Centre for Economic Performance

Discussion Paper

ISSN 2042-2695

No. 1955
November 2023

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

Bridget Kauma
Giordano Mion

The London School of Economics and Political Science

UKRI Economic and Social Research Council

Abstract
We propose a new data resource that attempts to overcome limitations of standard firm-level datasets for the UK (like the ARD/ABS) by building on administrative data covering the population of UK firms with at least one employee. We also construct a similar dataset for France and use both datasets to: 1) Provide some highlights of the data and an overall picture of the evolution of aggregate UK and French productivity and markups: 2) Analyse the spatial distribution of productivity in both countries at a fine level of detail – 228 Travel to Work Areas (TTWAs) for the UK and 297 Zones d'emploi (ZEs) for France – while focusing on the role of economic density. Our findings suggest that differences in firm productivity across regions are magnified in the aggregate by an increasing productivity return of density along the productivity distribution.

Keywords: firm-level dataset, merging, BSD, FAME; VAT, FICUS, FARE, productivity, markups, UK, France, regional disparities, density.
JEL Codes: R12; D24

This paper was produced as part of the Centre's Urban Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

We are grateful to Josh Martin, Mary O'Mahony, Nick Oulton, Rebecca Riley, Bart Van Ark, Tony Venables as well as participants of various seminars and conferences for helpful comments. The authors acknowledges financial support from Labex MME-DII (France) and The Productivity Institute (UK). HM Revenue & Customs (HMRC) agrees that the figures and descriptions of results in the attached document may be published. This does not imply HMRC's acceptance of the validity of the methods used to obtain these figures, or of any analysis of the results. Please note that all statistical results remain Crown Copyright, and should be acknowledged either as such and/or as "Source: Calculations based on HMRC administrative datasets". Copyright of the statistical results may not be assigned. Written work intended for publication should include a note to the effect that: "This work contains statistical data from HMRC which is Crown Copyright. The research datasets used may not exactly reproduce HMRC aggregates. The use of HMRC statistical data in this work does not imply the endorsement of HMRC in relation to the interpretation or analysis of the information."

Bridget Kauma, University of Sussex. Giordano Mion, ESSEC Business School, University of Nottingham and Centre for Economic Performance at London School of Economics.

Published by
Centre for Economic Performance
London School of Economic and Political Science
Houghton Street
London WC2A 2AE

1. Introduction

Productivity quantifies how efficiently an economy uses the resources it has available, by relating the inputs to output. On the one hand, productivity provides the basis for the living standards of a country. On the other hand, productivity is 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.

In this respect, a stylized fact of economic geography is that the productivity of firms increases with city size and urban density (Combes and Gobillon, 2015), and a large literature going back to Marshall (1890) explores the question of why cities have this productivity advantage. Micro-foundations put forward for these agglomeration externalities are typically grouped under the headings sharing, matching, learning and sorting (Duranton and Puga, 2004, Combes et al., 2008) and include different forms of knowledge spillovers between firms, costly trade, pro-competitive effects of city size, and sorting of workers (Syverson, 2011). The empirical literature suggests a rather consistent, across countries and years, range for the elasticity of productivity with respect to city size. 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.[1] 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).

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 regions. Crucially, the two approaches do not seem to provide the same magnitudes regarding, for example, the elasticity of productivity with respect to local density. More specifically, Jacob and Mion (2020a) provide evidence for French manufacturing firms highlighting the importance of weighting in going from the firm-level (micro) to the regional-level (macro) productivity. They find smaller values for the elasticity of productivity with respect to population density when using unweighted firm-level regressions while getting quite larger values when considering revenue- or employment-weighted firm-level regressions.

The productivity of a region is clearly the productivity of its firms. However, the aggregate productivity of a region is the weighted average (typically by employment) of the productivity of the firms located in the region and not the simple average. When running unweighted regressions, in which firms are the unit of analysis, to measure productivity differences across space one is essentially comparing the average firm across different locations irrespective of

[1] See also Combes et al. (2008), Mion and Naticchioni (2009) and De La Roca and Puga (2017) for estimates of the elasticity of worker-level wages with respect to density.

the firm size distribution, and its link to productivity, within regions. The link between micro and macro is restored if one runs weighted regressions and the coefficients from the unweighted and weighted regressions do not need to be the same. One reasons they could differ is a varying (across regions) correlation between firm size and firm productivity. For example, if denser regions are characterized by a higher correlation between firm size and firm productivity, unweighted differences in productivity across space will be magnified when weighting. Another reason for differences between coefficients is the heterogeneity (along the productivity dimension) of the elasticity of productivity with respect to density. For example, if more productive firms enjoy disproportionately from the density of economic activities, unweighted differences in productivity across space will again be magnified when weighting because (on average) more productive firms are larger.

In what follows we extend the analysis of Jacob and Mion (2020a) beyond manufacturing to the whole private sector for both France and the UK while digging into the above mentioned explanations for the larger values of the elasticity of productivity with respect to population density when using weighted firm-level regressions as compared to unweighted firm-level regressions. In order to achieve this, we first construct two large datasets spanning the entire population of French and UK firms with at least an employee allowing us to retrieve different measures of productivity – including labour productivity and total factor productivity (TFP) – and investigate the links between productivity and geography at a fine spatial level: 228 Travel to Work Areas (TTWAs) for the UK and 297 Zones d'emploi (ZEs) for France. Considering the last year of the data, i.e., 2017 the datasets we constructed span over 814,407 firms employing 17,441,714 workers for the UK and over 900,026 firms employing 12,406,277 workers for France. In both cases, the availability of the location of the different establishments belonging to each firm allows us to link productivity to space and perform our investigations.

After describing the construction of the two datasets which has, particularly for the UK, involved a considerable amount of effort, we provide some data highlights regarding productivity, markups and the financial crisis period, which is included in the time span of our analysis (2000-2017 for France and 2004-2017 for the UK). Considering the UK, while total factor productivity has only been both very lightly and very briefly affected by the financial crisis, the same is not true for markups, apparent labour productivity and labour productivity, which is consistent with evidence provided in analyses based on the smaller ARD/ABS datasets like Harris and Moffat (2017) and Jacob and Mion (2020b). Inspection of markups reveals that they recovered their pre-financial crisis level around 2015 while for labour productivity the recovery year is 2016. As for France, it is not entirely clear whether total factor productivity has by 2017 picked up its pre-financial crisis level. 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 French firms struggle to achieve pre-financial crisis profit margins.

Turning to the spatial analysis we focus in most of our investigation on 'single region firms', i.e, firms that we can uniquely associate to one region. Such firms may thus have more than one establishment, but such establishments need to be located in the same region. The reason we are particularly interested in single region firms is that for such firms there is no issues in, for example, attributing their productivity and their employment to a particular region. Single region firms represents the vast majority of firms and account for about half of overall employment. We also provide some robustness results including multi region firms in the analysis, while attributing the same productivity to all of the establishments of a given multi region firm and using establishment-level employment for weighting. Such robustness results largely confirm our finding based on single region firms.

Our results can be summarised as follows. First, for both France and the UK we find a larger productivity return to density when weighting observations by employment as compared to unweighted regressions. Digging deeper into this reveals, in both cases, that: 1) The correlation between firm size and productivity within a region is quite low (and sometimes negative) across regions particularly for the UK; 2) The relationship between these correlations and region density is not positive and actually slightly negative. These findings indicate that, if anything, the varying correlation between firm size and productivity across regions should *reduce* and not amplify spatial productivity differences when going from the micro to the macro. On the other hand, in both cases, we find evidence that the productivity return of density is increasing along the deciles of the productivity distribution. This finding is reminiscent of the 'dilating' of the productivity distribution in larger regions found for France in Combes et al. (2012) and it is such an increasing productivity return of density that magnifies firm-level productivity differences for the UK and France, when going from the micro to the macro, and not a varying correlation between firm size and productivity across space.

In terms of the comparison between the UK and France we find the following. 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 instead the negative productivity return of density for the least productive firms, i.e., denser places in France nurture too many low productive firms and this creates a big divide between the un-weighted and weighted productivity return of density.

The rest of the paper is organized as follows. Sections 2 and 3 present the data sources we use and describe how we cleaned and combined the data together for the UK and France, respectively. Section 4 provides details of the productivity and markups estimations while Section 5 presents some data highlights and an overall picture of the evolution of aggregate UK and French productivity and markups. Section 6 contains our spatial analysis. Section 7 concludes. Some complementary Tables and Figures are reported in the Appendix.

2. UK data

2.1 *Data sources*

2.1.1 BSD

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. Organisations that are registered for VAT or pay at least one member of staff through the Pay As You Earn (PAYE) tax system, will appear on this register.

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)) and foreign ownership. A firm in the BSD is identified by a unique code to which we refer to as the 'BSD firm id'. 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. An establishment in the BSD is identified by a unique code to which we refer to as the 'BSD establishment id'.

2.1.2 VAT

The Value Added Tax (VAT) panel database 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, among other, the value of purchases operated in a given (fiscal) year as well as the value of sales. A firm in the VAT panel database is identified by her unique VAT code, which is anonymised within the HMRC datalab environment, and to which we refer to as the 'VAT firm id'.

2.1.3 FAME

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, most notably the national official bodies in charge of collecting company accounts data, and are then compiled and organised by Bureau van Dijk (BvD). FAME is available within the HMRC Datalab where original company identifiers are anonymised.

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, and 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 anonymised CHR number to which we refer to as the 'FAME firm id'.

2.2 Cleaning and combining the data

In what follows we explain how we cleaned and merged the data. The data are organised by fiscal year and, for example, when referring to the year 2017 we actually mean the fiscal year 2017-18.

2.2.1 Data cleaning

BSD. For the BSD we first worked on the industry classification to consistently have information on the SIC 2007 primary code of each firm. The SIC 2007 industry affiliation is not available in the 2004 and 2005 vintages of the data (only SIC 2003 is available) but we exploited the fact that both SIC 2003 and SIC 2007 are available from 2006 onwards to build a correspondence table that we applied to earlier years.

We have subsequently eliminated firms involved in financial and insurance activities (SIC 2007 codes 64, 65 and 66) and restricted the sample to firms with at least one employee and with a live vat status. This latter restriction allows to deal with an otherwise inexplicable drop of the number of firms around 2010. A firm in the data is identified by the BSD firm id and the data spans from 2004 to 2017.

VAT. Again we applied some cleaning to the industry classification (which is time varying in the VAT panel dataset) and eliminated firms involved in financial and insurance activities (SIC 2007 codes 64, 65 and 66). We checked for values meaning and consistency across years and kept only firms for which values of sales and acquisitions are both non missing and greater than zero. We also kept information about sales and acquisitions related to EU countries for future use. A firm in the data is identified by the VAT firm id and the data spans from 2004 to 2017.

FAME. We cleaned the data from some duplicates and keep only observations for which the variable fixed assets is not missing. As for the other datasets, we applied some cleaning to the industry classification (which is consistently SIC 2003 in the dataset) and eliminated firms involved in financial and insurance activities (SIC 2003 codes 65, 66 and 67). A firm in the data is identified by the FAME firm id and the data spans from 2004 to 2017.

2.2.2 Data matching

Each of the 3 datasets has a different firm identifier and the correspondence between any pair of identifiers is in some cases many to many. The HMRC datalab provides a lookup Table across the 3 identifers but the many to many correspondence issue still needs to be addressed. A simple example highlighting the many to many issue, and how we deal with it, is reported in Table 1 below.

Table 1: Example of correspondences

BSD firm id	VAT firm id
A	1
A	2
A	3
A	4
A	5
B	3
B	4
B	6
B	7
C	8
D	8

The example in Table 1 is related to the correspondence between the BSD firm id (for which we use letters) and the VAT firm id (for which we use numbers). Table 1 indicates that the BSD firm id A is linked to many VAT firm id and in particular to 1, 2, 3, 4 and 5. This would not be a problem (being simply a case of one to many) if VAT firm id 3 and 4 were not also linked to the BSD firm id B, which is also connected to VAT firm id 6 and 7. On the other hand, the case of BSD firm id C and D is simpler because they are both related to the VAT firm id 8, which in turn is not related to other BSD firm id (a simple case of many to one). For our analyses we have devised a looping code that would 'aggregate' BSD and VAT codes in such a way to get, in the case of Table 1, two 'combined firm id' (for which we use Greek letters). The first combined firm id α would correspond to BSD firm id A and B as well as to VAT firm id 1, 2, 3, 4, 5, 6, 7. The second combined firm id β would correspond to BSD firm id C and D as well as to the VAT firm id 8. Once solved the issue of the many to many cases for the BSD firm id and the VAT firm id, we apply the same procedure using the correspondence between the combined firm id and the FAME firm id, which will yet generate another more aggregate firm id, encompassing the three different firm identifiers, to which we refer to as the 'final firm id'. At the end of the procedure, each original firm id (BSD, VAT and FAME) will be associated to a unique final firm id.

Armed with this notion we then aggregate the information coming from the three datasets at the final firm id level. For example, we sum the sales of the different VAT codes corresponding to a given final firm id and impute as SIC 2007 code of a final firm id the SIC 2007 code corresponding to the BSD firm id with the largest employment among the different BSD firm id linked to the final firm id considered.

We first match the BSD with the VAT data and keep only firms present in both datasets. This entails a drop of about 4 to 5 million employees per year that is concentrated in sectors where public employment is more prevalent. We then match FAME, which at this stage entails a minimal loss in terms of firms lost in the match, and apply some final cleaning and polishing to the capital stock variable to increase coverage. Finally, we define industries as two digit SIC 2007 codes and apply some grouping (detailed below) in preparation for TFP estimations.

2.2.3 *Adding information on location*

In order to retrieve the location(s) of a firm we use the information on local units from the establishments files of the BSD. Each BSD establishment id is uniquely linked to a BSD firm id and so to a unique final firm id. For each final firm id in our data we can then identify the related establishments and for each such establishment the BSD provides information on location (up to the postcode level) and employment. In view of conducting a meaningful spatial analysis, we use an 'economic' partition of the UK geography and in particular the 2011 version of the Travel To Work Areas (TTWAs). The 2011 version of TTWAs breaks down the UK (including Northern Ireland) into 228 areas.

In order to go from postcodes to 2011 TTWAs for the whole of our sample period we use a postcode directory provided by the ONS. The match between the postcode directory and the postcodes in the data works very well and only requires minor adjustments. Starting from the year 2017 (fiscal year 2017-18), only the first part of the postcode is available in the BSD data but fortunately information on the corresponding TTWA 2011 version is also provided.

Equipped with the information above we are thus able to identify what we refer to as 'single TTWA firms', i.e, firms that we can uniquely associate to one TTWA. Such firms may thus have more than one establishment, but such establishments need to be located in the same TTWA. The reason we are particularly interested in single TTWA firms is that for such firms there is no issues in, for example, attributing their productivity and their employment to a particular TTWA. By contrast, for multi TTWA firms (Tesco for example) it is less clear how to allocate productivity (which can only be measured at the level of the firm) to the different TTWAs in which the firm has her establishments. Single TTWA firms represents the vast majority of firms (around 97%) and account for about 43% of overall employment in our dataset.

3. French data

3.1 Data sources

3.1.1 FICUS

FICUS is an administrative database containing detailed accounting information (employment, sales, intermediates, capital, industry affiliation, etc.) for the population of French firms. The database is part of the SUSE (Système unifié de statistiques d'entreprises) framework. SUSE constitutes a coherent set of statistical data on firms obtained from the joint use of two sources of information: tax declarations of companies to the General Directorate of Taxes (DGI); the annual business surveys (EAE).

FICUS includes both balance sheet and profits and losses account information and, for the purpose of our analysis, we use information from the year 2000 till the year 2007, when the dataset stops because of being replace by the companion database FARE (see below). Each firm in the dataset is uniquely identified by a 9-digit code (SIREN code).

3.1.2 FARE

FARE is an administrative database containing detailed accounting information (employment, sales, intermediates, capital, industry affiliation, etc.) for the population of French firms. The database is part of the ESANE (Élaboration des statistiques annuelles d'entreprises) framework. The ESANE framework succeeds to the previous framework (SUSE) and, since 2008, this new system has jointly exploited, via a specific estimation procedure, administrative data and data from the ESA and EAP surveys in order to produce the most accurate sectoral statistics possible.

FARE includes both balance sheet and profits and losses account information and, for the purpose of our analysis, we use information from the year 2008 till the year 2017. Each firm in the dataset is uniquely identified by a 9-digit code (SIREN code).

3.1.3 Stocks d'Établissements

The Stocks d'Établissements database is a demography product of establishments providing identity data on the characteristics of establishments. It includes establishments active on 31 December of year N. The date are compiled from the Directory of Companies and Establishments (REE).

The Stocks d'Établissements database contains information on the location of establishments (up to the municipality level) as well as on their employment. Each establishment in the dataset is uniquely identified by a 14-digit code (SIRET code) which can be uniquely attached to a SIREN (firm) code.

3.2 *Cleaning and combining the data*

In what follows we explain how we cleaned and merged the data.

3.2.1 *Data cleaning*

FICUS and FARE. For both FICUS and FARE we apply the following cleaning to the data. First, we discard observations without information on the municipality where the firm is located and/or with missing SIREN code or industry affiliation. Second, we operate some cleaning to the business start year and the number of employees variables. Third, we use a correspondence table between the NACE rev1 and NACE rev2 to consistently obtain information on the NACE rev2 affiliation of firms for the whole period 2000-2017. Finally, we eliminate firms involved in financial and insurance activities (NACE rev2 codes 64, 65 and 66) and restrict the sample to firms with at least one employee.

Stocks d'Établissements. Considering the Stocks d'Établissements dataset, we simply discard observations with missing SIRET and/or municipality code.

3.2.2 *Data matching*

Data matching is quite straightforward with French firm data because of the unique firm-identifier (SIREN code). We thus simply append the information for each year, coming from either FICUS or FARE, thus obtaining a firm panel data over the period 2000-2017. Finally, we define industries as two digit NACE rev2 codes and apply some grouping (detailed below) in preparation for TFP estimations.

3.2.3 *Adding information on location*

Each establishment in the Stocks d'Établissements database is identified by a unique 14-digit code (SIRET code) whose first 9 digits correspond to the SIREN code of the firm. This greatly facilitates the task of adding location information.

In view of conducting a meaningful spatial analysis, we use an 'economic' partition of the French geography and in particular the 2010 version of the Zone d'Emplois (ZEs). In our analysis we do not consider French overseas territories (DOMs) as well as Corsica. This leaves 297 areas for continental France. In order to go from municipalities to ZEs for the whole of our sample period we use a correspondence table provided by the INSEE. The match between municipalities and ZEs works quite well and only requires minor adjustments.

Equipped with the information above we are thus able to identify, as in the case of the UK, single ZE firms, i.e, firms that we can uniquely associate to one ZE. Such firms may thus have more than one establishment, but such establishments need to be located in the same ZE. As stated above, the reason we are particularly interested in single ZE firms is that for such firms there is no issues in, for example, attributing their productivity and their employment to a particular ZE. By contrast, for multi ZE firms (Carrefour for example) it is less clear how to allocate productivity (which can only be measured at the level of the firm) to the different ZEs in which the firm has her establishments. Single ZE firms represents the vast majority of firms (around 93%) and account for about 53% of overall employment in our dataset.

4. Productivity and markups estimation

In order to estimate productivity and markups we use a production function approach. For the UK we use sales from the VAT data as a measure of output/revenue, purchases from the VAT data as a measure of intermediates expenditure, tangible fixed assets from FAME as a measure of the capital stock, and employment (count of employees plus the owner(s)) from the BSD as a measure of the labour input. For France we use firm turnover as a measure of output/revenue, purchases of goods and services as a measure of intermediates expenditure, tangible fixed assets as a measure of the capital stock, and employment (count of employees) as a measure of the labour input.

First we deflate revenue, intermediates and capital using corresponding indexes provided by the ONS (for the UK) and the INSEE (for France) with the base year being 2017.[2] Second, we apply some trimming to the data. More specifically, we discard observations where the value of intermediates is higher than the value of sales and further apply a bottom and top trimming of 0.5% (by industry) based on the ratios of: i) intermediates to sales; ii) capital to labour; iii) revenue to labour.[3] Third, we use a second-order polynomial in intermediates, capital and labour to smooth revenue and purge it from measurement error as suggested in De Loecker et al. (2016) and Forlani et al. (2023) among others.

[2]For France we use the same industry deflators (base year 2017) for revenue and intermediates while using specific deflators for capital. For the UK we employ a double deflation method to deflate all monetary variables to constant 2017 prices. We construct output, intermediates and capital deflators from series provided by the ONS. For the output price deflators, we use the ONS, 'Experimental Industry Level Deflators'. These are available at the 2-, 3- and 4-digit Industry level. They are produced by aggregating industry product deflators based on their use of products in line with the National Accounts supple-use framework. Where the deflators are not available at the 2-digit level, we average the 3- or 4-digit deflators to construct them. To construct deflators for intermediate inputs we make use of the Supply and Use Tables (SUTs) produced by the ONS. The SUTs for 1997 to 2020 are consistent with the UK National Accounts in Blue Book 2022. We use the industries' intermediate consumption values in 2010 to create weights by dividing each 2-digit industry's demand by the total demand for each 2-digit industry. This generates a Leontief matrix which we use as weights to derive input deflators for each 2-digit industry from the output deflators. Capital stock deflators are constructed from the ONS series of annual estimates of net capital stocks and consumption of fixed capital in the UK which is provided by asset and sector. These are available in both current prices and chained volume measures. We construct the deflators at the 2-digit Industry level. Since tangible fixed assets from FAME are provided as a net value, we obtain the corresponding deflator by dividing the current prices by the chained volume measures of net capital stock at the 2-digit Industry level.

[3]Post-TFP estimations we also discard those (very few) observations with markups below 0.6 and above 20.

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. Considering the log production function we thus have:

$$q_{it} = \alpha_L l_{it} + \alpha_M m_{it} + \alpha_K k_{it} + a_{it}, \tag{1}$$

where small case letters indicate logs (for example $k_{it} = \log K_{it}$). In line with the productivity literature, we assume that the TFP process is driven by an autoregressive component:

$$a_{it} = \phi_a \; a_{it-1} + \nu_{ait}, \tag{2}$$

where ν_{ait} denotes productivity shocks that represent innovations with respect to the information set of the firm in $t-1$ and are iid across firms and time.

In line with the literature, we assume capital k_{it} to be predetermined in the short-run, i.e., the current capital level has been chosen in t-1 and cannot immediately adjust to current period shocks ν_{ait}.[4] We further assume, as standard in the literature, that intermediates m_{it} are a variable input free of adjustment costs. This means that intermediates can be optimally chosen in t based on, among others, the particular realization of ν_{ait}. In this respect, we will see later on that intermediates being fully adjustable in the short-run allows for a simple rule to pin-down the markup of firm i. Concerning labor, we assume it to be a semi-flexible input meaning that it can, to some extent, adjust to current shocks in t but not to the optimal cost-minimizing level determined only by wages and marginal productivity.[5]

At time t firms have already chosen capital and labor and so these inputs are considered as given in their decision process along with the cost of intermediates W_{Mit}. At the same time, productivity a_{it} becomes known at time t. We assume firms in t use the above information and constraints to choose intermediates in order to minimize production costs and choose quantity or price (depending upon the features of competition) in order to maximize profits. In this respect, as first highlighted in Hall (1986) and further implemented in De Loecker

[4]Intuitively, the restriction behind this assumption is that it takes a full period for new capital to be ordered, delivered, and installed. Note this means that k_{it} is uncorrelated with current period shocks ν_{ait}. However, this does not mean that k_{it} is uncorrelated with the current productivity *level* a_{it}. For example, investment decisions in $t-1$ are likely to be determined by both the level of capital in $t-1$ and the level of productivity in $t-1$. In this light, k_{it} should be correlated with a_{it-1} and so with a_{it}. See Ackerberg et al. (2015) for more details.

[5]In sum, l_{it} should be correlated (like intermediates m_{it}) with shocks ν_{ait} but the amount of labour in t does not simply reflect wages and marginal productivity implying that it cannot be used to recover markups. As far as the timing is concerned, we assume l_{it} is chosen by firm i at time $t-b$ ($0 < b < 1$), after k_{it} being chosen in $t-1$ but prior to m_{it} being chosen in t.

and Warzynski (2012), De Loecker et al. (2016) and Forlani et al. (2023) among others, cost-minimization of a variable input free of adjustment costs provides a simple rule to pin down markups. The marginal cost is:

$$\frac{\partial C_{it}}{\partial Q_{it}} = \frac{\partial C_{it}}{\partial M_{it}}\frac{\partial M_{it}}{\partial Q_{it}} = W_{Mit}\frac{\partial M_{it}}{\partial Q_{it}}.$$

Now define the markup as:

$$\mu_{it} \equiv \frac{P_{it}}{\frac{\partial C_{it}}{\partial Q_{it}}}.$$

We thus have:

$$\frac{P_{it}}{\mu_{it}} = W_{Mit}\frac{\partial M_{it}}{\partial Q_{it}}.$$

Multiplying by Q_{it} and dividing by M_{it} on both sides implies that:

$$\frac{P_{it}Q_{it}}{M_{it}\mu_{it}} = \frac{R_{it}}{M_{it}\mu_{it}} = W_{Mit}\frac{\partial M_{it}}{\partial Q_{it}}\frac{Q_{it}}{M_{it}} = W_{Mit}\frac{\partial m_{it}}{\partial q_{it}}.$$

Re-arranging we finally have:

$$\mu_{it} = \frac{\frac{\partial q_{it}}{\partial m_{it}}}{\frac{W_{Mit}M_{it}}{R_{it}}} = \frac{\frac{\partial q_{it}}{\partial m_{it}}}{s_{Mit}}.$$

This simple rule to pin-down markups is consistent with many hypotheses on product market structure (monopolistic competition, monopoly and standard forms of oligopoly) and consists in taking the ratio of the output elasticity of intermediates ($\frac{\partial q_{it}}{\partial m_{it}}$) to the share of intermediates in revenue ($s_{Mit} \equiv \frac{W_{Mit}M_{it}}{R_{it}}$). Considering our production function (1) we simply have:

$$\mu_{it} = \frac{\alpha_M}{s_{Mit}}. \tag{3}$$

Therefore, provided estimates of the parameters of the production function (1), and in particular of α_M, as well as data on intermediates expenditure and revenue, one can simply compute the firm-specific markup μ_{it} using (3).

In terms of estimating the parameters of the production function (1) we use the intuition developed in Wooldridge (2009), i.e, we: i) substitute for a_{it} in equation (1) using (2); ii) substitute for a_{it-1} using a polynomial in k_{it-1} and m_{it-1}; iii) in the final augmented production function equation we do not instrument capital k_{it} but instrument labour and intermediates l_{it} and m_{it} with time lags.[6] We estimate the parameters of the production function separately for each industry while adding as controls a battery of time dummies (and information on foreign ownership for the UK). Standard errors are clustered at the firm-level. Last but not least, we also perform simple OLS estimations of the production function (1) to provide robustness.

[6] We use l_{it-1}, l_{it-2}, m_{it-2} and k_{it-2}.

5. Some data highlights

Tables A-1 to A-4 in the Appendix provide (for both the UK and France) estimates of the parameters of the production function (for each industry) obtained with our instrumental variables approach à la Wooldridge (2009) (we label such estimates WLD). Inspection of Tables A-1 to A-4 reveals that coefficients are quite precisely estimated and have the expected magnitude for a 3-inputs production function, namely an elasticity of intermediates around 0.7-0.8, an elasticity of labour around 0.2 and an elasticity of capital around 0.02-0.05. Furthermore, the under-identification tests and the weak identification F-statistics clearly indicate that our instruments are strong.

Table 2: UK Data: 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).

Table 3: French Data: 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.

In order to get a taste of the size and coverage of our dataset we provide in Tables 2 and 3 below some key summary stats across all years. Considering the UK, our dataset spans over 9,954,131 observations across the time frame 2004-2017. The average firm has a 4.3 million pounds revenue,[7] a 3.2 million value of intermediates, a 2.4 million capital stock and 22 workers. Standard deviation values are almost 2 orders of magnitudes higher than mean values indicating that our data covers both very small and very large firms. This is confirmed by looking at the 5th and 95th percentiles. Considering capital, for example, the firm in the 5th percentile has a capital stock of just over one thousand pounds while the 95th percentile firm has a capital stock of about 664 thousand pounds, which is still below the mean of 2.4 million pounds with the latter being driven up by the presence of a few very big firms. As for France, our dataset spans over 17,641,530 observations across the time frame 2000-2017. The average firm has a 3 million euros revenue, a 2.1 million value of intermediates, a 1.8 million capital stock and 13 workers. Again, the data covers both very small and very large firms. Considering capital, for example, the firm in the 5th percentile has a capital stock of just over six thousand euros while the 95th percentile firm has a capital stock of about 1.7 million euros, which is still below the mean of 1.8 million euros with the latter being driven up by the presence of a few very big firms.

Tables 4 and 5 below provides a breakdown of the number of firms (and the related overall employment) in our two datasets by year. For the UK the number of firms rises from 642,748 in 2004 to 814,407 in 2017. Overall employment covered by our dataset is in between 14 and 17 million.[8] Considering France, the number of firms varies by year with a maximum of 1,144,423 in 2007 and a minimum of 871,200 in 2013. Overall employment covered by our dataset is rather stable across years and above 12 million. Tables A-5 and A-6 in the Appendix contain instead an industry breakdown of the number of firms and related employment for the year 2017.

Finally, Tables 6 and 7 deliver average (employment weighted) apparent labour productivity, labour productivity, OLS TFP, WLD TFP and markups by year. Considering the UK, Table 6 indicates that, while total factor productivity (both OLS TFP and WLD TFP) has only been both very lightly and very briefly affected by the 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 like Harris and Moffat (2017) and Jacob and Mion (2020b). Inspection of markups reveals that they recovered their

[7] We have checked the correlation between revenue from VAT data and revenue coming from the BSD and Fame. The correlation between revenue from VAT and revenue from the BSD is quite poor standing at about 0.4. Visual inspection reveals that turnover from the BSD is frequently made up of round numbers that are often not updated across time within a firm. The correlation between turnover from VAT and turnover for FAME (the latter being available for the medium and large firms only) is better and around 0.5 to 0.6. Again, there seems to be a good amount of rounding in FAME turnover figures.

[8] Firms involved in financial and insurance activities are excluded from our dataset and account for around 1.2 million workers as reported by ONS official figures.

Table 4: UK Data: 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.

pre-financial crisis level around 2015[9] while for labour productivity the recovery year is 2016.[10] In this respect, Tables A-8 and A-9 in the Appendix show that results are similar if we split the sample into single-TTWA firms (essentially small and medium firms) and multi-TTWA firms (essentially large firms) with the recovery being stronger for multi-TTWA firms. For single-TTWA firms labour productivity in 2017 is still below its pre-financial crisis level.

Moving to France, Table 7 suggests that 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

[9]Our findings for markups are not incompatible with those obtained by Black (2022). There are, first of all, some obvious differences in the two analyses. For example, Black (2022) uses the ABS/ABI, as compared to our more comprehensive database, and so our results also speak about those medium and small firms which are missing from the ABS/ABI. On the other hand, Black (2022) embraces a longer time span (1997-2019) in which the financial crisis episode is overall dwarfed by time trends. Finally, some of the markups measures developed in Black (2022) also point to a fall in markups, followed by a recovery, around the financial crisis.

[10]Table A-7 in the Appendix provides complementary information for the UK about the evolution of apparent labour productivity and labour productivity. More specifically, in Table A-7 we still provide employment-weighted apparent labour productivity and labour productivity but rather than deflating current values we simply use those current values to compute averages.

Table 5: French Data: 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.

crisis.[11] Inspection of markups reveals that they have not yet recovered their pre-financial crisis level suggesting that firms struggle to achieve pre-financial crisis profit margins. Results are similar if we split the sample into single-TTWA firms and multi-TTWA firms (Tables A-10 and A-11 in the Appendix) with the recovery being stronger for multi-TTWA firms. For multi-TTWA firms, total factor productivity has definitely picked up its pre-financial crisis level.

[11] Interestingly, when considering apparent labour productivity and labour productivity (the most comparable productivity measures between the two countries) while taking the last year of the data, the UK does not appear to be much less productive than France, which is contrast with some macro comparisons suggesting the UK is behind France in terms of productivity. This is of course limited to our data and so to those more structured firms employing at least one worker.

Table 6: UK Data. 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

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. Revenue, intermediates and capital have been deflated using indexes provided by the ONS with the base year being 2017. Apparent labour productivity is computed as firm revenue (in 2017 pounds) over firm employment. Labour productivity is computed as firm value added (in 2017 pounds) over firm employment. OLS TFP is firm-level total factor productivity obtained from a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by SIC industry) using the OLS estimator. WLD TFP is firm-level total factor productivity obtained from a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by SIC industry) using a methodology consistent with Wooldridge (2009). Markups are estimates of the firm-level price to marginal cost ratio and are obtained from WLD TFP estimations and the share of intermediates in revenue as developed in De Loecker and Warzynski (2012). All firm-level variables have been aggregated using firm employment as weight.

Table 7: French Data. 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

Notes: Employment is number of employees. Revenue, intermediates and capital have been deflated using indexes provided by the INSEE with the base year being 2017. Apparent labour productivity is computed as firm revenue (in 2017 euros) over firm employment. Labour productivity is computed as firm value added (in 2017 euros) over firm employment. OLS TFP is firm-level total factor productivity obtained from a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by NACE rev2 industry) using the OLS estimator. WLD TFP is firm-level total factor productivity obtained from a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by NACE rev2 industry) using a methodology consistent with Wooldridge (2009). Markups are estimates of the firm-level price to marginal cost ratio and are obtained from WLD TFP estimations and the share of intermediates in revenue as developed in De Loecker and Warzynski (2012). All firm-level variables have been aggregated using firm employment as weight.

6. Regional productivity difference: conceptual framework

A stylized fact of economic geography is that the productivity of firms increases with city size and urban density (Combes and Gobillon, 2015), and a large literature going back to Marshall (1890) explores the question of why cities have this productivity advantage. Micro-foundations put forward for these agglomeration externalities are typically grouped under the headings sharing, matching, learning and sorting (Duranton and Puga, 2004, Combes et al., 2008) and include different forms of knowledge spillovers between firms, costly trade, pro-competitive effects of city size, and sorting of workers (Syverson, 2011). The empirical literature suggests a rather consistent, across countries and years, range for the elasticity of productivity with respect to city size. 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.[12] 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).

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 regions. Crucially, the two approaches do not seem to provide the same magnitudes regarding, for example, the elasticity of productivity with respect to local density. More specifically, Jacob and Mion (2020a) provide evidence for French manufacturing firms highlighting the importance of weighting in going from the firm-level (micro) to the regional-level (macro) productivity. They find smaller values for the elasticity of productivity with respect to population density when using unweighted firm-level regressions while getting quite larger values when considering revenue- or employment-weighted firm-level regressions.

The productivity of a region is clearly the productivity of its firms. However, the aggregate productivity of a region is the weighted average (typically by employment) of the productivity of the firms located in the region and not the simple average. When running unweighted regressions, in which firms are the unit of analysis, to measure productivity differences across space one is essentially comparing the average firm across different locations irrespective of the firm size distribution, and its link to productivity, within regions. The link between micro and macro is restore if one runs weighted regressions (es explained better below) and the coefficients from the unweighted and weighted regressions do not need to be the same. One reasons they could differ is a varying (across regions) correlation between firm size and firm productivity. For example, if denser regions are characterized by a higher correlation between firm size and firm productivity, unweighted differences in productivity across space

[12] See also Combes et al. (2008), Mion and Naticchioni (2009) and De La Roca and Puga (2017) for estimates of the elasticity of worker-level wages with respect to density.

will be magnified when weighting. Another reason for differences between coefficients is the heterogeneity (along the productivity dimension) of the elasticity of productivity with respect to density. For example, if more productive firms enjoy disproportionately from the density of economic activities, unweighted differences in productivity across space will be again magnified when weighting because (on average) more productive firms are larger.

In what follows we extend the analysis of Jacob and Mion (2020a) beyond manufacturing to the whole private sector for both France and the UK while digging into the above mentioned explanations for the larger values of the elasticity of productivity with respect to population density when using weighted firm-level regressions as compared to unweighted firm-level regressions. We are interested in the variation of TFP across regions and how it is affected by aggregation/weighting. The baseline estimation equation is:

$$\bar{a}_{it} = \gamma density_{r(it)} + I_{r(it)} + I_t + \epsilon_{it}, \tag{4}$$

where

- $\bar{a}_{it}$ is (log) WLD 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,[13]
- $I_{r(it)}$ and I_t are macro region and year dummies
- ϵ_{it} is an error term.

We perform both un-weighted and weighted (by the share of employment of firm i in total region r employment at time t) OLS estimations of equation (4) and cluster standard errors at the region-year (ZE for France and TTWA for the UK) 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). More formally, consider first aggregating firm productivity $\bar{a}_{it}$ (using our weights) at the region r and year t level and then regressing this average region-year productivity on density, as well as macro region and year dummies, while using robust standard errors. The resulting OLS coefficients and standard errors of this 'aggregate' regression will be, by the properties of OLS, identical to those obtained from weighted OLS estimations of equation (4) at the firm-level with region-year clustering of the standard errors. In this light, unweighted and weighted regressions of equation (4) allow to go from the micro firm-level to the macro regional-level. At the same time, the R2 of the weighted regression of equation (4) and the aggregate regression will be different because in the latter heterogeneity in productivity across firms within a region-year has been eliminated.

[13]We use population density by TTWA in 2015 for the UK and population density by ZE in 2009 for France.

7. Results

In what follows we focus on the samples of single TTWA firms (for the UK) and single ZE firms (for France), i.e, firms that we can uniquely associate to one region. Such firms may thus have more than one establishment, but such establishments need to be located in the same TTWA/ZE. The reason we are particularly interested in single TTWA/ZE firms is that for such firms there is no issues in, for example, attributing their productivity and their employment to a particular region. By contrast, for multi TTWA/ZE firms it is less clear how to allocate productivity (which can only be measured at the level of the firm) to the different regions in which the firm has her establishments. Towards the end of this Section we provide some robustness results including multi TTWA/ZE firms in the analysis, while attributing the same productivity to all of the establishments of a given multi TTWA/ZE firm and using establishment-level employment for weighting. Such robustness results largely confirm our finding based on single TTWA/ZE firms.

Table 8 provides estimates of equation (4) for the UK. The first column contains unweighted estimates while the second column delivers weighted results. The weighted γ is around 2.1% and so in line with the literature while the unweighted coefficient of density stands at about 1.8%. At the same time, macro region dummies are (the reference category being London) all negative and strongly significant while being typically larger in magnitude in the case of weighted regressions. These results suggest some amplification of unweighted differences in productivity across space when considering weighting.[14] For example, our 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. Furthermore, both unweighted and weighted estimates point to a substantial productivity gap, over and beyond what can be explained by density, of all UK regions with respect to London.

As suggested above 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; 2) A productivity return on density being stronger for the most productive firms. In order to analyse the first hypothesis we compute the correlation between firm size and productivity for each of the 228 TTWAs in the UK and plot in Figure 1 these correlations against the density of the region. Inspection of Figure 1 reveals that: 1) The correlation between firm size and productivity within a region is quite low and sometimes negative across UK TTWAs; 2) The relationship between these correlations and region density is not positive and actually slightly negative (red regression line in the Figure). These findings

[14] At the bottom of the second column of Table 8 we report the R-squared of the equivalent, to our weighted firm-level regression, 'aggregate' regression. Such R-squared is much higher than the one based on the weighted firm-level regression (0.2613 vs. 0.0071) and in line with previous findings in the literature (Combes and Gobillon, 2015).

Table 8: UK: Density regressions

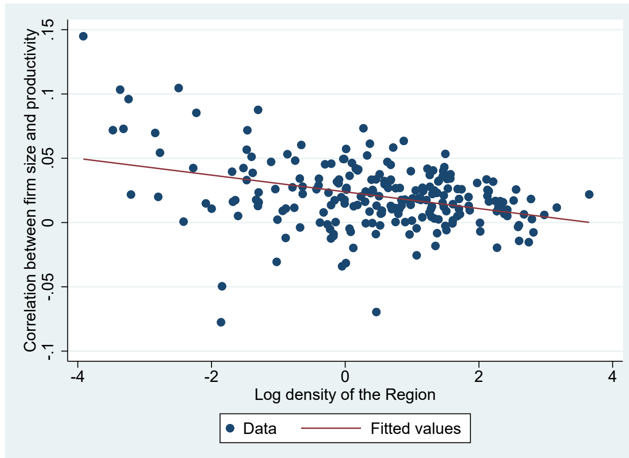
	Unweighted	Weighted
log density	0.0178***	0.0208***
	(0.0007)	(0.0009)
	Reference category is London	
East Midlands	-0.0676***	-0.1004***
	(0.0027)	(0.0072)
East of England	-0.0421***	-0.1056***
	(0.0031)	(0.0082)
North East	-0.0765***	-0.1021***
	(0.0030)	(0.0084)
North West	-0.0767***	-0.1252***
	(0.0029)	(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

Notes: The dependent variable is firm (log) WLD TFP demeaned by the corresponding industry average. Log density is the 2015 population density of the TTWA r where firm i is located at time t. Year dummies (not reported) and macro region dummies (London being the reference category) are included in the regressions. Column one provides simple OLS regressions of equation (4) while column two shows weighted OLS regressions of equation (4) where the weight is the share of the employment of firm i and time t in overall regional employment in year t. R-squared 'aggregate' refers to the R-squared of the equivalent, to the weighted firm-level regression in column two, 'aggregate' regression at the region-year level. Standard errors are clustered by TTWA-year. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

indicate that, if anything, the varying correlation between firm size and productivity across regions should *reduce* and not amplify spatial productivity differences when going from the micro to the macro.

With the aim of exploring room for the second explanation we report in Table 9 the results of quantile estimations (for each decile of the distribution of productivity) of equation (4) while focusing on the coefficient of density. Table 9 does indicate that the productivity return of density is increasing along the deciles of the productivity distribution ranging from 1.1% for the first decile to 3.5% in the 9th decile. This finding is reminiscent of the 'dilating' of the productivity distribution in larger regions found for France in Combes et al. (2012) and

Figure 1: UK: Correlation between firm productivity and size within each region



Notes: The Figure provides a scatter plot of two variables for each TTWA. The variable on the y-axis in the correlation (across firms and years within a region) between firm employment and (log) WLD TFP demeaned by the corresponding industry average. The variable on the x-axis is the 2015 population density of the TTWA r. The red line indicates the regression line.

it is such an increasing productivity return of density that magnifies firm-level productivity differences across the UK space, when going from the micro to the macro, and not a varying correlation between firm size and productivity across space.

Table 9: 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)

Notes: The dependent variable is firm (log) WLD TFP demeaned by the corresponding industry average. Log density is the 2015 population density of the TTWA r where firm i is located at time t. Year dummies and macro region dummies (not reported) are included in the regressions. Columns one to nine provides quantile regressions of equation (4) on deciles one to nine. Robust standard errors provided. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Moving forward Table 10 provides the equivalent information of Table 8 for France. As can be noticed, the weighted γ is around 2% and so also in line with the literature while the unweighted coefficient of density is much smaller standing at about 0.4%. At the same time, macro region dummies are (the reference category being Île-de-France, i.e., Paris) all negative but small in magnitude and often not significant. On the one hand these results indicate, contrary to the UK, the absence of a strong productivity gap (over and beyond what can be attributed to density) between the core region of France and the rest of the country. On the other hand, they also suggest a stronger (for France compared to the UK) amplification of unweighted differences in productivity across space when considering weighting. For example, our 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 only about 36% of the aggregate difference.

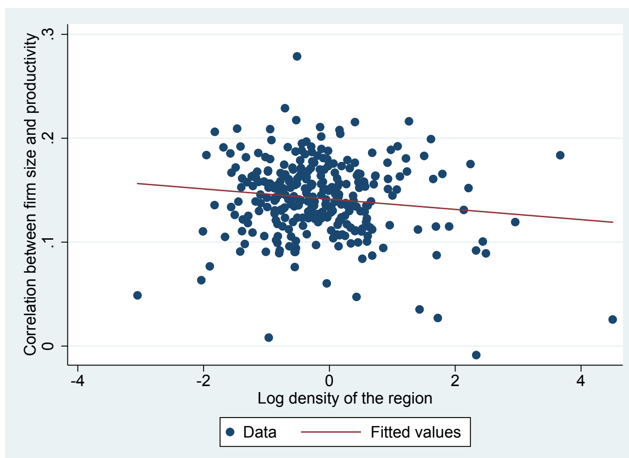
Table 10: France: Density regressions

	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

Notes: The dependent variable is firm (log) WLD TFP demeaned by the corresponding industry average. Log density is the 2009 population density of the ZE r where firm i is located at time t. Year dummies (not reported) and macro region dummies (Paris being the reference category) are included in the regressions. Column one provides simple OLS regressions of equation (4) while column two shows weighted OLS regressions of equation (4) where the weight is the share of the employment of firm i and time t in overall regional employment in year t. R-squared 'aggregate' refers to the R-squared of the equivalent, to the weighted firm-level regression in column two, 'aggregate' regression at the region-year level. Standard errors are clustered by ZE-year. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Figure 2 further qualifies our results by showing (as in the case of the UK) that the relationship between the correlation of firm size and productivity within a region and region density is not positive and, if anything, slightly negative (red regression line in the Figure). Though, a comparison of Figures 1 and 2 also reveals that the correlation between firm size and productivity is more on the positive side across French ZEs as compared to UK TTWAs and this extends to the nationwide correlation between firm employment and productivity standing at 0.0243 for France (0.0712 when using our weights instead of firm employment) and 0.0122 for the UK (-0.0046 when using our weights instead of firm employment). Furthermore, Table 11 shows the results of quantile estimations (for each decile of the distribution of productivity) of equation (4) for France while focusing on the coefficient of density. As in the case of the UK, the productivity return of density is increasing along the deciles of the productivity distribution and this is the key driver of the magnification of productivity differences across space when going from the micro to the macro. Contrary to the UK though, the productivity return of density is actually negative for the first few deciles of the productivity distribution signalling an issue that France seems to have about large agglomeration and low productive firms. Interestingly, this finding has no counterpart in the productivity distribution analysis of Combes et al. (2012).

Figure 2: France: Correlation between firm productivity and size within each region



Notes: The Figure provides a scatter plot of two variables for each ZE. The variable on the y-axis in the correlation (across firms and years within a region) between firm employment and (log) WLD TFP demeaned by the corresponding industry average. The variable on the x-axis is the 2009 population density of the ZE r. The red line indicates the regression line.

Key observations. Some of the key highlights of our results so far can be summarised as follows. 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 instead the negative productivity return of density for the least productive firms, i.e., denser places in France nurture too many low productive firms and this creates a big divide between the un-weighted and weighted productivity return of density.

Table 11: 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)

Notes: The dependent variable is firm (log) WLD TFP demeaned by the corresponding industry average. Log density is the 2009 population density of the ZE r where firm i is located at time t. Year dummies and macro region dummies (not reported) are included in the regressions. Columns one to nine provides quantile regressions of equation (4) on deciles one to nine. Robust standard errors provided. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Robustness using both single-region and multi-region firms. As anticipated above, we have produced equivalent results to those presented in this Section for the full sample of single-region and multi-region firms. In doing so we exploit information on the different establishments of a firm and allocate the same productivity to all of the establishments of a firm while using establishment-level employment to weigh our regressions. Indeed, the unit of the analysis in these robustness exercises, where density regressions are presented in Tables A-12 and A-13, switches from the firm to the establishment. Moving to our findings Figures A-1 and A-2 in the Appendix convey a very similar message as Figures 1 and 2, i.e., that the relationship between the correlation of establishment size and productivity within a region and region density is not positive and, if anything, slightly negative. At the same time, French ZEs are characterised by a more positive correlation between establishment size and productivity than UK TTWAs. Furthermore, Tables A-14 and A-15 in the Appendix indicate that the productivity return of density is (overall) increasing along the establishment productivity deciles. Still, France features some negative productivity returns on density for the first deciles as in the case of Table 11.

8. Conclusions

We propose a new data resource that attempts to overcome limitations of standard firm-level datasets for the UK (like the ARD/ABS) by building on administrative data covering the population of UK firms. More specifically, we merge the BSD, VAT and FAME datasets and create a common firm definition encompassing the different firm identifiers used in the three datasets. 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. We also construct a similar dataset for France and use both datasets to: 1) Provide some highlights of the data and an overall picture of the evolution of aggregate UK and French productivity and markups: 2) Analyse the spatial distribution of productivity in both countries at a very fine level of detail – 228 Travel to Work Areas (TTWAs) for the UK and 297 Zones d'emploi (ZEs) for France – while focusing on the role of economic density.

Considering the UK, while total factor productivity has only been both very lightly and very briefly affected by the financial crisis, the same is not true for markups, apparent labour productivity and labour productivity. Inspection of markups reveals that they recovered their pre-financial crisis level around 2015 while for labour productivity the recovery year is 2016. As for France, it is not entirely clear whether total factor productivity has by 2017 picked up its pre-financial crisis level. 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 French firms struggle to achieve pre-financial crisis profit margins.

In terms of spatial analysis we obtain the following results. First, for both France and the UK we find a larger productivity return to density when weighting observations by employment as compared to unweighted regressions. Digging deeper into this reveals, in both cases, that: 1) The correlation between firm size and productivity within a region is quite low (and sometimes negative) across regions particularly for the UK; 2) The relationship between these correlations and region density is not positive and actually slightly negative. These findings indicate that, if anything, the varying correlation between firm size and productivity across regions should *reduce* and not amplify spatial productivity differences when going from the micro to the macro. On the other hand, in both cases, we find evidence that the productivity return of density is increasing along the deciles of the productivity distribution. This finding is reminiscent of the 'dilating' of the productivity distribution in larger regions found for France in Combes et al. (2012) and it is such an increasing productivity return of density that magnifies firm-level productivity differences for the UK and France, when going from the micro to the macro, and not a varying correlation between firm size and productivity across space.

In terms of the comparison between the UK and France we document a number of striking differences. 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 instead the negative productivity return of density for the least productive firms, i.e., denser places in France nurture too many low productive firms and this creates a big divide between the un-weighted and weighted productivity return of density.

In terms of directions for future research we look forward to seeing more studies, especially studies covering countries other than France and the UK, tackling the issue of aggregation in measuring the return to density. Indeed, despite a number of common features between France and the UK, our analysis also reveals important differences so highlighting the importance of country specificities.

References

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

Black, R. (2022). Estimates of markups, market power and business dynamism from the Annual Business Survey, Great Britain: 1997 to 2019. ONS paper.

Ciccone, A. (2002). Agglomeration effects in Europe. *European Economic Review*, 46(2):213–227.

Ciccone, A. and Hall, R. E. (1996). Productivity and the Density of Economic Activity. *The American Economic Review*, 86(1):54–70.

Combes, B. P.-p., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012). The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection. *Econometrica*, 80(6):2543–2594.

Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2):723–742.

Combes, P.-P. and Gobillon, L. (2015). The Empirics of Agglomeration Economies. In *Handbook of Regional and Urban Economics*, chapter 5, pages 247–348. Elsevier.

De La Roca, J. and Puga, D. (2017). Learning by working in big cities. *The Review of Economic Studies*, 84(1):106–142.

De Loecker, J., Goldberg, P. K., Khandelwal, A. K., and Pavcnik, N. (2016). Prices, markups, and trade reform. *Econometrica*, 84(2):445–510.

De Loecker, J. and Warzynski, F. (2012). Markups and Firm-Level Export Status. *American Economic Review*, 102(6):2437–71.

Duranton, G. and Puga, D. (2004). Micro-Foundations of Urban Agglomeration Economies. In *Handbook of Regional and Urban Economics*, volume 4, chapter 48, pages 2063–2117. Elsevier.

Forlani, E., Martin, R., Mion, G., and Muûls, M. (2023). Unraveling firms: Demand, productivity and markups heterogeneity. *The Economic Journal*, 133(654):2251–2302.

Hall, R. E. (1986). Market Structure and Macroeconomic Fluctuations. *Brookings Papers on Economic Activity*, 2.

Harris, R. and Moffat, J. (2017). The UK productivity puzzle, 2008 to 2012: evidence using plant-level estimates of total factor productivity. *Oxford Economic Papers*, 69(3):529–549.

Jacob, N. and Mion, G. (2020a). On the productivity advantage of cities. CEPR Discussion Paper 14644.

Jacob, N. and Mion, G. (2020b). The UK's Great Demand and Supply Recession. CEPR Discussion Paper 15516.

Marshall, A. (1890). *The Principles of Economics*. McMaster University Archive for the History of Economic Thought.

Mion, G. and Naticchioni, P. (2009). The spatial sorting and matching of skills and firms. *Canadian Journal of Economics/Revue canadienne d'économique*, 42(1):28–55.

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the Nature and Sources of Agglomeration Economies. In *Handbook of Regional and Urban Economics*, volume 4, chapter 49, pages 2119–2171. Elsevier.

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature*, 49(2):326–365.

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters*, 104(3):112–114.

Appendix: Additional Tables and Figures

Table A-1: UK Data. WLD TFP production function estimations by industry: industries 01 to 59.

VARIABLES	Industry 01	Industry 0X	Industry 0Y	Industry 13	Industry 16	Industry 17	Industry 18	Industry 1X	Industry 1Y
log intermediates	0.8317***	0.4732***	0.8027***	0.7580***	0.7609***	0.7949***	0.7634***	0.8435***	0.8065***
	(0.0026)	(0.0172)	(0.0678)	(0.0111)	(0.0053)	(0.0076)	(0.0057)	(0.0059)	(0.0213)
log employment	0.1535***	0.2506***	0.1393***	0.2176***	0.2154***	0.1726***	0.2162***	0.1438***	0.1847***
	(0.0010)	(0.0073)	(0.0145)	(0.0036)	(0.0017)	(0.0024)	(0.0017)	(0.0015)	(0.0040)
log capital	-0.0094***	0.0575***	0.0154	0.0214***	0.0100***	0.0211***	0.0198***	0.0170***	0.0070
	(0.0012)	(0.0071)	(0.0165)	(0.0019)	(0.0013)	(0.0016)	(0.0013)	(0.0015)	(0.0043)
Observations	157,181	15,816	4,629	20,268	37,806	14,727	61,529	40,829	16,959
R-squared	0.9968	0.9605	0.9909	0.9980	0.9986	0.9994	0.9975	0.9992	0.9935
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
KP rk LM stat under-identif.	1411	205.7	26.83	70.33	238.7	98.44	406.2	196.2	31.62
Under-identif. p-value	0	0	6.39e-06	0	0	0	0	0	6.30e-07
KP rk Wald F stat weak identif.	716.4	113.5	8.871	26.87	97.07	32.18	176.6	77.66	11.94

VARIABLES	Industry 22	Industry 23	Industry 24	Industry 25	Industry 26	Industry 27	Industry 28	Industry 29	Industry 2X
log intermediates	0.7900***	0.7600***	0.8445***	0.6958***	0.7476***	0.7802***	0.7523***	0.7878***	0.8659***
	(0.0037)	(0.0069)	(0.0295)	(0.0041)	(0.0075)	(0.0080)	(0.0042)	(0.0145)	(0.0179)
log employment	0.1864***	0.2063***	0.1783***	0.2630***	0.2192***	0.1906***	0.2283***	0.1947***	0.1309***
	(0.0014)	(0.0023)	(0.0045)	(0.0014)	(0.0022)	(0.0019)	(0.0017)	(0.0036)	(0.0035)
log capital	0.0235***	0.0208***	-0.0022	0.0232***	0.0235***	0.0212***	0.0208***	0.0180***	0.0145***
	(0.0007)	(0.0017)	(0.0049)	(0.0010)	(0.0015)	(0.0012)	(0.0009)	(0.0026)	(0.0028)
Observations	40,781	21,864	8,600	140,055	34,787	17,174	50,149	14,607	16,339
R-squared	0.9995	0.9986	0.9984	0.9965	0.9983	0.9994	0.9988	0.9988	0.9991
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
KP rk LM stat under-identif.	216.8	139.9	33.77	1153	169.9	65.55	343.1	50.87	41.05
Under-identif. p-value	0	0	2.21e-07	0	0	0	0	5.21e-11	6.39e-09
KP rk Wald F stat weak identif.	97.72	59.08	9.473	521.1	62.63	23.84	142.4	15.17	13.87

VARIABLES	Industry 30	Industry 31	Industry 32	Industry 33	Industry 3X	Industry 41	Industry 42	Industry 43	Industry 45
log intermediates	0.7208***	0.7936***	0.7942***	0.7098***	0.8139***	0.7628***	0.7221***	0.7129***	0.8596***
	(0.0317)	(0.0036)	(0.0053)	(0.0133)	(0.0173)	(0.0047)	(0.0057)	(0.0023)	(0.0047)
log employment	0.1910***	0.1980***	0.1884***	0.2586***	0.1836***	0.1223***	0.1628***	0.1986***	0.1529***
	(0.0110)	(0.0015)	(0.0018)	(0.0035)	(0.0041)	(0.0016)	(0.0022)	(0.0008)	(0.0009)
log capital	0.0179***	0.0137***	0.0220***	0.0162***	0.0025	0.0440***	0.0224***	0.0388***	0.0080***
	(0.0053)	(0.0008)	(0.0010)	(0.0028)	(0.0047)	(0.0009)	(0.0019)	(0.0006)	(0.0007)
Observations	6,777	30,479	37,265	27,531	27,315	172,303	98,091	606,002	255,910
R-squared	0.9963	0.9994	0.9983	0.9935	0.9954	0.9907	0.9915	0.9903	0.9968
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
KP rk LM stat under-identif.	42.31	170.5	215.2	171.9	118.6	1259	847.4	5055	526.3
Under-identif. p-value	3.44e-09	0	0	0	0	0	0	0	0
KP rk Wald F stat weak identif.	18.45	65.08	87.30	57.25	36.51	403	308.4	1807	217.2

VARIABLES	Industry 46	Industry 47	Industry 49	Industry 52	Industry 53	Industry 55	Industry 56	Industry 58	Industry 59
log intermediates	0.9026***	0.8733***	0.7384***	0.7723***	0.7158***	0.8310***	0.8402***	0.6098***	0.6869***
	(0.0047)	(0.0029)	(0.0047)	(0.0120)	(0.0233)	(0.0104)	(0.0034)	(0.0128)	(0.0102)
log employment	0.1258***	0.1244***	0.2306***	0.2238***	0.1800***	0.1358***	0.1476***	0.2905***	0.1841***
	(0.0007)	(0.0004)	(0.0012)	(0.0029)	(0.0040)	(0.0022)	(0.0006)	(0.0038)	(0.0035)
log capital	0.0066***	0.0121***	0.0135***	0.0005	0.0556***	0.0191***	-0.0117***	0.0375***	0.0209***
	(0.0006)	(0.0004)	(0.0017)	(0.0018)	(0.0055)	(0.0021)	(0.0011)	(0.0026)	(0.0032)
Observations	475,117	679,662	140,715	56,558	20,137	66,695	422,517	34,010	43,395
R-squared	0.9968	0.9976	0.9944	0.9924	0.9862	0.9913	0.9945	0.9909	0.9803
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
KP rk LM stat under-identif.	1179	1238	1164	221.7	91.55	413.1	1925	208.2	475.4
Under-identif. p-value	0	0	0	0	0	0	0	0	0
KP rk Wald F stat weak identif.	367.9	405.9	493.3	73.70	30.05	152.2	741.1	72.83	159.3

Notes: Employment is number of employees count including the owner(s). Revenue, intermediates and capital have been deflated using indexes provided by the ONS with the base year being 2017. Estimations refer to a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by SIC industry) using a methodology consistent with Wooldridge (2009). 'KP rk LM stat under-identif.' (and the related p-value) refer to an under-identification test while 'KP rk Wald F stat weak identif.' is the F-test statistic of weak identification. Firm-level clustered standard errors in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A-2: UK Data. WLD TFP production function estimations by industry: industries 5X to 9Z.

VARIABLES	Industry 5X	Industry 62	Industry 63	Industry 68	Industry 69	Industry 6X	Industry 70	Industry 71	Industry 72
log intermediates	0.8060***	0.4096***	0.6480***	0.6659***	0.5145***	0.6963***	0.4994***	0.4568***	0.6928***
	(0.0242)	(0.0045)	(0.0406)	(0.0053)	(0.0033)	(0.0271)	(0.0041)	(0.0045)	(0.0288)
log employment	0.1771***	0.3941***	0.2905***	0.2413***	0.4831***	0.1909***	0.3238***	0.3984***	0.2625***
	(0.0075)	(0.0025)	(0.0066)	(0.0021)	(0.0014)	(0.0065)	(0.0018)	(0.0022)	(0.0072)
log capital	0.0540***	0.0464***	0.0281***	0.0531***	0.0387***	0.0235***	0.0422***	0.0253***	-0.0047
	(0.0075)	(0.0016)	(0.0080)	(0.0009)	(0.0011)	(0.0060)	(0.0013)	(0.0015)	(0.0062)
Observations	6,769	285,543	12,322	138,560	189,071	13,884	254,544	210,952	9,669
R-squared	0.9926	0.9680	0.9875	0.9820	0.9941	0.9902	0.9719	0.9776	0.9932
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
KP rk LM stat under-identif.	22.67	2529	37.80	1356	1143	63.85	2338	2129	53.58
Under-identif. p-value	4.73e-05	0	3.12e-08	0	0	0	0	0	0
KP rk Wald F stat weak identif.	10.37	898.1	9.856	605.4	431.2	18.76	820.4	826.5	15.38

VARIABLES	Industry 73	Industry 74	Industry 75	Industry 77	Industry 78	Industry 79	Industry 80	Industry 81	Industry 82
log intermediates	0.7249***	0.6049***	0.7023***	0.7458***	0.6113***	0.7014***	0.5234***	0.6165***	0.6520***
	(0.0118)	(0.0081)	(0.0252)	(0.0085)	(0.0138)	(0.0118)	(0.0167)	(0.0055)	(0.0061)
log employment	0.2404***	0.3170***	0.2773***	0.1987***	0.3336***	0.2365***	0.3947***	0.3148***	0.2525***
	(0.0023)	(0.0023)	(0.0054)	(0.0024)	(0.0018)	(0.0025)	(0.0029)	(0.0011)	(0.0018)
log capital	0.0236***	0.0209***	0.0029	0.0330***	0.0858***	0.0324***	0.0587***	0.0453***	0.0232***
	(0.0025)	(0.0021)	(0.0045)	(0.0026)	(0.0036)	(0.0016)	(0.0052)	(0.0018)	(0.0016)
Observations	57,336	110,370	16,568	54,982	79,813	30,385	20,836	92,107	133,749
R-squared	0.9917	0.9747	0.9939	0.9923	0.9855	0.9947	0.9846	0.9904	0.9862
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
KP rk LM stat under-identif.	274.6	810.7	76.17	455.1	214.7	125.6	125.6	817.5	785.8
Under-identif. p-value	0	0	0	0	0	0	0	0	0
KP rk Wald F stat weak identif.	78.38	265.5	30.04	203.7	58.42	38.53	41.05	287.5	267.4

VARIABLES	Industry 84	Industry 85	Industry 88	Industry 8X	Industry 94	Industry 95	Industry 9X	Industry 9Y	Industry 9Z
log intermediates	0.8645***	0.6960***	0.6767***	0.6717***	0.8723***	0.7004***	0.7026***	0.8008***	0.6757***
	(0.0436)	(0.0077)	(0.0162)	(0.0167)	(0.0120)	(0.0190)	(0.0079)	(0.0089)	(0.0045)
log employment	0.1513***	0.2425***	0.2720***	0.2778***	0.1187***	0.2655***	0.1044***	0.1702***	0.2412***
	(0.0164)	(0.0020)	(0.0032)	(0.0041)	(0.0027)	(0.0042)	(0.0025)	(0.0016)	(0.0012)
log capital	0.0783	-0.0025	-0.0226***	-0.0230***	-0.0193***	0.0262***	0.0168***	-0.0132***	0.0097***
	(0.1105)	(0.0016)	(0.0039)	(0.0037)	(0.0027)	(0.0037)	(0.0020)	(0.0022)	(0.0012)
Observations	2,517	64,504	14,099	28,015	30,706	19,741	52,550	78,835	187,084
R-squared	0.9967	0.9955	0.9909	0.9925	0.9916	0.9911	0.9862	0.9905	0.9849
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
KP rk LM stat under-identif.	56.62	303.2	115.5	133.1	279.2	61.05	509.2	551.5	1407
Under-identif. p-value	0	0	0	0	0	0	0	0	0
KP rk Wald F stat weak identif.	22.71	95.58	42.34	41.18	126	19.32	180.7	235.2	565.6

Notes: Employment is number of employees count including the owner(s). Revenue, intermediates and capital have been deflated using indexes provided by the ONS with the base year being 2017. Estimations refer to a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by SIC industry) using a methodology consistent with Wooldridge (2009). 'KP rk LM stat under-identif.' (and the related p-value) refer to an under-identification test while 'KP rk Wald F stat weak identif.' is the F-test statistic of weak identification. Firm-level clustered standard errors in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A-3: French Data. WLD TFP production function estimations by industry: industries 0Y to 5X.

VARIABLES	Industry 0Y	Industry 0Z	Industry 13	Industry 16	Industry 17	Industry 18	Industry 1X	Industry 1Y	Industry 22
log intermediates	0.6750***	0.6461***	0.6369***	0.6927***	0.6282***	0.6632***	0.6868***	0.5209***	0.7121***
	(0.0286)	(0.0168)	(0.0322)	(0.0128)	(0.0446)	(0.0175)	(0.0060)	(0.0315)	(0.0184)
log employment	0.1763***	0.2149***	0.2738***	0.2459***	0.2334***	0.2771***	0.1973***	0.3488***	0.2233***
	(0.0066)	(0.0044)	(0.0057)	(0.0033)	(0.0079)	(0.0035)	(0.0008)	(0.0042)	(0.0034)
log capital	0.0825***	0.0573***	0.0313***	0.0319***	0.0540***	0.0325***	0.0760***	0.0629***	0.0325***
	(0.0088)	(0.0043)	(0.0073)	(0.0034)	(0.0109)	(0.0029)	(0.0013)	(0.0076)	(0.0053)
Observations	17,995	33,010	30,159	62,893	15,945	90,801	502,829	44,665	49,645
R-squared	0.9947	0.9762	0.9884	0.9919	0.9963	0.9859	0.9941	0.9817	0.9952
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
KP rk LM stat under-identif.	68.92	238.4	123.2	308.3	38.35	266.5	1640	145.3	192.6
Under-identif. p-value	0	0	0	0	2.38E-08	0	0	0	0
KP rk Wald F stat weak identif.	22.31	81.94	37.4	107.7	10.04	82.56	503.7	43.53	63.94

VARIABLES	Industry 23	Industry 24	Industry 25	Industry 26	Industry 27	Industry 28	Industry 29	Industry 2X	Industry 30
log intermediates	0.6834***	0.7388***	0.5603***	0.6339***	0.6749***	0.6410***	0.7893***	0.7919***	0.6436***
	(0.0178)	(0.0283)	(0.0094)	(0.0230)	(0.0194)	(0.0107)	(0.0288)	(0.0221)	(0.0899)
log employment	0.1903***	0.1872***	0.2728***	0.2688***	0.2294***	0.2472***	0.1932***	0.1562***	0.2285***
	(0.0031)	(0.0067)	(0.0021)	(0.0052)	(0.0057)	(0.0036)	(0.0059)	(0.0054)	(0.0129)
log capital	0.0421***	0.0113	0.0630***	0.0357***	0.0299***	0.0337***	0.0234***	0.0313***	0.0385**
	(0.0036)	(0.0113)	(0.0024)	(0.0055)	(0.0050)	(0.0028)	(0.0063)	(0.0053)	(0.0166)
Observations	51,874	10,820	205,693	31,707	23,889	73,345	19,682	31,995	4,119
R-squared	0.9942	0.9969	0.9894	0.9926	0.9955	0.9935	0.9973	0.9971	0.9965
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
KP rk LM stat under-identif.	235.7	31.89	1087	174.3	164.4	485.5	78.79	118.4	16.73
Under-identif. p-value	0	5.52e-07	0	0	0	0	0	0	0.000803
KP rk Wald F stat weak identif.	74.23	9.921	351.2	52.64	57.94	178.1	23.37	38.20	4.286

VARIABLES	Industry 31	Industry 32	Industry 33	Industry 3X	Industry 41	Industry 42	Industry 43	Industry 45	Industry 46
log intermediates	0.7097***	0.3535***	0.6009***	0.4617***	0.7195***	0.7552***	0.6417***	0.7133***	0.7022***
	(0.0100)	(0.0232)	(0.0093)	(0.0250)	(0.0087)	(0.0223)	(0.0021)	(0.0061)	(0.0076)
log employment	0.2284***	0.3180***	0.3035***	0.2017***	0.2073***	0.1683***	0.2420***	0.1598***	0.1793***
	(0.0029)	(0.0030)	(0.0023)	(0.0049)	(0.0037)	(0.0075)	(0.0006)	(0.0011)	(0.0013)
log capital	0.0303***	0.1249***	0.0264***	0.1466***	0.0278***	0.0223***	0.0357***	0.0470***	0.0232***
	(0.0025)	(0.0045)	(0.0021)	(0.0057)	(0.0022)	(0.0073)	(0.0005)	(0.0012)	(0.0014)
Observations	59,520	88,117	127,324	57,119	88,693	17,471	2,024,420	589,557	845,391
R-squared	0.9920	0.9750	0.9852	0.9829	0.9811	0.9919	0.9837	0.9934	0.9860
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
KP rk LM stat under-identif.	434.2	311	1073	333.7	794.7	141.5	15790	1709	2642
Under-identif. p-value	0	0	0	0	0	0	0	0	0
KP rk Wald F stat weak identif.	158.4	110.5	362.5	95.38	264.9	51.37	5453	514.1	760.4

VARIABLES	Industry 47	Industry 49	Industry 52	Industry 53	Industry 55	Industry 56	Industry 58	Industry 59	Industry 5X
log intermediates	0.7449***	0.4500***	0.4528***	0.3397***	0.7719***	0.8399***	0.5624***	0.7943***	0.7780***
	(0.0033)	(0.0097)	(0.0415)	(0.1181)	(0.0109)	(0.0046)	(0.0263)	(0.0173)	(0.0643)
log employment	0.1384***	0.2487***	0.2446***	0.2888***	0.2071***	0.1423***	0.3305***	0.2436***	0.2790***
	(0.0005)	(0.0019)	(0.0050)	(0.0129)	(0.0022)	(0.0007)	(0.0062)	(0.0071)	(0.0182)
log capital	0.0472***	0.0712***	0.0733***	0.0575***	0.0806***	0.0536***	0.0505***	0.0159***	0.0237***
	(0.0006)	(0.0014)	(0.0081)	(0.0142)	(0.0020)	(0.0007)	(0.0047)	(0.0046)	(0.0091)
Observations	1,942,421	347,549	54,988	4,778	279,770	1,034,863	56,898	46,673	7,690
R-squared	0.9937	0.9874	0.9786	0.9724	0.9786	0.9876	0.9807	0.9684	0.9842
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
KP rk LM stat under-identif.	5413	2607	178.5	28.11	595.3	2700	236	403.5	23.48
Under-identif. p-value	0	0	0	3.44e-06	0	0	0	0	3.21e-05
KP rk Wald F stat weak identif.	1598	668.5	47.11	7.298	221.4	800.7	74.96	142.4	7.073

Notes: Employment is number of employees. Revenue, intermediates and capital have been deflated using indexes provided by the INSEE with the base year being 2017. Estimations refer to a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by NACE rev2 industry) using a methodology consistent with Wooldridge (2009). 'KP rk LM stat under-identif.' (and the related p-value) refer to an under-identification test while 'KP rk Wald F stat weak identif.' is the F-test statistic of weak identification. Firm-level clustered standard errors in parenthesis. *** p<0.01, ** p<0.05, * p<0.1.

Table A-4: French Data. WLD TFP production function estimations by industry: industries 62 to 9Z.

VARIABLES	Industry 62	Industry 63	Industry 68	Industry 69	Industry 6X	Industry 70	Industry 71	Industry 72	Industry 73
log intermediates	0.3428***	0.5113***	0.6833***	0.5289***	0.6658***	0.4430***	0.5873***	0.7230***	0.5670***
	(0.0220)	(0.0421)	(0.0094)	(0.0114)	(0.0563)	(0.0226)	(0.0081)	(0.0387)	(0.0214)
log employment	0.4749***	0.3893***	0.2754***	0.3953***	0.2646***	0.4151***	0.3640***	0.2999***	0.3220***
	(0.0038)	(0.0073)	(0.0028)	(0.0019)	(0.0084)	(0.0039)	(0.0021)	(0.0095)	(0.0035)
log capital	0.0172***	0.0338***	0.0586***	0.0465***	0.0452***	0.0473***	0.0259***	-0.0319***	0.0338***
	(0.0039)	(0.0088)	(0.0020)	(0.0018)	(0.0115)	(0.0036)	(0.0017)	(0.0084)	(0.0045)
Observations	139,397	25,538	297,426	351,062	14,025	135,638	332,281	11,848	97,471
R-squared	0.9600	0.9736	0.9529	0.9427	0.9896	0.9196	0.9661	0.9868	0.9714
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
KP rk LM stat under-identif.	703	105.1	1637	1242	61.78	666.9	2086	96.34	377.8
Under-identif. p-value	0	0	0	0	0	0	0	0	0
KP rk Wald F stat weak identif.	189.2	31.62	588.7	387.8	16.34	191.8	659.3	27.89	108.8

VARIABLES	Industry 74	Industry 75	Industry 77	Industry 78	Industry 79	Industry 80	Industry 81	Industry 82	Industry 8U
log intermediates	0.6532***	0.8136***	0.6912***	-0.3174***	0.6674***	0.2909***	0.3579***	0.6518***	0.3711***
	(0.0373)	(0.0379)	(0.0311)	(0.1115)	(0.0188)	(0.0596)	(0.0179)	(0.0177)	(0.0140)
log employment	0.3608***	0.1784***	0.1778***	0.6389***	0.2781***	0.4909***	0.3853***	0.3514***	0.2548***
	(0.0067)	(0.0042)	(0.0051)	(0.0065)	(0.0049)	(0.0052)	(0.0015)	(0.0027)	(0.0015)
log capital	0.0147**	0.0177***	0.1028***	0.1611***	0.0198***	0.0684***	0.0717***	0.0022	0.0630***
	(0.0059)	(0.0038)	(0.0059)	(0.0196)	(0.0043)	(0.0141)	(0.0034)	(0.0038)	(0.0014)
Observations	29,211	23,466	56,873	34,882	40,475	32,751	164,499	124,400	395,301
R-squared	0.9467	0.9769	0.9822	0.8200	0.9835	0.9466	0.9616	0.9635	0.9500
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
KP rk LM stat under-identif.	164.7	48.95	268.8	164.2	136.8	84.88	986.3	411.6	1243
Under-identif. p-value	0	1.34e-10	0	0	0	0	0	0	0
KP rk Wald F stat weak identif.	46.54	15.71	70.77	42.93	43.97	22.40	294.4	129.3	389.8

VARIABLES	Industry 8Z	Industry 94	Industry 95	Industry 9X	Industry 9Y	Industry 9Z
log intermediates	0.5644***	0.7301***	0.6184***	0.8103***	0.6646***	0.5486***
	(0.0195)	(0.0165)	(0.0167)	(0.0108)	(0.0328)	(0.0068)
log employment	0.3146***	0.2838***	0.2694***	0.2235***	0.2652***	0.3344***
	(0.0029)	(0.0061)	(0.0029)	(0.0035)	(0.0043)	(0.0011)
log capital	0.0109***	0.0017	0.0514***	0.0073**	0.0603***	0.0429***
	(0.0033)	(0.0062)	(0.0032)	(0.0031)	(0.0066)	(0.0015)
Observations	138,241	20,737	57,404	42,362	48,485	599,775
R-squared	0.9631	0.9716	0.9781	0.9794	0.9766	0.9488
Year Dummies	YES	YES	YES	YES	YES	YES
KP rk LM stat under-identif.	611.2	113.6	245.6	454.1	168.8	2347
Under-identif. p-value	0	0	0	0	0	0
KP rk Wald F stat weak identif.	183.6	60.49	90.83	191.4	46.10	920.4

Notes: Employment is number of employees. Revenue, intermediates and capital have been deflated using indexes provided by the INSEE with the base year being 2017. Estimations refer to a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by NACE rev2 industry) using a methodology consistent with Wooldridge (2009). 'KP rk LM stat under-identif.' (and the related p-value) refer to an under-identification test while 'KP rk Wald F stat weak identif.' is the F-test statistic of weak identification. Firm-level clustered standard errors in parenthesis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A-5: UK Data: number of firms and total employment covered by SIC industry for the year 2017

SIC industry	SIC details	Number of firms	Total employment
1		21,430	172,647
0X	Covers SIC 02,03	1,813	12,016
0Y	Covers SIC 05, 06, 07, 08, 09	491	41,445
13		2,162	42,633
16		4,172	56,917
17		835	50,963
18		5,900	81,024
1X	Covers SIC 10, 11, 12	5,179	407,367
1Y	Covers SIC 14, 15	1,761	28,076
22		3,788	142,603
23		2,156	83,046
24		944	61,138
25		13,354	248,194
26		2,883	92,866
27		1,710	60,602
28		4,324	155,690
29		1,438	142,688
2X	Covers SIC 19, 20, 21	1,758	116,300
30		758	164,224
31		3,374	71,939
32		3,695	53,971
33		5,452	69,313
3X	Covers SIC 35, 36, 37, 38, 39	4,452	274,109
41		25,971	242,454
42		9,336	180,940
43		88,057	543,997
45		34,900	476,269
46		51,914	950,915
47		89,585	2,414,936
49		17,181	435,846
52		5,179	291,990
53		3,518	205,136
55		7,833	371,451
56		56,363	1,419,417
58		4,088	102,830
59		7,233	71,423
5X	Covers SIC 50, 51	682	49,312
62		46,770	391,303
63		2,225	46,138
68		18,093	298,456
69		22,901	495,478
6X	Covers SIC 60, 61	3,455	207,608
70		48,479	363,535
71		32,261	332,440
72		1,542	86,103
73		8,682	128,911
74		19,630	102,025
75		1,882	47,921
77		6,837	117,508
78		11,655	824,150
79		2,937	78,529
80		3,344	156,139
81		13,212	541,699
82		18,356	265,039
84		469	128,529
85		8,165	836,069
88		1,806	227,389
8X	Covers SIC 86, 87	4,040	555,582
94		3,287	94,823
95		2,674	26,899
9X	Covers SIC 90, 91	7,729	98,227
9Y	Covers SIC 92, 93	10,485	431,448
9Z	Covers SIC 96, 99	17,822	173,079

Notes: Employment is number of employees count including the owner(s). The year 2017 corresponds to the fiscal year 2017-18. SIC industries in column 1 correspond either to a unique two-digit SIC 2007 code or are obtained from aggregation of two-digits SIC 2007 codes as indicated in column 2. Financial and insurance activities (SIC 2007 codes 64, 65 and 66) are excluded in our analysis.

Table A-6: French Data: number of firms and total employment covered by NACE rev2 industry for the year 2017

NACE rev2 industry	NACE rev2 details	Number of firms	Total employment
0Y	Covers NACE rev2 05, 06, 07, 08, 09	902	19,577
0Z	Covers NACE rev2 01, 02,03	–	–
13		1,595	34,530
16		3,526	53,333
17		951	61,811
18		4,246	47,461
1X	Covers NACE rev2 10, 11, 12	35,518	548,370
1Y	Covers NACE rev2 14, 15	1,733	59,981
22		2,813	146,251
23		2,718	95,778
24		618	65,876
25		13,101	283,745
26		1,556	114,036
27		1,415	100,655
28		3,377	172,053
29		1,263	114,183
2X	Covers NACE rev2 19, 20, 21	1,920	226,699
30		454	138,084
31		2,310	38,132
32		3,591	31,773
33		10,100	168,978
3X	Covers NACE rev2 35, 36, 37, 38, 39	4,562	195,171
41		10,797	121,066
42		2,363	157,685
43		143,083	980,980
45		41,687	343,432
46		55,551	965,145
47		141,963	527,186
49		24,922	599,885
52		3,133	173,059
53		–	–
55		18,438	178,418
56		103,574	674,056
58		4,661	129,113
59		5,270	61,974
5X	Covers NACE rev2 50, 51	677	68,340
62		9,195	291,941
63		1,772	43,692
68		24,481	220,593
69		15,677	186,791
6X	Covers NACE rev2 60, 61	1,357	164,652
70		20,675	248,623
71		26,521	341,965
72		1,078	47,153
73		6,623	106,436
74		5,951	34,914
75		1,853	10,415
77		4,437	102,369
79		2,716	32,717
80		1,888	145,981
81		16,821	416,157
82		8,645	214,575
8U	Covers NACE rev2 86, 87, 88	18,877	466,001
8Z	Covers NACE rev2 84, 85	12,887	110,387
94		679	8,132
95		4,178	30,338
9X	Covers NACE rev2 90, 91	2,978	24,940
9Y	Covers NACE rev2 92, 93	7,729	87,957
9Z	Covers NACE rev2 96, 99	47,454	166,629

Notes: – indicates cell suppressed because of disclosure. Employment is number of employees. NACE rev2 industries in column 1 correspond either to a unique two-digit NACE rev2 code or are obtained from aggregation of two-digits NACE rev2 codes as indicated in column 2. Financial and insurance activities (NACE rev2 codes 64, 65 and 66) are excluded in our analysis.

Table A-7: UK Data. Average (employment weighted) apparent labour productivity and labour productivity by year using current values

Year	Apparent Lab. Prod.	Lab. Prod.	N. of firms
2004	141,054	37,055	642,748
2005	155,780	39,822	681,104
2006	159,100	40,686	695,050
2007	169,178	41,118	717,933
2008	165,131	39,987	701,827
2009	160,519	42,065	684,485
2010	174,337	42,682	681,465
2011	184,016	44,227	700,898
2012	185,985	46,793	692,865
2013	187,130	47,988	716,939
2014	195,882	50,800	728,632
2015	190,949	53,653	740,365
2016	198,554	57,215	755,413
2017	206,930	59,777	814,407

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. Apparent labour productivity is computed as firm revenue (in current values) over firm employment. Labour productivity is computed as firm value added (in current values) over firm employment. All firm-level variables have been aggregated using firm employment as weight.

Table A-8: UK Data. Single Travel To Work Area firms – 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	155,912	47,857	3.645	3.255	1.749	622,287
2005	178,309	45,695	3.663	3.269	1.783	660,857
2006	174,077	48,661	3.704	3.311	1.770	674,797
2007	166,897	48,712	3.710	3.311	1.761	697,730
2008	153,124	43,848	3.697	3.299	1.732	681,862
2009	146,794	41,011	3.674	3.278	1.726	664,516
2010	158,407	39,850	3.679	3.276	1.684	661,541
2011	153,598	37,966	3.683	3.278	1.681	680,680
2012	145,641	39,646	3.693	3.286	1.677	672,137
2013	146,041	39,686	3.709	3.299	1.706	695,721
2014	161,022	40,157	3.726	3.317	1.716	707,245
2015	150,702	42,404	3.741	3.331	1.750	718,654
2016	159,238	46,713	3.737	3.328	1.746	733,595
2017	157,602	46,516	3.765	3.357	1.800	792,036

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. Revenue, intermediates and capital have been deflated using indexes provided by the ONS with the base year being 2017. Apparent labour productivity is computed as firm revenue (in 2017 pounds) over firm employment. Labour productivity is computed as firm value added (in 2017 pounds) over firm employment. OLS TFP is firm-level total factor productivity obtained from a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by SIC industry) using the OLS estimator. WLD TFP is firm-level total factor productivity obtained from a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by SIC industry) using a methodology consistent with Wooldridge (2009). Markups are estimates of the firm-level price to marginal cost ratio and are obtained from WLD TFP estimations and the share of intermediates in revenue as developed in De Loecker and Warzynski (2012). All firm-level variables have been aggregated using firm employment as weight. The number of firms refers to observations remaining after imposing that a firm has all of her establishments located in a single Travel To Work Area.

Table A-9: UK Data. Multiple Travel To Work Area firms – 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	219,049	64,496	3.379	2.881	1.422	20,461
2005	220,125	65,719	3.407	2.901	1.408	20,247
2006	221,162	64,168	3.432	2.922	1.384	20,253
2007	240,128	62,408	3.424	2.912	1.416	20,203
2008	212,727	50,747	3.424	2.908	1.392	19,965
2009	201,217	52,428	3.429	2.917	1.405	19,969
2010	209,986	47,854	3.460	2.943	1.399	19,924
2011	218,034	47,774	3.454	2.932	1.397	20,218
2012	223,496	51,579	3.463	2.944	1.422	20,728
2013	221,983	53,142	3.511	2.995	1.406	21,218
2014	228,860	58,095	3.611	3.098	1.439	21,387
2015	233,734	64,310	3.678	3.165	1.433	21,711
2016	238,469	67,817	3.678	3.162	1.475	21,818
2017	244,674	69,924	3.714	3.199	1.482	22,371

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. Revenue, intermediates and capital have been deflated using indexes provided by the ONS with the base year being 2017. Apparent labour productivity is computed as firm revenue (in 2017 pounds) over firm employment. Labour productivity is computed as firm value added (in 2017 pounds) over firm employment. OLS TFP is firm-level total factor productivity obtained from a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by SIC industry) using the OLS estimator. WLD TFP is firm-level total factor productivity obtained from a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by SIC industry) using a methodology consistent with Wooldridge (2009). Markups are estimates of the firm-level price to marginal cost ratio and are obtained from WLD TFP estimations and the share of intermediates in revenue as developed in De Loecker and Warzynski (2012). All firm-level variables have been aggregated using firm employment as weight. The number of firms refers to observations remaining after imposing that a firm has her establishments located in more than one Travel To Work Area.

Table A-10: French Data. Single Zone d'Emploi firms – 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	183,898	55,307	1.633	2.205	1.317	972,123
2001	184,032	56,064	1.647	2.217	1.323	958,766
2002	184,258	56,077	1.626	2.195	1.324	964,952
2003	183,785	56,948	1.636	2.208	1.329	987,099
2004	181,815	56,751	1.638	2.211	1.338	1,017,682
2005	190,321	58,815	1.673	2.395	1.335	987,519
2006	190,365	57,963	1.666	2.378	1.331	1,052,693
2007	193,393	59,526	1.675	2.394	1.331	1,082,347
2008	184,422	56,420	1.507	2.126	1.273	867,675
2009	179,569	57,560	1.599	2.362	1.288	868,914
2010	179,716	56,233	1.594	2.354	1.286	877,688
2011	197,508	58,259	1.608	2.364	1.275	880,028
2012	193,976	58,607	1.589	2.221	1.288	862,354
2013	198,991	60,687	1.599	2.234	1.293	814,473
2014	196,765	59,744	1.587	2.224	1.286	851,840
2015	192,035	58,316	1.595	2.238	1.294	826,891
2016	199,450	60,923	1.588	2.226	1.289	880,895
2017	201,225	60,982	1.581	2.215	1.294	841,416

Notes: Employment is number of employees. Revenue, intermediates and capital have been deflated using indexes provided by the INSEE with the base year being 2017. Apparent labour productivity is computed as firm revenue (in 2017 euros) over firm employment. Labour productivity is computed as firm value added (in 2017 euros) over firm employment. OLS TFP is firm-level total factor productivity obtained from a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by NACE rev2 industry) using the OLS estimator. WLD TFP is firm-level total factor productivity obtained from a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by NACE rev2 industry) using a methodology consistent with Wooldridge (2009). Markups are estimates of the firm-level price to marginal cost ratio and are obtained from WLD TFP estimations and the share of intermediates in revenue as developed in De Loecker and Warzynski (2012). All firm-level variables have been aggregated using firm employment as weight. The number of firms refers to observations remaining after imposing that a firm has all of her establishments located in a single Zone d'Emploi.

Table A-11: French Data. Multiple Zone d'Emploi firms – 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	274,342	77,239	1.681	2.851	1.199	53,419
2001	274,826	78,139	1.673	2.841	1.173	54,086
2002	280,770	78,713	1.680	2.817	1.187	56,666
2003	293,173	79,097	1.619	2.492	1.165	57,864
2004	286,339	78,703	1.680	2.808	1.180	59,321
2005	289,803	79,803	1.650	2.620	1.168	59,187
2006	298,155	80,884	1.656	2.637	1.165	60,948
2007	302,826	79,321	1.653	2.620	1.167	62,076
2008	293,040	77,959	1.619	2.605	1.179	60,032
2009	275,972	80,042	1.632	2.629	1.195	58,683
2010	283,365	79,203	1.621	2.611	1.186	59,686
2011	294,289	78,237	1.628	2.636	1.176	56,025
2012	286,605	78,139	1.656	2.827	1.186	57,038
2013	278,555	76,735	1.657	2.835	1.190	56,727
2014	285,563	78,087	1.658	2.856	1.185	57,474
2015	278,323	78,419	1.653	2.802	1.192	58,500
2016	281,486	79,833	1.658	2.800	1.191	59,833
2017	289,665	79,015	1.664	2.805	1.188	58,610

Notes: Employment is number of employees. Revenue, intermediates and capital have been deflated using indexes provided by the INSEE with the base year being 2017. Apparent labour productivity is computed as firm revenue (in 2017 euros) over firm employment. Labour productivity is computed as firm value added (in 2017 euros) over firm employment. OLS TFP is firm-level total factor productivity obtained from a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by NACE rev2 industry) using the OLS estimator. WLD TFP is firm-level total factor productivity obtained from a 3 inputs (intermediates, labour and capital) Cobb-Douglas production function where revenue is the output measure and coefficients are estimated (separately by NACE rev2 industry) using a methodology consistent with Wooldridge (2009). Markups are estimates of the firm-level price to marginal cost ratio and are obtained from WLD TFP estimations and the share of intermediates in revenue as developed in De Loecker and Warzynski (2012). All firm-level variables have been aggregated using firm employment as weight. The number of firms refers to observations remaining after imposing that a firm has her establishments located in more than one Zone d'Emploi.

Table A-12: UK: Establishment density regressions

	Unweighted	Weighted
log density	0.0173***	0.0116***
	(0.0007)	(0.0011)
	Reference category is London	
East Midlands	-0.0597***	-0.0893***
	(0.0022)	(0.0060)
East of England	-0.0405***	-0.0863***
	(0.0026)	(0.0063)
North East	-0.0321***	-0.0586***
	(0.0067)	(0.0090)
North West	-0.0633***	-0.1047***
	(0.0026)	(0.0073)
Northern Ireland	-0.0444***	-0.0974***
	(0.0029)	(0.0076)
Scotland	-0.0104***	-0.0522***
	(0.0035)	(0.0072)
South East	-0.0265***	-0.0571***
	(0.0029)	(0.0061)
South West	-0.0606***	-0.0980***
	(0.0026)	(0.0058)
Wales	-0.0718***	-0.0985***
	(0.0026)	(0.0067)
West Midlands	-0.0562***	-0.0693***
	(0.0034)	(0.0068)
Yorkshire and The Humber	-0.0699***	-0.1029***
	(0.0025)	(0.0058)
Observations	14,035,253	14,030,268
R-squared	0.0072	0.0039

Notes: The dependent variable is establishment (log) WLD TFP demeaned by the corresponding industry average. Log density is the 2015 population density of the TTWA r where establishment i is located at time t. Year dummies (not reported) and macro region dummies (London being the reference category) are included in the regressions. Column one provides simple OLS regressions of equation (4) while column two shows weighted OLS regressions of equation (4) where the weight is the share of the employment of establishment i and time t in overall regional employment in year t. Standard errors are clustered by TTWA-year. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A-13: France: Establishment density regressions

	Unweighted	Weighted
log density	0.0049***	0.0331***
	(0.0004)	(0.0009)
	Reference category is Île-de-France	
Auvergne-Rhône-Alpes	-0.0048**	-0.0229***
	(0.0021)	(0.0038)
Bourgogne-Franche-Comté	-0.0071***	-0.0119***
	(0.0021)	(0.0046)
Bretagne	0.0090***	-0.0376***
	(0.0020)	(0.0037)
Centre-Val de Loire	0.0009	0.0041
	(0.0021)	(0.0052)
Grand Est	-0.0075***	-0.0059
	(0.0020)	(0.0040)
Hauts-de-France	-0.0012	-0.0108***
	(0.0019)	(0.0041)
Normandie	0.0013	-0.0303***
	(0.0019)	(0.0038)
Nouvelle-Aquitaine	-0.0161***	-0.0345***
	(0.0021)	(0.0040)
Occitanie	-0.0304***	-0.0415***
	(0.0023)	(0.0041)
Pays de la Loire	0.0026	-0.0292***
	(0.0021)	(0.0041)
Provence-Alpes-Côte d'Azur	-0.0292***	-0.0286***
	(0.0020)	(0.0041)
Multi-region	-0.0129***	-0.0257***
	(0.0029)	(0.0047)
Observations	18,459,242	18,459,242
R-squared	0.0054	0.0110

Notes: The dependent variable is establishment (log) WLD TFP demeaned by the corresponding industry average. Log density is the 2009 population density of the ZE r where establishment i is located at time t. Year dummies (not reported) and macro region dummies (Paris being the reference category) are included in the regressions. Column one provides simple OLS regressions of equation (4) while column two shows weighted OLS regressions of equation (4) where the weight is the share of the employment of establishment i and time t in overall regional employment in year t. Standard errors are clustered by ZE-year. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A-14: UK: Establishments 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	0.0141***	0.0145***	0.0127***	0.0119***	0.0123***	0.0146***	0.0176***	0.0221***	0.0292***
	(0.0003)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0003)

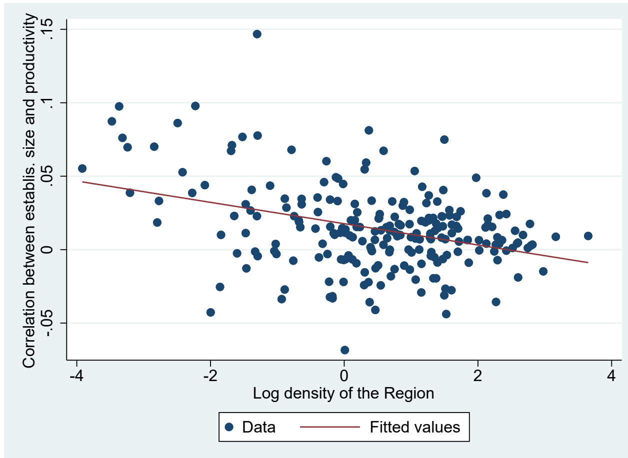
Notes: The dependent variable is establishment (log) WLD TFP demeaned by the corresponding industry average. Log density is the 2015 population density of the TTWA r where establishment i is located at time t. Year dummies and macro region dummies (not reported) are included in the regressions. Columns one to nine provides quantile regressions of equation (4) on deciles one to nine. Robust standard errors provided. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A-15: France: Establishment 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	-0.0060***	-0.0017***	0.0010***	0.0032***	0.0049***	0.0070***	0.0096***	0.0129***	0.0185***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)

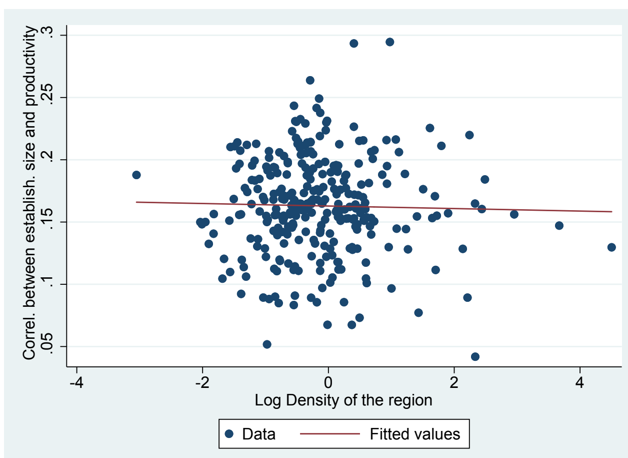
Notes: The dependent variable is establishment (log) WLD TFP demeaned by the corresponding industry average. Log density is the 2009 population density of the ZE r where establishment i is located at time t. Year dummies and macro region dummies (not reported) are included in the regressions. Columns one to nine provides quantile regressions of equation (4) on deciles one to nine. Robust standard errors provided. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Figure A-1: UK: Correlation between establishment productivity and size within each region



Notes: The Figure provides a scatter plot of two variables for each TTWA. The variable on the y-axis in the correlation (across establishments and years within a region) between establishment employment and (log) WLD TFP demeaned by the corresponding industry average. The variable on the x-axis is the 2015 population density of the TTWA r. The red line indicates the regression line.

Figure A-2: France: Correlation between establishment productivity and size within each region



Notes: The Figure provides a scatter plot of two variables for each ZE. The variable on the y-axis in the correlation (across establishments and years within a region) between establishment employment and (log) WLD TFP demeaned by the corresponding industry average. The variable on the x-axis is the 2009 population density of the ZE r. The red line indicates the regression line.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1954	Tim Obermeier	Individual welfare analysis: A tale of consumption, time use and preference heterogeneity
1953	Kirill Borusyak Xavier Jaravel	The distributional effects of trade: Theory and evidence from the United States
1952	Ariela Caglio Sebastien Laffitte Donato Masciandaro Gianmarco Ottaviano	Has financial fair play changed European football?
1951	Eugenie Dugoua Todd D. Gerarden	Induced innovation, inventors and the energy transition
1950	Christian Krekel Johannes Rode Alexander Roth	Do wind turbines have adverse health impacts?
1949	Réka Juhász Claudia Steinwender	Industrial policy and the great divergence
1948	Virginia Minni	Making the invisible hand visible: Managers and the allocation of workers to jobs
1947	Eugenie Dugoua	Induced innovation and international environmental agreements: Evidence from the ozone regime
1946	Marco Bertoni Gabriel Heller-Sahlgren Olmo Silva	Free to improve? The impact of free school attendance in England
1945	Matthias Mertens Bernardo Mottironi	Do larger firms exert more market power? Markups and markdowns along the size of distribution
1944	Ria Ivandić Anne Sophie Lassen	Gender gaps from labor market shocks

1943	Ihsaan Bassier Alan Manning Barbara Petrongolo	Vacancy duration and wages
1942	Cheng Keat Tang Stephen Gibbons	Are friends electric? Valuing the social costs of power lines using house prices
1941	Luis Bauluz Pawel Bukowski Mark Fransham Annie Seong Lee Margarita López Forero Filip Novokmet Sébastien Breau Neil Lee Clément Malgouyres Moritz Schularick Gregory Verdugo	Spatial wage inequality in North America and Western Europe: Changes between and within local labour markets 1975-2019
1940	Felipe Carozzi Edward Pinchbeck Luca Repetto	Scars of war: The legacy of WW1 deaths on civic capital and combat motivation
1939	Christos Genakos Costas Roumanias Tommaso Valletti	Is having an expert "friend" enough? An analysis of consumer switching behavior in mobile telephony
1938	Andrew E. Clark Maria Cotofan	Are the upwardly mobile more left-wing?
1937	Priya Manwaring Tanner Regan	Public disclosure and tax compliance: Evidence from Uganda
1936	Philippe Aghion Celine Antonin Luc Paluskiewicz David Stromberg Raphael Wargon Karolina Westin	Does Chinese research hinge on US co-authors? Evidence from the China initiative

The Centre for Economic Performance Publications Unit
Tel: +44 (0)20 7955 7673 Email info@cep.lse.ac.uk
Website: http://cep.lse.ac.uk Twitter: @CEP_LSE