitanie	-0.0410***	-0.0367**
	(0.0068)	(0.0184)
Pays de la Loire	-0.0001	-0.001
	(0.0053)	(0.0198)
Provence-Alpes-Côte d'Azur	-0.0375***	-0.031
	(0.0062)	(0.0206)
Multi-region	-0.0182*	-0.0132
	(0.0095)	(0.0226)
Observations	16,595,355	16,595,355
R-squared	0.0050	0.0089
R-squared 'aggregate'		0.2458

Results: France

The weighted γ is around 2% and in line with the literature. Though the un-weighted is much smaller.

(Contrary to the UK!) Regions in France do not suffer, beyond density, any strong productivity gap with respect to Paris

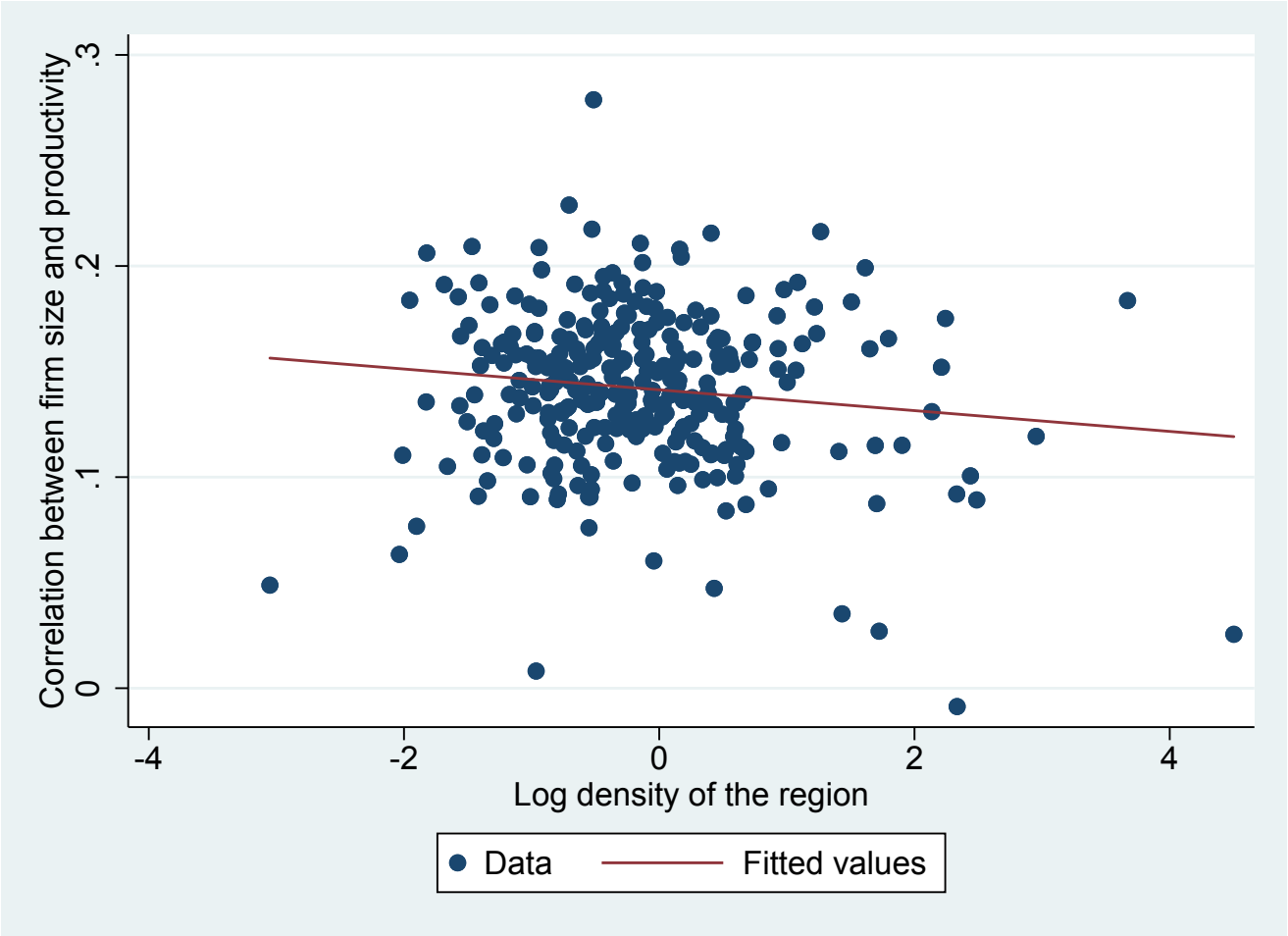
These estimates imply that the aggregate productivity difference between the median density region (Saint-Dié-des-Vosges, Grand Est) and Paris is 10.5% while the unweighted productivity difference between firms in the two regions is 3.77%, i.e., the latter accounts for about 36% of the aggregate difference.

The difference between the two sets of estimates could be driven by:

1. A correlation between firm size and productivity varying across regions and in particular increasing with density (NOPE! Actually slightly decreasing)
2. A productivity return on density being stronger for the most productive firms (YES!)

Results: France

Figure: France: Correlation between firm productivity and size within each region



Results: France

Table: France: Quantile regressions and the heterogeneous impact of density across the productivity distribution

VARIABLES	1st decile	2nd decile	3rd decile	4th decile	5th decile	6th decile	7th decile	8th decile	9th decile
log density	-0.0100***	-0.0052***	-0.0018***	0.0007***	0.0030***	0.0050***	0.0075***	0.0113***	0.0189***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)

Different problems

- In the UK the correlation between firm size and productivity within a region is low (compared to France) and sometimes negative. If the UK had the French correlations aggregate productivity would be higher. Also the UK has a problem of productivity being quite unequal across space beyond density (big London gap while little Paris gap).

- The problem with France is the NEGATIVE return on density for the least productive firms, i.e., denser places have too many low productive firms. This creates a big divide between the un-weighted and weighted return on density.

Conclusions

- We assemble two comparable datasets for France and the UK covering the population of firms with one employee or more spanning over the period 2000/2004-2017.
- This delivers us with enough information to estimate TFP and markups for an unprecedentedly large number of firms allowing for comprehensive longitudinal analyses and granular regional-level investigations for both countries.
- We provide some figures about the overall evolution of labour productivity, TFP and markups in the two countries highlighting similarities and differences.
- We provide some preliminary insights into the source of regional productivity differences for the two countries.

Thank you for your attention