ISSN 2042-2695

CENTRE *for* ECONOMIC PERFORMANCE

CEP Discussion Paper No 1619

May 2019

Tales of the City: What Do Agglomeration Cases Tell Us About Agglomeration in General?

Giulia Faggio
Olmo Silva
William C. Strange

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Abstract

This paper considers the heterogeneous microfoundations of agglomeration economies. It studies the co-location of industries to look for evidence of labor pooling, input sharing, and knowledge spillovers. The novel contribution of the paper is that it estimates single-industry models using a common empirical framework that exploits the cross-sectional variation in how one industry co-locates with the other industries in the economy. This unified approach yields evidence on the relative importance of the Marshallian microfoundations at the single-industry level, allowing for like-for-like cross-industry comparisons on the determinants of agglomeration. Using UK data, we estimate such microfoundations models for 97 manufacturing sectors, including the classic agglomeration cases of automobiles, computers, cutlery, and textiles. These four cases – as with all of the individual industry models we estimate – clearly show the importance of the Marshallian forces. However, they also highlight how the importance of these forces varies across industries – implying that extrapolation from cases should be viewed with caution. The paper concludes with an investigation of the pattern of heterogeneity. The degree of an industry's clustering (localization), dynamism, incumbent firm size, and worker education are shown to contribute to the pattern of heterogeneous microfoundations.

Key words: Agglomeration, microfoundations, heterogeneity, industrial clusters
JEL Codes: R10; R12; L52; L60

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

We thank the Economics and Social Research Council (ESRC) UK and the Social Sciences and Humanities Research Council for financial support. We also thank Filippo Boeri for outstanding research assistance. We are also grateful for the useful comments and suggestions we received from Stephan Heblich, Simona Iammarino, Dan McMillen, Henry Overman, Matt Turner and participants at the 2018 UEA Conference (New York), the 2018 NARSC Meetings (San Antonio), and the 2018 Singapore Management University Conference on Urban and Regional Economic. We are responsible for any errors or omissions.

Giulia Faggio, City, University of London and Centre for Economic Performance, London School of Economics. Olmo Silva, London School of Economics and Centre for Economic Performance, LSE. William Strange, University of Toronto.

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

I. Introduction

The economic literature on agglomeration has proceeded along the lines of most economic research: theories generate predictions, and these predictions are brought to data for quantitative econometric testing. Of course, the abstraction inherent to this sort of research program leaves out important specific details. A more qualitative literature on agglomeration considers specific cases. These case studies have considered a range of industries – including the classic 'tales' of cutlery, textiles, automobiles, and computers. Because the case studies embrace much of the detail that economic theory and econometric analysis are forced to abstract away from, they are highly valuable complements to more quantitative economic research. However, despite their usefulness, taken in aggregate, these studies raise a question: how much can be learned in general from particular agglomeration cases?

A prior paper of ours, Faggio et al. (2017), begins to address this question by considering patterns in the microfoundations of agglomeration economies. The paper builds on Ellison et al. (2010), who consider the coagglomeration of industry pairs. In this approach, a tendency for pairs linked in the supply chain to coagglomerate is consistent with input sharing, while coagglomeration of industry pairs linked in patent citations is consistent with knowledge spillovers. Lastly, a co-location tendency of an industry pair drawing on the same worker skills (as evidenced by the pattern of employment by occupation) is consistent with labor pooling. The key result in Ellison et al. is that Marshall's three forces – input sharing, knowledge spillovers and labor pooling – are all positively associated with the tendency of industries to coagglomerate. Faggio et al. (2017) extends Ellison et al. (2010) by documenting systematic variation in microfoundations across industry pairs. The coagglomeration of industry pairs is sometimes driven by input market linkages, sometimes by labor market relationships, and some other times by patterns of knowledge spillovers. This variation is important for its own sake as well as for the light it sheds on the microfoundations of agglomeration economies. For instance, industry pairs both characterised by the presence of smaller firms show stronger effects from input linkages – a result in the spirit of Chinitz (1961).

The present paper further explores the heterogeneity in microfoundations. It does so by using a common empirical framework that identifies the importance of the Marshallian foundations at the single-industry level by exploiting the cross-sectional variation in how one industry co-locates with all other industries in the economy. This unified approach yields novel evidence on the relative importance of labour pooling, input sharing and knowledge spillovers, allowing for like-for-like cross-industry comparisons on the determinants of agglomeration. Furthermore, this approach allows us to characterize the forces that drive one industry's coagglomeration (e.g., the industry is dominated by small firms or high levels of firms entry) rather than the forces the drive the coagglomeration of industry pairs with specific characteristics (e.g., both industries in the pair have small firms or considerable entry; as in Faggio et al., 2017).

Our evidence delivers the important cautionary insight that particular cases do not generalize easily and directly to the universe of industries – or to other industries – and that evidence gathered by pooling data across all sectors masks very significant differences. Indeed, the pattern we document reveals a stark heterogeneity in the relative importance of the three Marshallian forces for industrial co-location – with few industries impacted by the agglomeration forces in the same way.

These findings have the potential to inform policy interventions aimed at stimulating the emergence of economic hubs. A local planner interested in promoting the development of a cluster in a given industry should be especially careful in acting on lessons learned from another industry with very different microfoundations. Given the renewed interest in 'active' industrial policy in the UK and elsewhere in the world to engender local economic growth and stimulate productivity and innovation, our evidence is highly topical.[1] See for example the 'Industrial Strategy 2018' White Paper of the UK Government or the Franco-German 2019 manifesto for 'European Industrial Policy' (Chatterji et al., 2014, Duranton, 2011, and Neumark and Simpson, 2015 present a critical account of similar initiatives for the US).

In order to carry out our analysis, we employ confidential firm level data for 97 manufacturing industries from the UK's Business Structure Database (BSD) covering the years 1997-2008. We match this information with a range of other data on industry characteristics in order to arrive at proxies for the Marshallian agglomeration forces. We consider agglomeration at the Travel-to-Work-Area (TTWA) level. These areas are constructed to be self-contained labor markets. In that sense, they correspond to U.S. Metropolitan Statistical Areas or Canadian Census Metropolitan Areas.

The estimates of the individual industry models shed light on the nature of industry heterogeneity in agglomeration. We devote much of our attention to four classic cases: cutlery, textiles, cars and computers. Cutlery was considered by Marshall (1890). For this industry, we find evidence that input linkages and labor market pooling are important – while knowledge spillovers are not. The picture is different when considering textiles – which has also been of historical importance for the development of manufacturing in the UK (Landes, 1969). For this industry, we find large and significant labor pooling effects and significant but smaller knowledge spillovers. Conversely, the impact of input sharing is small and insignificant. These results show that one would not want to generalize from cutlery to textiles, and illustrate more generally the limits of extrapolation.

Without doubt, the computer and the car industries are amongst the most salient 'tales' in the agglomeration literature – though for opposite reasons. Saxenian (1994) offers an important and often quoted analysis of the Silicon Valley – and its glowing success. Conversely, Glaeser (2011) provides an informative discussion about the car industry's declining cluster surrounding Detroit – contrasting it to the thriving computer agglomeration in Greater San Jose. Our evidence shows that for the computer industry, knowledge spillovers are very important – while input sharing and labor market pooling seem unrelated to the co-location pattern of this industry with other sectors in the economy. Conversely, for the automobile industry, labor pooling has a large and significant effect, while knowledge spillovers have a smaller but

[1] For example, previous work by Devereux et al. (2007) shows that government subsidies are more effective in attracting firms to locations where large agglomerations pre-exists.

still significant coefficient. Input sharing instead has a small and insignificant impact. In short, we see a very different pattern of agglomeration effects across these four exemplary industries.

A similar heterogeneity appears when considering the drivers of coagglomeration for the rest of the individual industries. A handful of industries display relatively similar patterns to one of the four classic cases. But more often than not industries are characterised by individual patterns in terms of agglomeration microfoundations. Once again, this illustrates the limits to generalization. It also clarifies that, because of the substantial heterogeneity that we document, pooled regressions are not a valuable tool for identifying the microfoundations of agglomeration for individual sectors (nevertheless, pooled regressions are informative about common patterns that hold across industries).

As mentioned, to understand the pattern of heterogeneity, we consider the relationship between a range of industry characteristics and the industry-level coefficients on the Marshallian forces. We do this across all the industries in our sample and consider the following attributes: an industry's agglomeration (localization), its dynamism (proxied by its new firms' creation rate), its incumbent firms' size, and its workers' education. In studying the patterns, we are forced to deal with various ambiguous predictions arising from theory. For instance, different theoretical frameworks suggest that the agglomeration (localization) of a given industry could be a substitute or a complement for that industry's coagglomeration with other industries. In fact, it is not hard to imagine a model of either sort.[2] On the one hand, there may be a substitution effect of agglomeration if the presence of own industry activity fosters agglomeration – making cross-industry coagglomeration not as valuable. The complementarity argument, on the other hand, would suggest that industries that benefit from labor pooling, input sharing, and knowledge spillovers will seek to enjoy these benefits within and across sectors – i.e., by both agglomerating and coagglomerating. We see similar ambiguities in considering other channels by which agglomeration economies are created. For example, industry dynamism might strengthen agglomeration effects associated with matching – potentially raising the coefficients on any of the three Marshallian microfoundations. However, entry might deny market participants the trust needed for agglomeration effects to arise, therefore inhibiting the diffusion of knowledge (as in Helsley and Strange, 1994). There are similar ambiguities associated with the likely effect of the size of incumbent firms and with workers' education – a standard, but imperfect, proxy for skills.

In order to (empirically) resolve some of these ambiguities, we regress our estimated coefficients for the Marshallian agglomeration forces on our proxies for own-industry agglomeration, industry dynamism, incumbent firm size, and worker education. Regarding whether agglomeration and coagglomeration are complements or substitutes, we find that for labor pooling and knowledge spillovers, complementarity dominates. However, we find little effects on input sharing. Regarding dynamism, we find that more dynamic industries have larger labor pooling coefficients. Conversely, input sharing coefficients are smaller for the most dynamic industries. We do not find a significant relationship with the

[2] Duranton and Puga (2004) characterize agglomeration economies as arising from sharing, matching, and learning. It is possible for any of these forces to operate within industries, across industries, or both.

knowledge spillover coefficients. Next, we find no impact of incumbent size on labor pooling, while input sharing is less important with large existing firms – a Chinitz (1961) effect. For knowledge spillovers, we find instead a sort of anchor effect – with smaller firms having smaller effects. Finally, for education, the labor pooling effect is strongest for the less educated workers but input sharing is strongest with a more educated workforce – a pattern suggestive of the nursery effects discussed by Vernon (1960). We find no significant effect for knowledge spillovers.

The bottom line of all of the analysis is that the individual industry models both deepen the cases (which makes them more valuable) and clarify the limitations of extrapolating from the cases (which makes the cases less valuable). Taken together, the individual industry models coupled with the regressions exploring the regularities in the heterogeneity pattern can assist in the use of cases by suggesting situations in which a given case might apply to with reasonable accuracy.

The rest of the paper is structured as follows. In Section II we present the data and describe the main variables we use. In Section III, we present our findings on the four classic cases of agglomeration – namely cutlery, textile, cars and computers. Section IV discusses the heterogeneity we find when we explore all the manufacturing industries in our sample. Finally, Section V presents our attempt at rationalising this heterogeneity. Some concluding remarks are presented in Section VI.

II. Data

A. Data and variable construction

The core data we use to carry out our analysis is the UK Business Structure Database (BSD) covering the period 1997 to 2008. The data is an annual snapshot (taken in April at the closing of the fiscal year) of the Inter-Departmental Business Register (IDBR), which consists of constantly-updated administrative business data collected for taxation purposes. Businesses liable for value-added taxation (VAT) and/or with at least one employee registered for tax collection appears on the IDBR. In 2004, the Office for National Statistics (ONS) estimated that businesses listed on the IDBR accounted for approximately 99 per cent of economic activity in the UK.

Businesses tracked in the dataset are structured into enterprises and local units, where the first refers to the overall business organization while the second can be thought of as a plant or establishment. In the majority of cases (70 per cent), enterprises only have one local unit. In our work, we make use of data at the local unit level including plants belonging to both single- and multi-plant enterprises and located in England, Wales and Scotland. We neglect Northern Ireland because of poor data coverage.

The initial raw data includes approximately three million local units every year. However, in order to prepare the data for our analysis, we carry out a series of checks and drop a number of units. These mainly deal with inconsistencies in terms of anomalous opening/closing dates of establishments, and outliers in terms of concentration of establishments in very small-scale geographical units. The Web Appendix to Faggio et al. (2017) provides more detail of our sample selection and data cleaning procedures.[3]

For our analysis, we focus on three-digit industries of the UK Standard Industry Classification (SIC) 1992 and restrict our attention to manufacturing (SIC151-SIC372). We exclude, however, a few industries for the following reasons. First, 'Manufacturing of tobacco products' (SIC160) is dropped because of its limited number of plants throughout the sample period (e.g., 43 in 1997). Second, we disregard five industries for which we cannot measure one of our key variables of interest – namely, knowledge spillovers: 'Reproduction of recorded media' (SIC223); 'Manufacturing of machine tools' (SIC294); 'Manufacturing of weapons & ammunition' (SIC296); 'Recycling of metal waste and scrap' (SIC371;) and 'Recycling of non-metal waste and scrap' (SIC372).[4] After these restrictions, our sample covers 97 manufacturing 3-digit sectors for a total of 4,656 unique pairwise industry combinations for twelve years (1997-2008). The complete dataset thus contains 55,872 industry-pair-by-year observations.

In terms of geographical units of aggregation, we use Travel-to-Work Areas (TTWAs) – which are designed to guarantee that at least 75% of the resident population works in the area and that 75% of the people working in the area are resident there. These delineate areas that can be considered as self-contained labor markets and economically relevant aggregates. In 2007, there were 243 TTWAs within the United Kingdom. We focus on Britain (excluding Northern Ireland), split TTWAs into urban and rural ones, and only consider 84 urban TTWAs with population in excess of 100,000 residents. More detail is provided in the Web Appendix to our previous work (Faggio et al., 2017).[5]

To measure coagglomeration, we use the Ellison et al. (2010) metric calculated using the total employment shares of the selected 97 three-digit industries contained in the 84 urban TTWAs. More formally, let us denote total employment in industry i by N_i; and denote the employment in metropolitan area m and industry i by n_{mi}. The share of a given industry i's employment in metropolitan area m is defined as $s_{mi} = n_{mi} / N_i$, while the metropolitan area's share of national employment is denoted by x_m. For industries i and j, the Ellison et al. (2010) coagglomeration measure is defined as:

$$\gamma^C_{ij} = \frac{\sum_{m=1}^{M} (s_{mi} - x_m)(s_{mj} - x_m)}{1 - \sum_{m=1}^{M} (x_m)^2}. \tag{1}$$

[3] Accessible at: http://personal.lse.ac.uk/silvao/Heterogeneous%20Agglomeration%20Web%20Appendix.pdf

[4] Note that we checked whether our results substantially change if we include these sectors despite the missing data on knowledge spillovers. Broadly speaking, we found this was not the case. However, we prefer working with a clean dataset in which each variable we use is measured for every sector we consider. Similarly, we have checked that our results carry through if we use the 94 sectors we considered in our previous work (obtained by re-aggregating sectors with low employment/firm counts; see Faggio et al., 2017).

[5] In some extensions, we experimented with keeping rural areas or excluding London from our analysis. Overall, we find similar results. These are not tabulated for space reasons, but are available from the authors.

This measure is related to the covariance of industries across metropolitan areas. In order to study how this tendency of industry to co-locate is affected by the three standard Marshallian agglomeration forces of labour pooling, input sharing and knowledge spillovers, we construct the following proxies.

To measure the importance of labor pooling, we use the UK Labour Force Survey (LFS) data between 1995 and 1999 – at the beginning of our observation window – and investigate whether industries use similar types of workers. The LFS is a representative quarterly survey of households living in the UK sampling between 64,000 (earlier years) and 52,000 (later years) households every quarter, equivalent to about 120,000-150,000 individuals. In our work, we focus on 16-59 aged women and 16-64 aged men, and on individuals either working as employees or as self-employed. We only consider individuals with non-missing information on educational qualifications, industry of employment and occupation. Furthermore, we only keep those who live in English, Welsh or Scottish TTWAs – while we drop Northern Ireland (as we did for our main BSD data). Finally, we select individuals living in urban areas and working in manufacturing – leaving us with a sample of about 35,000 workers a year. We then use the 331 occupation groups defined by the three-digit LFS Standard Occupation Classification (SOC 1990, which categorizes occupations on the basis of skill level and skill content at a very detailed level) in conjunction with the 97 manufacturing industries defined at the three-digit SIC level to calculate $Share_{io}$ and $Share_{jo}$. These measure the shares of employees of occupation o in the total employment of industry i and j, respectively. Using this information, we proxy for labour pooling by measuring the similarity of employment in industries i and j computed as the correlation between $Share_{io}$ and $Share_{jo}$.

To assess the importance of this input sharing, we use the ONS Input-Output Analytical Tables (I-O henceforth) for 1995 to 1999. This allows us to measure the extent to which industries buy and sell intermediate inputs from one another. Specifically, we first calculate the shares of inputs that each industry within a pair buys from each other as fractions of their total intermediate inputs, and the shares of outputs that they sell to each other as fractions of their total output (excluding sales directly to consumers). We then follow Ellison et al. (2010) and Faggio et al. (2017) and proxy input sharing by taking the maximum of either the upstream linkages (i.e., the largest between the share of inputs that sector i buys from sector j and vice versa) or the downstream linkages (i.e., the largest between the share of output that sectors i sells to sector j, and vice versa) between two industries.[6]

Lastly, in order to obtain a proxy for knowledge spillovers, we track patent citation flows using information on UK inventors contained in the European Patent Office (EPO) data for the years 1997 to

[6] The sector classification used in the I-O Tables is more aggregated than the three-digit SIC industrial classification used in the BSD, and only includes 77 manufacturing industries. We assign input-output shares to SIC three-digit sectors belonging to the same I-O sector code using an apportioning procedure based on their employment share within the group (averaged between 1995 and 1999).

2009.[7] Approximately 144,000 patents were filed by 160,000 UK inventors (multiple inventors can be recorded for each patent). These generated a stream of more than 77,000 citations of UK patents over the observed time-window. In order to construct knowledge spillover measures we exclude self-citations from the same inventor (or the company at which he/she is based), as well as citing patents filed after 2000/before 1981, and cited patents filed after 1997. The latter two restrictions guarantee that on average cited patents are at least three years older than citing ones, and allows us to center our knowledge-spillover measures in the initial years of our sample (i.e. up to 2000) – so that they are measured at a similar time as the labor-pooling and input sharing metrics. Using these data, we measure the extent to which patents associated with industry *i* cite patents associated with industry *j* and vice-versa. One challenge lies with creating a mapping between sectors and patents – which are categorized using technological classes. Following the literature, we use a probabilistic mapping based on the Industry of Manufacture (IOM).[8] After applying these mapping procedures, we investigate the number of citations that a patent in sector *i* is receiving from patents in sector *j*, and the number of patents in sector *j* that a patent in sector *i* is citing. Our final indicator for knowledge spillovers consider the maximum patent-citation flow between sector *i* and sector *j* – normalized by total citations in that industry.

Using the various data sources discussed above, we construct an additional set of sector/sector-pair characters that we deploy in our analysis. To begin with, we create proxies that capture industry-pairs' similarity in terms of their reliance on natural and other geographically concentrated resources. These variables are used in our analysis of the agglomeration microfoundations to control for the tendency of certain industries to co-locate simply because of their reliance on resources and inputs that are unevenly distributed across space. In particular, we measure industries' use of primary inputs as a share of total inputs (using the I-O Tables) considering their purchasing from the following 'natural resources' sectors: agriculture, forestry, fishing, mining and quarrying. We also consider their usage of water and energy separately, as well as the share of inputs bought from transport-related sectors. In addition, we create a proxy for access to business services by considering the share of inputs bought from this compartment. Using these shares, we then construct proxies for the dissimilarity of industry pairs in terms of their reliance on these resources by measuring (one half of) the absolute value of the difference in the shares of these various inputs used by the pair.

Furthermore, we characterize sectors (not sector pairs) along four dimensions, which we use to study the heterogeneity in the strength of the agglomeration microfoundations that we document. First, we calculate an industry's agglomeration (localization) as measured by the Ellison and Glaeser (1997) index of spatial concentration at the three-digit sectoral level (obtained from the BSD data). Second, we consider industry dynamism by measuring the entry share of new firms in the sector (i.e., the incidence of new firms

[7] We acknowledge that patent citations are an imperfect proxy for knowledge spillovers (see, for a discussion, Breschi et al., 2005). However, alternative proxies for KS (e.g., based on innovation surveys or on various proximity metrics) have similar limitations. More details about the EPO dataset can be found in Breschi and Lissoni (2004).

[8] These probabilistic correspondences were developed by Statistics Canada and are discussed in Silverman (2002).

at time *t* in the total number of firms in that year; using the BSD data). Third, we consider data on the share of college graduates in each industry to measure average education levels (obtained using the LFS data). Lastly, we characterize sectors by measuring the average size of its incumbents – i.e., the employment of firms operating both at time *t* and *t-1* (based on the BSD data).

B. Key descriptive statistics

Descriptive statistics for our industry-pair dataset are presented in Table 1. The first row of the table shows that our measure of coagglomeration γ^C is centered on zero with a standard deviation of 0.008, a minimum of -0.043 and a maximum of 0.175. These figures are similar to the patterns in Faggio et al. (2017) – where we used 94 manufacturing industries (instead of 97) – and consistent with the findings of Ellison et al. (2010), who similarly report a right-skewed distribution of γ^C for the US displaying a higher co-location tendency.

The next three rows present descriptive evidence for our proxies for the Marshallian forces. The mean value for labor pooling is 0.225 with a standard deviation of 0.187, a maximum value of 0.968 and a minimum of -0.033. We find instead that the mean values of our input sharing and knowledge spillover proxies are much closer to zero – at 0.013 and 0.016, respectively – but the distributions have a strong right skew – with maximum values of 0.782 and 0.420, respectively. This confirms the pattern we found in Faggio et al. (2017) suggesting that most industries do not share intermediates or 'knowledge' – but a few are very highly interlinked.

The bottom half of Table 1 presents descriptive statistics on our proxies for industry-pairs' (dis)similarity in their use of natural and other non-manufacturing resources. The largest mean value is found for the dissimilarity in the use of natural resources (at 0.053), while the smallest relates to use of water (0.001). The other three measures have similar mean values at around 0.015-0.020.

The attributes we use to characterize industry are presented graphically in Appendix Figure 1. The top-left plot presents the Ellison-Glaeser Index (EGI) of agglomeration. Its mean value is 0.032 with a standard deviation of approximately 0.06. However, more than 40% of industries have values close to zero, and the distribution is clearly right skewed. Consistently, the EGI median is substantially smaller at 0.008. The top-right figure displays the distribution of the entry share of new firms, which has mean and median both at around 0.100, and a standard deviation of approximately 0.033. Next, the bottom-left plot shows that distribution of the industries' share of highly educated workers – with a mean/median of 0.099/0.078 and a standard deviation of 0.08. Lastly, the bottom-right plot presents the incumbent firms' size distribution. It should be noted that the figure excludes the sector 'Processing of nuclear fuel' (SIC233), which is clearly an outlier with 399 employees on average. Without this industry, the mean/median employment of incumbent firms is 23.7/19.2 with a standard deviation of 18.3.

III. Four classic agglomeration tales

In order to study the microfoundations of agglomeration economies, we link the proxies for the three standard Marshallian forces discussed above to our measure of industrial coagglomeration γ^C using the following empirical model:

$$\gamma^C_{ijt} = \alpha + \beta_{LP}\, LP_{ij} + \beta_{IO}\, IO_{ij} + \beta_{KS}\, KS_{ij} + \Sigma_{k=1}^{5}\, \lambda_k\, Diss_{ij}^k + \varepsilon_{ijt}, \qquad (2)$$

where γ^C_{ijt} is the Ellison et al. (2010) measure of coagglomeration between sectors *i* and *j* at time *t*; LP_{ij}, IO_{ij} and KS_{ij} denote the measure of labor-pooling (LP), input sharing (IO) and knowledge spillovers (KS) between sectors *i* and *j*; $Diss_{ij}^k$ is one of the measures of dissimilarity between sectors *i* and *j* in terms of use of primary resources and non-manufacturing inputs; and ε_{ijt} is an error term uncorrelated with all other variables. Throughout the analysis, we standardize variables to have unitary standard deviation at the level of the full dataset – i.e., when considering all manufacturing sectors. This eases comparison of the relative strength of the three Marshallian forces.

This approach is based on the idea that more coagglomeration between industry pairs will take place when the links between industries are stronger. Ellison et al. (2010, Mathematical Appendix) establish this formally in a particular model of agglomeration. O'Sullivan and Strange (2018) reach a similar conclusion in the context of an agent-based model.

We begin our analysis by estimating equation (2) including all manufacturing sectors in our data. In this case, the sample includes 4,656 industry pairs repeated over 12 years, giving rise to 55,872 observations. Standard errors are clustered at the industry-pair level.[9] Results are reported in the top row of Table 2 and confirm prior findings that all three Marshallian forces are significant determinants of co-location but labor pooling has a much stronger effect than input sharing and knowledge spillovers. In particular, the standardized effect of LP is approximately 10% – two and a half times the impact of IO (at 3.7%) and five times larger than the impact of KS (at 2%).[10]

It is worth emphasizing that this estimation is across the universe of industry pairs (as in Ellison et al., 2010, and as in some specifications in our prior paper, Faggio et al., 2017) rather than for individual industries – which is instead our focus here. In our previous paper, we thoroughly assessed the robustness of these findings – for example, by excluding London, by controlling for average population or employment of the TTWAs in which the industry pairs are located, or by accounting for the industries' own agglomeration – and established their causal nature. In the current work, we have carried out a similar

[9] Note that while there is time-variation in γ^C_{ijt}, our proxies for LP_{ij} IO_{ij} and KS_{ij} are fixed and averaged at the beginning of our observations window (1995-1999). Because standard errors are clustered at the sector-pair level, our results are equivalent to collapsing the dataset to one observation per industry pair – i.e., to 4,565 observations. However, we work with the expanded data set because in some extensions we investigate whether our results change if we only consider the first/second half of our time window, or exclude the last two years (corresponding to the 'Great Recession'). We found broadly comparable results irrespective of the exact years considered.

[10] One possible explanation for the weaker KS results is that knowledge spillovers are more difficult to define and measure that other Marshallian forces (see footnote 6).

(albeit less extensive) set of checks and reached similar conclusions: our findings most likely capture the causal impact of the three Marshallian forces on co-location – holding fixed other potential confounders. Although these results are not reported for brevity, they are available upon request.

We now turn to individual industry models, specifically for our four salient classic tales of agglomeration – namely, the textile (SIC171-SIC177), cutlery (SIC286), computer (SIC300) and automobile (SIC341) sectors. When considering microfoundations for these specific single sectors (or sets of sectors in the case of textiles), the empirical model in equation (2) is identified by exploiting variation in how one of these industries co-locates with the remaining 96 industries in the economy. Note that while these models allow for maximal heterogeneity, the results are noisier given the limits imposed on this approach by the data.

To begin with, the second row of Table 2 presents our evidence for the textile industry. This set of sectors has been of historical importance for the development of manufacturing in the UK (Landes, 1969). Our findings reveal very large and significant labor pooling effects at 0.367 – three and a half times larger than the average for the whole of manufacturing (at 0.101). We also find significant but smaller knowledge spillovers (at 0.143). Both estimates are significantly larger than the corresponding LP and KS for all other sectors in our data (excluding the textile group) with p-values on the null of no difference at 0.014 and 0.000, respectively. Conversely, the coefficient of input sharing is small (at 0.012) and insignificant. As for our sectoral characterisation, textile industries are more agglomerated (localized) than the average manufacturing sector (the EGI index is 0.081 vs. 0.032 across all manufacturing sectors), and have a less educated workforce than average (the share of college graduate is 0.050 compared to a manufacturing-wide average of 0.099). The sector also has average levels of new firms' entry and close-to-average size of incumbent firms.

The picture is different when considering another salient industry – namely cutlery. The industry was considered by Marshall (1890) who used it as a classic example of agglomeration driven by sharing of inputs and services. Indeed, he wrote that "many cutlery firms … put out grinding and other parts of their work, at piece-work prices, to working men who rent the steam power which they require, either from the firm from whom they take their contract or from someone else'' (Marshall, 1890, p. 172). The results in Table 2 support these intuitions. The coefficient on input linkages is very large and significant (at 0.599 – sixteen times larger than for the average manufacturing sector). Testing the equality of the IO coefficients between cutlery and all other sectors in our data leads us to reject the null of no difference with a p-value of 0.007. For cutlery, the IO coefficient is also twice as large as the coefficient on labor market pooling (at 0.238), which is also significant (and significantly different from the rest of manufacturing; p-value on the significance of the difference: 0.034). Conversely, knowledge spillovers are not significant and slightly negative. Clearly, these results show that one would not want to generalize from cutlery to textiles, and illustrate the perils of extrapolation. Testing whether the strength of IO (as well as KS) between cutlery and textiles is the same clearly leads us to reject the null – though from a statistical point of view LP is found to have the same importance in both sectors. In terms of its attributes, cutlery is characterised by average sectoral agglomeration, relatively low entry rates and low educated workers, further emphasising that even in terms of underlying organizational structure cutlery and textiles are not comparable – despite being both classic examples of the historical development of the UK industrial clusters. Finally, the small size of its incumbent firms – at 11 employees (40% of the manufacturing average) – coupled with the very large impact of the IO proxy provides support to the intuitions in Chinitz (1961), who emphasised the importance of input sharing among small firms as a driver of agglomeration.

Without doubt, the computer industry is the salient industry in the agglomeration literature. One example is Saxenian's (1994) highly impactful work on the Silicon Valley. Given previous discussion of this sector in the literature, it is no surprise that the regression results in Table 2 show a very large and significant coefficient on knowledge spillovers at 0.215 – ten times larger than for the average manufacturing sector (and statistically different from KS in the rest of the economy; p-value on the significance of the difference: 0.000). The input sharing coefficient is also positive but substantially smaller and non-significant (at 0.017 – half the size of the impact for manufacturing overall, though the difference is not statistically significant). The computer industry is somewhat unusual in displaying a slightly negative and insignificant coefficient on labor market pooling (at -0.058; statistically different from the rest of manufacturing). It is worth noting that the latter result does not mean that there is no labor market pooling in this industry. Instead, there could be significant labor pooling taking place within the computer sector itself. Note also that, although the computer industry is similar to cutlery in having small incumbents, the 'organizational' Chinitz-type IO effects are dominated here by the importance of knowledge effects. This is possibly due to its highly educated workforce and high entry share – features that distinguish this sector from the previous 'tale'. It is thus manifestly problematic to extrapolate from the cutlery industry to the computer industry. Similarly, however appealing it might be to use the computer industry to illustrate agglomeration economies in general, the logic of extrapolating from the computer industry is similarly strained.

The car industry is also highly salient in the agglomeration literature. In the US, this industry's declining cluster centered around Detroit is often contrasted to the prosperous computer cluster in Great San Jose. A very informative discussion along these lines can be found in Glaeser (2011). Somewhat surprisingly, our sectoral characterisation uncovers similarities between these two sectors in the UK. Looking at Table 2, the car industry has more highly educated workers than manufacturing on average – like the computer sector (though less markedly so) – and a high entry share. It is also not markedly more agglomerated than the average manufacturing sector (unlike textiles; although computers are clearly less agglomerated). The only remarkable difference with respect to the computer industry is that incumbents in the car sector are very large. Despite these broad similarities, the pattern of the regression coefficients differs. Labor pooling has a large and significant effect (at 0.291), while knowledge spillovers have a much smaller but still significant coefficient (at 0.100). Input sharing instead has a small insignificant impact (at 0.017). This evidence shows once again that even within sectors that share some features in terms of their organization and characteristics, the microfoundations of agglomeration can be different. Indeed, tests on whether LP and KS have the same impact across the two sectors reject the null – though the rejection is borderline for knowledge spillovers (largely due to the imprecision of the estimate for the car sector).

In a nutshell, the evidence on the four tales shows that agglomeration is very heterogeneous. In the next section, we further substantiate this claim by exploring the variation in the strength of the three Marshallian forces across all the 97 manufacturing sectors in our data.

IV. Individual industry models

In this section, we characterise the heterogeneity in the microfoundations of agglomeration more completely by estimating single-industry co-location models for all the manufacturing sectors covered by our data. Stated differently, we estimate the empirical model in equation (2) industry-by-industry – i.e., considering all 97 sectors we track and not just the four classic tales. As already discussed, these models are identified by the variation in the coagglomeration patterns of one industry with the remaining 96 in the economy.

The most important findings from this exercise are presented in Table 3 and Figure 1. The full set of estimates is presented in Appendix Table 1. While it is not possible to discuss every industry's pattern in the text of the paper, we believe these results have the potential to be of interest to a wide range of scholars and policymakers.

The first result that emerges is the striking heterogeneity in the strength of the Marshallian forces across manufacturing industries as clearly displayed in the panels of Figure 1. This heterogeneity is not only visually sizeable but is also statistically significant: F-tests on whether the LP, IO and KS estimates are identical across sectors clearly reject the null (similarly, F-tests for the joint significance of the three sets of Marshallian forces reject the hypothesis that they are jointly equal to zero).

Looking at estimates for LP across all the industries, we find a mean estimated effect of 0.138,[11] but a substantially smaller median impact at 0.059 (see top panel of Table 3). This difference is due to a spread-out distribution of estimates (standard deviation of 0.268) with an evident stretch towards positive values (skew 1.44). The top-right plot of Figure 1 further reveals that the estimated effect distribution easily covers the (-0.5,0.5) interval but stretches well above this range on the positive side of the horizontal axis – reaching values above one (i.e., a unitary standardized effect). However, not all of these estimates are significant. The bottom right plot of the figure also reveals that the associated standard errors – though mainly concentrated in the (0,0.25) interval – are relatively spread out giving rise to approximately 36% (35 out of 97) significant estimates.

It is also interesting to reflect on the nature of the 'top four' and 'bottom four' sector with the highest/lowest LP estimates. All industries in the top four belong to the broad textile sector – which we already analysed as one of our classic tales. The bottom four sectors are instead very different from each other. 'Manufacture of musical instruments' (SIC363) and 'Cutting, shaping, and finishing of stone'

[11] Note that the average of the Marshallian force effects estimated industry-by-industry does not necessarily coincide with the corresponding effects estimated by pooling data for all industries – i.e., for the average sector.

(SIC267) are likely to have highly specialized workers – a situation where pooling with another industry may not be possible. In contrast, 'Manufacture of games and toys' (SIC365) and 'Manufacturing of vegetable and animal oils and fats' (SIC154) seem likely to involved standardized labor, where pooling with specific partner industries may not be needed. Interestingly, none of these sectors has a highly educated workforce – with shares of college graduate ranging between 0.03 and 0.11.

Regarding IO, we find a much smaller average effect at 0.056 and an even smaller median impact at 0.017 (see Table 3). This reflects the fact that nearly 40% of the sectors have IO effects very close to zero (see Figure 1). On the other hand, the distribution is significantly spread out (std.dev. 0.251) with an even more pronounced right skew than LP (2.905 vs. 1.444). This lends support to our previous claim that the majority of sectors are not tightly related via input-output linkages – but some industries are clearly very interconnected. For these industries, IO is a very strong determinant of co-location. We also find relatively stretched out standard errors – once again giving rise to 35 (out of 97) significant estimates.

We further note that one of our classic tales – cutlery – is among the 'top four' sectors with the largest IO estimates. The other three are two industries in the same sector division – i.e., the manufacture of ceramics (SIC262 and SIC263) – and 'Manufacture of coke oven products' (SIC231). While the last seems at first only loosely associated with the previous two, it turns out it is not: this sector uses the output of the ceramic industry for the maintenance and repairs of its oven (in a process called 'ceramic welding'). It is also interesting to note that the ceramic compartment is arguably another classic tale. For example, the Italian district of Sassuolo has been known for being a world-leading ceramic-production cluster for decades (Porter, 1990). Similarly, in England, Stoke-on-Trent – and more generally Staffordshire – hosts highly concentrated and specialised pottery and ceramic-related productive activities. Conversely, little stands out when we look at the 'bottom four' sectors – although for two of them (SIC181 and SIC183, relating to the processing of leather and fur) we would have expected higher IO effects (considering that they might feed into the textile and fashion manufacturing).

Lastly, we find much smaller mean and median values for the estimated effects of KS – at 0.018 and 0.011, respectively (see Table 3). The distribution is also less spread out (std.dev. 0.089) than for LP and IO, and has a small, negative skew of -0.339. Yet, visually the distribution is clearly much more symmetric than what we found for LP and IO (see Figure 1) with values mostly concentrated in the (-0.2,0.2) interval and just above 25% of the industries displaying KS effects very close to zero. As a result of this relative compression and associated standard errors, we find that only approximately 16% of the estimates (16 out of 97) are significant.

When looking at the 'top four' KS sectors, we find that one of the classic tales – i.e., the computer industry – clearly ranks very high in terms of its responsiveness to knowledge flows. More surprisingly, we also find 'Preparation/spinning of textile fibres' (SIC171) among the 'top four', and two sectors in the ceramic-and-related compartment – namely 'Manufacture of ceramic goods other than for construction' (SIC262) and 'Manufacture of cement, lime and plaster' (SIC265). The former also has the highest IO coefficient while the latter – despite belonging to the same compartment – has a small and negative IO estimate (-0.15). Interestingly, neither sector has a high share of college graduates (0.057 and 0.107, respectively) making them different from the computers (with a 0.294 share of college graduates). Nevertheless, all three industries significantly respond to the benefits of sharing knowledge. At the opposite end, the 'bottom four' sectors form a disparate group which includes a high skilled sector – 'Processing of nuclear fuel' (SIC233; share of graduates 0.317) – and a very unskilled one – 'Manufacture of leather clothes' (SSIC181; with zero college graduates).

Next, we carry out an attempt at identifying industries that resemble the four classic tales discussed above in terms of the relative strength and significance of their estimated microfoundations. To do so, we proceed as follows: *i*- we sort the data contained in Appendix Table 1 on the basis of the strength of the Marshallian force that best identifies a given tale – for example, on the basis of the IO effect which, at 0.599, characterises cutlery; *ii*- we focus on a relative tight neighbourhood around the estimate of the force that characterises the tale – i.e., we focus on IO values two standard errors up or down from 0.599; *iii*- we mainly consider sectors that report a *statistically significant* coefficient *within the identified range* for the Marshallian force under consideration (e.g., IO for cutlery); and *iv*- we identify industries that resemble the tale under investigation on the basis of the other two forces – e.g., they are similar to cutlery along LP and KS in terms of both strength and statistical significance (bearing in mind that we are already focussing on industries with similar input-sharing effects by selecting industries with significant IO estimates around 0.599). While this is not an exact approach, it reveals potentially useful insights in terms of possible similarities between our classic cases and other manufacturing industries.

Starting with textiles, we basically find that no other sector reproduces the kind of pattern that characterises this industry (SIC171-SIC177; LP=0.367, significant; IO=0.012, insignificant; KS=0.143, significant). If anything, there is some heterogeneity within the textile group when considering its various sub-sectors. For example, 'Preparation/spinning of textile fabrics' (SIC171), 'Textile weaving' (SIC172), 'Manufacture of knitted and crocheted fabrics' (SIC176) and 'Manufacture of knitted and crocheted articles' (SIC177) are relatively similar with high estimate LP effects and still sizeable but smaller KS effects. However, SIC176 has a relatively low KS (at 0.045, still more than double the 0.02 manufacturing average) while SIC171 has a positive and significant impact of input sharing (at 0.095). Furthermore, 'Manufacture of textile articles, except apparel' (SIC174) displays a very different pattern with a very large effect of IO (at 0.169), but no impact for LP (0.029) and KS (0.0049).

When focussing on the pattern for cutlery (SIC286; LP=0.238, significant; IO=0.599, significant; KS=-0.039, insignificant), we find some similarities with 'Pressing of iron and steel' (SIC273; LP=0.329, significant; IO=0.418, significant; KS=-0.032, insignificant) and 'Manufacture of ceramic tiles' (SIC263; LP=0.376, significant; IO=1.109, significant; KS=0.106, insignificant). While the former resembles cutlery in terms of its core production processes, the latter is in a very different compartment – and in fact displays too large a coefficient on KS. We also find some similarities with 'Forging and pressing' (SIC284; LP=0.352, significant; IO=0.327, significant; KS=-0.077, insignificant) – clearly positioned in a closely related industry – although the relative strength of LP and IO in this sector makes it somewhat different from cutlery (where IO clearly dominates).

Next, we look for similarities to the computer industry (SIC300; LP=-0.058, insignificant; IO=0.017, insignificant; KS=0.215, significant), but struggle to find any. The closest sectors we identify are 'Manufacture of accumulators, cells and batteries' (SIC314, LP=0.007, insignificant; IO=-0.039, insignificant; KS=0.141, insignificant) and 'Manufacture of pesticides' (SIC242; LP=-0.007, insignificant; IO=0.045, insignificant; KS=0.118, significant) – although for the former the impact of KS is not significant despite being the dominant force.

Finally, when we hone in on cars (SIC341; LP=0.291, significant; IO=0.017, insignificant; KS=0.100, significant), we find no other sector that displays a similarly large and significant LP effect while also having a positive and significant impact of KS and no effect (neither positive nor negative) of IO. If anything, the industries with pattern closest to automobiles are in the textile compartment (for example, SIC177-'Manufacture of knitted and crocheted articles') although these tend to display substantially larger LP effects and insignificant KS.

We conclude this section by investigating the correlations between the three Marshallian forces across the industrial sectors in our sample. Our findings are presented diagrammatically in Figure 2. This displays linear prediction from univariate regressions of one of the three Marshallian forces on the other two – one at the time (for example LP on IO and then LP on KS). The figure shows that sectors with high LP also tend to have high IO – but these same sectors tend to have low KS. Conversely, we find that the association between IO and KS is positive.[12]

All in all, the evidence from this section confirms our previous conclusions. The forces that govern agglomeration are very heterogeneous. Extrapolation from salient cases to other sectors should be carried out carefully, as should 'interpolation' from regressions that pool all manufacturing sectors to characterise the behaviour of specific industries. We believe that the individual industry models have the potential to guide this sort of analysis: a local planner interested in promoting the emergence of a cluster in a given industry should be especially careful in acting on lessons learned from another industry with very different microfoundations. Put the other way, the planner should only attempt to extrapolate from industries that are similar in their agglomeration tendencies.

In the next section, we try to systematise the patterns we have found so far by relating the strength of the Marshallian forces in different sectors to some of their underlying characteristics in ways that are informed by the theories of agglomeration.

[12] When pairing up Marshallian forces, we used the force with the smallest amount of variation as right-hand side variable to guarantee that the predictions plotted in the graph cover the actual variation taken by this force (and do not 'predict' out-of-sample). Unsurprisingly, the graphs display the same tendency when we run regressions swapping right- and left-hand side variables though the actual slopes are different.

V. Understanding the patterns

A. Theoretical foundations and empirical approach

This section considers the patterns of microfoundations that are present in the individual industry models documented in Sections III and IV and develops a simple approach to systematize our findings. This involves estimating models of the relationship between four key industry characteristics – namely, localization, new firm entry, workforce education and size of incumbent firms – and the estimated individual industry coefficients on Marshallian microfoundations that we discussed above. This approach builds on the analysis in Faggio et al. (2017), where the characteristics of industry pairs were related to Marshallian coefficients estimated across the universe of industry pairs. The key difference between the current work and our previous analysis is that here we consider individual industries – as opposed to industry pairs. This delivers direct evidence on the correlates of one industry's microfoundations and generates insights that can be used to guide policy.

In order to conduct our investigation, we focus on some fundamental questions that are theory-grounded and related to the fundamental nature of agglomeration forces. First, we investigate whether coagglomeration is a substitute or a complement to localization – namely the 'own' agglomeration of an industry. This question is related to the old 'urbanization' vs. 'localization' issue, where the focus is whether agglomeration economies depend primarily on the scale of the entire city (urbanization) or that of an individual industry (localization). The older contributions in this literature employ cross-sectional reduced form approaches to looking at which effect is strongest. See, for instance, Glaeser et al (1992) and Henderson et al (1995), or the Rosenthal-Strange (2004) survey.

On the microfoundations side of this debate, some research shows effects that appear to operate within industries (e.g., Fallick et al., 2006), while other research shows effects operating between industries (e.g., Ellison *et al.*, 2010). Theoretically, it is straightforward to conceive of a model where both effects are at work. In such situation, the agglomeration of an industry may be a substitute or a complement to the coagglomeration of the industry with other sectors. The substitution effect of agglomeration would argue as follows: if the presence of own industry activity creates an external increasing return within the industry, then cross-industry coagglomeration is not as valuable. The potential complementarity argument would instead suggest that industries that benefit from Marshallian forces will seek to enjoy these benefits both by coagglomerating with other industries and by locating with other own-industry firms. Helsley and Strange (2002) provide a model where there is a potential for both substitute and complement relationships of this sort. In sum, there are theoretical arguments – and some empirical evidence – that both complementarity and substitution can be at work. We will further consider this issue by relating an industry's Ellison-Glaeser (1997) Index (EGI) of agglomeration to the industry-level coefficients that capture the strength of labor pooling, input sharing, and knowledge spillovers.

Second, we study how industry dynamism relate to the agglomeration microfoundations. Vernon (1960) argues that the distinction between stable and unstable industries is key to understanding the nature of increasing-returns productive activities. In Vernon's view, the dynamism found in unstable industries serves to strengthen microfoundations. This result is a clear comparative static in a range of models. For instance, in the Helsley and Strange's (1990) model of labor market matching, more instability would be reflected in a greater loss associated with poorly matched employers and workers. This would in turn raise the marginal benefit of market thickness, implying stronger agglomeration economies. Similarly, in Duranton and Puga's (2001) model of nursery cities, agglomeration is more valuable at the prototype stage than when the product is in ordinary production. In both cases, a dynamic industry will benefit more from coagglomeration with related industries. On the other hand, dynamism might instead weaken microfoundations. Helsley and Strange (2004) show that repeated interactions are needed to get knowledge sharing and, by extension, other microfoundations. To the extent that more dynamic industries make it less likely that interactions are repeated, this suggests that industry dynamism might be negatively associated with the strength of Marshallian forces. In short, the effect of dynamism could go either way and the relationship between an industry's dynamism and the strength of Marshallian agglomeration forces is an empirical question. In the analysis below, we proxy dynamism with entry share and we explore the relationship between the incidence of new firms and the estimated strength of the Marshallian forces at the industry level.

The third question that we consider is the relationship of workforce education to agglomeration. It is common to equate education with skill. Bacolod et al. (2009a, 2009b, 2010) show that this is somewhat misleading: education is an input in skills, but there is not a one-to-one relationship between the two. Skill is a heterogeneous concept, with cognitive and social skills more strongly related to education than the physical skills that dominate the skilled trades – like Marshall's cutlery workers, discussed earlier in the paper. If educated workers have more specialized skills, then labor pooling effects might be stronger in sectors with a more educated workforce. Nonetheless, since education is at best an imperfect proxy for skills, this relationship might not hold. Similarly, input sharing is also sometimes seen as operating especially strongly for high technology products (Porter, 1990), which might also mean that input sharing is stronger in industries with more educated workers. Having said this, there is no reason why input sharing could not apply to low technology products – suggesting that the relationship between the average education of an industry's workforce and input sharing could go either way. Similarly, although a worker must know something in order to have knowledge that might spill over, there is ambiguity: Marshall's cutlery workers – while clearly skilled – almost certainly did not hold university degrees. In order to consider these issues empirically, we investigate the relationship between education – specifically the share of graduates in the industry's workforce – and industry-by-industry Marshallian coefficients estimated in Section IV.

Fourth and finally, we explore how firm size is related to microfoundations of agglomeration. Again, prior research establishes the possibility of large firms either discouraging or encouraging agglomeration. On the one hand, Chinitz's (1961) classic paper argues that small firms have larger effects – in particular, by fostering input sharing linkages. See Rosenthal-Strange (2003, 2010) for empirical results consistent with this idea. On the other hand, other empirical work (see among others Feldman, 2003 and Agrawal and Cockburn, 2003) show 'anchor' effects whereby large firms have important externalities. In the analysis that follows, we reassess these questions by studying the relationship between an industry's mean employment of incumbent firms – i.e., those already in the market – and that industry's Marshallian coefficients.

The estimating equation we use to implement these ideas takes the following very simple form:

$$\beta_{ai} = \Sigma_j \delta_j X_{ji} + \varepsilon_i, \qquad (3)$$

where β_{ai} gives the value of the coefficient for Marshallian force a – with a $\in$ {LP, IS, IO} – for industry *i* (which we estimated using equation 2 sector-by-sector), while X_{ji} represents the value of the industry characteristic *j* – with j $\in$ {EGI, Entry, Education, Incumbent} – for industry *i*. We start by estimating univariate models where we enter one of these characteristics at a time, and then present multivariate models including all industry characteristics together. Furthermore, we also estimate simple linear models like those indicated by equation (3), as well as non-linear models in which dummies that capture quantiles of the underlying sectoral attributes are used to characterize industries. In the latter case, we estimate models where we regress the Marshallian coefficients on dummies for observations in the top 10% and bottom 10% of the various industry characteristics.

For all models, our preferred approach is to present and discuss estimates that come from specifications where industries are weighted by the inverse of the standard error of the Marshallian coefficients. This approach means more weight is placed on industries where Marshallian forces are estimated with greater precision, while our results are not 'pulled' by outlier industries with potentially large but far from significant estimates of LP, IO and KS. This correction is similar to the one routinely used in meta-analysis – where studies are generally weighted on the basis of the underlying sample size or the variance of the variables under consideration. As noted by Borenstein et al. (2009) the two approaches are almost equivalent – given that the (inverse of the) variance is proportional to the sample size. Since in our analysis all sectors occur an identical number of times in the industry-by-industry regression spelled out in equation (2) and so contribute in the same way to the estimation of the industry-specific LP, IO and KS, we cannot weigh by sample size. Instead, we weight by the inverse of the standard error – which we find is an intuitive way to account for the precision of our estimates given that significance levels are conventionally established by looking at the coefficient-to-standard error ratio (i.e., the t-test).[13] We also estimated unweighted models, which are reported in the Web Appendix (Table A2.1). While the results from this second approach yield similar intuitions, the estimates are noisier and the patterns less clear. However, unweighted models do not account for the fact that large estimated Marshallian coefficients need not to be statistically significant – and so should be 'discounted' in our analysis. So we consider this

[13] We also experimented with weights inversely proportional to the variance of our estimates. This approach returned similar patterns, but it was overly 'aggressive' in that it heavily penalized even coefficients that were quite precisely estimated.

approach less reliable. Finally, in all specifications we adjust standard errors for heteroscedasticity by using a standard 'robust' variance-covariance matrix correction.

B. Results

Table 4 presents results of weighted univariate regressions. For each Marshallian force, the first column reports the results of a continuous model, while the second column gives results of a model which includes dummies identifying observations in the top-10% and bottom-10% of the distribution of a given sector characteristic, as described above.

The top panels of the table address whether agglomeration and coagglomeration are substitutes or complements. The results suggest the latter is more likely the case. The labor pooling coefficient rises with the degree to which an industry is agglomerated according to the EGI measure. So does the knowledge spillover coefficient. The complementarity result holds for both the continuous measure and for the dummy approach. For input sharing, however, the results are weaker. The estimate in the continuous model is close to zero. However, the bottom-10% dummy is significantly negative – which is consistent with agglomeration and coagglomeration being complements.

The second set of results in Table 4 concerns industry dynamism as proxied by the entry share. With regard to labor pooling, there is a positive and significant relationship between the presence of new firms and the LP coefficient. The bottom-10% dummy is significant and negative, suggesting that the positive relationship is driven, at least in part, by the least dynamic sectors (i.e., those with the lowest rate of entry). Krugman (1991a; 1991b) showed that labor pooling can increase productivity in part by reducing unemployment when a city's employers experience labor demand shocks that are not perfectly correlated. The finding here seems similar in the sense that industries with a lot of entry (and possibly exit) exhibit LP to a greater extent. On the other hand, we fail to find a statistically significant relationship between dynamism and input sharing or knowledge spillovers. One explanation would be that input and knowledge relationships take longer to form. Alternatively, this could be the result of the ambiguities in the theoretical predictions discussed above – with 'negative' and 'positive' forces cancelling out and leaving it impossible to form tight predictions on the likely strength of these microfoundation forces on the basis of this sector specific characteristic.

The third set of results in Table 4 concerns the share of educated workers. The sharpest results we find are for knowledge spillovers. Industries with a high share of college graduates have larger KS coefficients. The dummy model (final column) shows that the top-10% dummy coefficient is positive and significant, suggesting that the relationship may be driven by the sectors with the very highest education levels. The coefficient on the top-10% dummy is also significant for input sharing. The continuous specification for input sharing, however, shows a positive but insignificant coefficient. Lastly, for LP, the bottom-10% dummy is positive and significant. This somewhat puzzling result echoes a similar finding in Faggio et al (2017), where low-education industry pairs showed a greater degree of labor pooling. In the current paper, we see that industries with the very least educated workers have the largest coefficients for labor pooling. This presumably reflects labor pooling operating strongly outside of sectors with highly educated workers. This is consistent with the argument above that education is not identical to skills.

Finally, the bottom panel of Table 4 presents results that focus on the size of incumbent firms. The results on input sharing show some parallels with Chinitz (1961) – though they are not especially strong. The negative and significant top-10% dummy means that the sectors with the largest incumbents have the least input sharing. This is consistent with large firms being more weakly linked to their local supply chains, as in Chinitz. Having said this, it is worth noting that the coefficient on incumbent employment in the continuous model is positive, though very small and insignificant. With regard to knowledge spillovers, the bottom-10% coefficient is negative and significant. This suggests that industries with the smallest incumbents have the smallest knowledge spillover coefficients. This is consistent with the anchor hypothesis offered by Feldman (2003) and Agrawal-Cockburn (2003). Finally, for labor pooling, we see a negative and significant coefficient on the bottom-10% dummy, consistent with the industries with the smallest firms showing the least LP. This could be explained by an organizational dimension of labor pooling: large firms can expand when their rivals are hit with negative shocks.[14]

So far, we have presented the results from the perspective of industry characteristics. It is however instructive to do the reverse, and consider the results from the perspective of Marshallian forces (i.e., columns rather than rows). It is clear that labor pooling is important for agglomerated industries, and especially dynamic ones. It appears to be strongest for the least educated workers and weakest for sectors with the smallest firms. There is some evidence that input sharing is most important for agglomerated industries. Furthermore, it becomes less important when incumbents are very large. It is also strongest for the most educated industries. Finally, knowledge spillovers are strongest for agglomerated industries with educated workers and weakest for the industries with small incumbents.

The models presented in Table 4 give the results of univariate estimation. Table 5 presents results for multivariate specifications. The results on the complementarity of agglomeration and coagglomeration continue to hold. Similarly, the results on worker education and knowledge spillovers and labor pooling are also fairly robust, while the association between share of college graduates and input sharing obtained in Table 4 is significantly weakened (the estimates point in the same direction, though they are smaller and associated with bigger standard errors). The results on the two industrial organization variables – namely, entry share and size of incumbents – are somewhat different now. Starting with entry share, the associations retain their signs for all three Marshallian forces but the estimated magnitudes are smaller and clearly not significant (with the exception of the coefficient on the bottom-10% dummy for input sharing which increases in size and turns significant, indicating that input sharing is important for less dynamic industries). Regarding the second, the significant associations between the size of incumbent firms and

[14] We also studied the association between the strength of the Marshallian forces and the age of the sector (measured as the difference between the last year in our data and the year in which the oldest firm in the industry was established; as an example, for computers-SIC300 this would be 2007-1961=46) but failed to find any striking patterns. Results are presented in Table A2.2 in the Web Appendix.

LP/IO/KS we observed in Table 4 are somewhat replicated in the multi-variate models of Table 5 – although the estimates lose some of their size and thus become insignificant.

While the results in Table 5 are an important check on the univariate associations presented in Table 4, the small number of observations in our analysis and some relatively strong patterns of correlation between our four industry attributes imply that there is a risk that collinearity causes the multi-variate estimates to lose precision. For example, the correlation between the share of skilled workers and the size of incumbent firms is 0.3697 (significant at the 5% level), while the share of graduates displays a 0.2028 correlation (5% significant) with the entry share and a negative -0.2116 association (5% significant) with the EGI index.[15] These patterns suggest the findings reported in Table 4 might be preferred.

Notwithstanding, the bottom line of our analysis is that we find a robust result on the relationship between agglomeration and coagglomeration. Industries that appear to benefit from the latter also seem to benefit from the former. This is true for all three Marshallian forces. For instance, an industry that agglomerates or clusters also tends to coagglomerate with other industries with a similar mix of occupations. We also find robust results on education, with industries with educated workforces tending to coagglomerate more with industries that are linked in innovation through patent citations. Industries with less educated workforces seem to show more tendency to labor market pooling. Finally, regarding industrial organization, univariate models show that dynamic industries see stronger labor pooling, while industries with large incumbents are less sensitive to input links to other industries. These results are however less robust to multivariate specifications.

VI. Conclusions

This paper employs UK data to consider the microfoundations of agglomeration economies. Using the variation in the other industries with which a given industry co-locates, we estimate the importance of Marshallian labor pooling, input sharing, and knowledge spillovers at the level of the individual industry.

The results support Marshall's analysis of agglomeration in a specific sense: each of the forces is shown to play an important role in the co-location patterns for a number of industries. However, the forces are not universal – something which Marshall himself never claimed to be the case. Some industries co-locate with other industries that have similar workforce needs. Others instead co-locate with industries to which they are linked via supply chains or in knowledge.

These findings are important for the understanding of the forces that drive agglomeration. The heterogeneity in the nature of the agglomeration process was noted previously by Faggio et al (2017). This previous paper looked at heterogeneity at the level of pairs of industries. The present paper, in contrast, provides evidence of heterogeneity in microfoundations at the level of the individual industry. The paper offers robust evidence that agglomeration is a complement to coagglomeration rather than being a substitute: an industry that co-locates with other industries linked in a particular way (e.g., in technology

[15] The figures refer to un-weighted correlations. Correlations weighted by the inverse of the LP, IO and KS standard errors (as in the regressions of Tables 4 and 5) provide similar intuitions.

and knowledge) will also have a tendency to cluster (which presumably gives additional valuable technological links). The paper further shows that an industry's dynamism, incumbent firm size, and worker education contribute to the pattern of heterogeneous microfoundations. Our strongest results are that industries with high levels of entry display high coefficients on labor pooling and that industries with high levels of worker education have larger coefficients on knowledge linkages.

These results have the potential to be important for policy design. It is natural, of course, for a policymaker interested in local economic development to make use of the experiences of other planners in other locations. As a general matter, the individual industry models show the peril of extrapolation from a one-industry agglomeration case to the larger phenomenon of agglomeration. Different industries manifestly differ in the importance of Marshallian forces, and a policy that is helpful to one industry may not be helpful to another. Making matters more concrete, our results clearly show that devising a policy based on the lessons of the computer industry in order to make an area attractive to automobile producers will most likely not be successful. Nevertheless, the individual industry models do have more positive implications for policy: extrapolation will more likely be on-target if the industries considered are similar – something that can be assessed using the paper's results.

Our results similarly show that policy makers should exercise caution when using results from pooled industry regressions to understand the microfoundations of agglomeration for specific industries. The substantial variation in microfoundations means that pooled industry regressions offer too blunt a tool for identifying an individual industry's reasons for clustering. Evidence based on single-industry models as those described in this paper can, instead, provide important insights on one industry's agglomeration patterns either by exploring the behaviour of the same industry in other locations or by investigating the behaviour of a set of industries that share similar characteristics with the industry in question.

References

Agrawal, A., and Cockburn, I., 2003. The anchor tenant hypothesis: exploring the role of large, local, R&D-intensive firms in regional innovation systems. *International Journal of Industrial Organization*, 21(9), 1227-1253.

Bacolod, M., Blum, B. S., and Strange, W. C., 2009a. Skills in the city. *Journal of Urban Economics*, 65(2), 136-153.

Bacolod, M., Blum, B. S., and Strange, W. C., 2009b. Urban interactions: soft skills versus specialization. *Journal of Economic Geography*, 9(2), 227-262.

Bacolod, M., Blum, B.S., and Strange W.C., 2010. Elements of Skill: Traits, Intelligences, and Agglomeration. *Journal of Regional Science*, 50(1), 245–280.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R., 2009. *Introduction to Meta-Analysis*. London, UK: Wiley.

Breschi, S. and Lissoni, F., 2004. Knowledge networks from patent data: methodological issues and research targets. *CESPRI Working Paper 150*.

Breschi, S., Lissoni, F., and Montobbio, F., 2005. The geography of knowledge spillovers: Conceptual issues and measurement problems. In S. Breschi, F. Malerba (eds.), *Clusters, Networks and Innovation*, Oxford: Oxford University Press, 343-378.

Chatterji, A., Glaeser, E., and Kerr, W., 2014. Clusters of Entrepreneurship and Innovation. Innovation Policy and the Economy 14, 129-166.

Chinitz, B., 1961. Contrasts in agglomeration: New York and Pittsburgh. *American Economic Review* 51, 279-289.

Devereux, M. P., Griffith R., and Simpson, H., 2007. Firm location decisions, regional grants and agglomeration externalities. *Journal of Public Economics*, 91, 413-435.

Duranton, G., 2011. California Dreamin': The Feeble Case for Cluster Policies. *Review of Economic Analysis*, 3, 3-45.

Duranton, G. and Puga, D., 2001. Nursery cities. *American Economic Review* 91(5), 1454-1477.

Duranton, G. and Puga, D., 2004. Micro-foundations of urban agglomeration economies. In Henderson, J.V. and Thisse, J.-F. (eds.), *Handbook of Urban and Regional Economics, Volume 4*, Amsterdam: Elsevier, 2063-2117.

Ellison, G. and Glaeser, E.L., 1997. Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy* 105, 889-927.

Ellison, G., Glaeser, E. L. and Kerr, W. R., 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review* 100, 1195-1213.

Fallick, B., C. Fleischman and J. B. Rebitzer, 2006. Job hopping in the Silicon Valley: The micro-foundations of an industry cluster. *Review of Economics and Statistics* 88, 472-481.

Feldman, M., 2003. The locational dynamics of the US biotech industry: knowledge externalities and the anchor hypothesis. *Industry and Innovation*, *10*(3), 311-329.

Faggio, G., Silva, O. and Strange, W.C., 2017. Heterogeneous agglomeration. *Review of Economics and Statistics*, 99(1), 80-94.

Glaeser, E.L., 2011. *The Triumph of the City, How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier and Happier,* New York, NY: Penguin US.

Glaeser, E.L., H. Kallal, J. Scheinkman, and A. Schleifer, 1992. Growth in cities. *Journal of Political Economy*, 100, 889-927.

Helsley, R. W., and Strange, W. C., 1990. Matching and agglomeration economies in a system of cities. *Regional Science and urban economics*, *20*(2), 189-212.

Helsley, R. W., and Strange, W. C., 2002. Innovation and input sharing. *Journal of Urban Economics*, *51*(1), 25-45.

Helsley, R. W., and Strange, W. C., 2004. Knowledge barter in cities. *Journal of Urban Economics* 56 (2), 327-345.

Henderson, J.V., Kuncoro, A. and Turner, M., 1995. Industrial development in cities. *Journal of Political Economy,* 103, 1067-1085.

Krugman, P., 1991a. Increasing returns and economic geography. *Journal of Political Economy* 99, 483-499.

Krugman, P., 1991b. *Geography and Trade*. Cambridge: MIT Press.

Landes, D. S., 1969. *The Unbound Prometheus*. Cambridge: Cambridge University Press.

Marshall, A., 1890. *Principles of Economics*. MacMillan: London.

O'Sullivan, A. and Strange W. C., 2018. The Emergence of Coagglomeration. *Journal of Economic Geography* 18, 293-317.

Porter, M., 1990. *The Competitive Advantage of Nations*. New York: The Free Press.

Rosenthal, S. S., and Strange, W. C., 2001. The determinants of agglomeration. *Journal of Urban Economics* 50, 191-229.

Rosenthal, S. S. and Strange, W. C., 2003. Geography, industrial organization, and agglomeration. *Review of Economics and Statistics* 85, 377-393.

Rosenthal, S. S. and Strange, W. C., 2004. Evidence on the Nature and Sources of Agglomeration Economies. In J. V. Henderson and J.-F. Thisse (eds.), *Handbook of Urban and Regional Economics, Volume 4,* Amsterdam: Elsevier, 2119-2172.

Rosenthal, S. S. and Strange, W. C., 2010. Big Establishment/Small Effects: Agglomeration, Industrial Organization, and Entrepreneurship, Chapter 10. In E. L. Glaeser, (ed.), *Agglomeration Economics - NBER Collections,* Chicago, IL: The University of Chicago Press, 277-302.

Saxenian, A.-L., 1994. *Regional Advantage: Cultural Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Silverman, B.S., 2002. *Technological Resources and the Logic of Corporate Diversification*. London, UK: Routledge.

Vernon, R., 1960. *Metropolis 1985*. Cambridge, MA: Harvard University Press.

Main Tables

Table 1: Descriptive statistics – Estimation sample for coagglomeration models

	Mean	Std. Dev.	Min.	Max.
Coagglomeration measures and Marshallian forces				
TTWA total employment coagglomeration (γ^C)	0.000	0.008	-0.043	0.175
Labor pooling (correlation)	0.225	0.187	-0.033	0.968
Input-output sharing (maximum)	0.013	0.044	0.000	0.782
Knowledge spillovers (maximum of inward/outward citation)	0.016	0.037	0.000	0.420
Additional Controls				
Energy dissimilarity index	0.015	0.018	0.000	0.099
Water dissimilarity index	0.001	0.001	0.000	0.006
Transport dissimilarity index	0.014	0.017	0.000	0.078
Natural Resources dissimilarity index	0.053	0.097	0.000	0.367
Services dissimilarity index	0.020	0.019	0.000	0.102

Note: Number of observations: 55,872. The sample contains non-repeated pairwise combination of 97 manufacturing SIC1992 3-digit industries over 12 years (1997-2008). The following sectors are not considered: Manufacturing of Tobacco Products (SIC160) because of a small number of plants throughout the period (43); Reproduction of recorded media (SIC223); Manufacturing of machine tools (SIC294), Manufacturing of weapons & ammunition (SIC296), Recycling of metal waste & scrap (SIC371) and Recycling of non-metal waste & scrap (SIC372) because of missing data on knowledge flows as measured by patent citations.

Sources: The coagglomeration index is computed using the ONS UK Business Structure Database 1997-2008. Labor correlation indices are computed from the UK Labour Force Survey 1995-1999. Input-Output measures are calculated using ONS UK Input-Output Tables for 1995-1999. Knowledge spillover measures are calculated using the UK data retrieved from the EPO-PATSTAT dataset made available to us by Bocconi University. Cited patents sampled for the years 1978 to 1997. Citing patents sampled for the years 1981 to 2000. Additional control measures are calculated using the UK Input-Output tables for 1995-1999.

Table 2: Coagglomeration and Marshallian forces – Whole Economy and Selected single-industry models

Sector Description	SIC Code	Effect of LP	Effect of IO	Effect of KS	EGI	Entry Share	Share Highly Educated	Incumbent Employment Size
All	All	.1014***	.0366**	.0199**	0.0321	0.1047	0.0986	27.61
		(.0144)	(.0149)	(.0092)				
Preparation, weaving & finishing of textiles	171-177	.3672***	.0119	.1426***	0.0810	0.1084	0.0505	20.83
		(.1126)	(.0566)	(.0349)				
Mfg. of cutlery, tools & general hardware	286	.2377***	.5987***	-.0392	0.0379	0.0774	0.0375	11.22
		(.0681)	(.2204)	(.0514)				
Mfg. of office machinery & computers	300	-.0577	0.0174	.2150***	0.0066	0.1658	0.2938	10.69
		(.0844)	(.0187)	(.0518)				
Mfg. of cars engines & bodies for vehicles	341	.2914***	.0172	.1005*	0.0451	0.1664	0.1029	34.58
		(.0658)	(.0470)	(.0536)				

Note: Regression coefficients come from single-industry regressions that exploit the variation in the coagglomeration of the industry in question with other industries (mutually exclusive pairs only) over twelve years. Number of observations as follows. All sectors: 55872. Textiles (SIC171-177): 7812. Cutlery (SIC286), Computers (SIC300) and Cars (SIC241): 1152. Standard errors clustered at the industry pair level. The Ellison-Glaeser Agglomeration Index (EGI) reported is an average across industries and years.

Table 3: Summary statistics for the Marshallian forces

	Mean	Median	SD	Skewness	Top 4	Top 4 industry description	Bottom 4	Bottom 4 industry description
Labour pooling	0.1380	0.0587	0.2677	1.444	SIC176	Mfg. of knitted & crocheted fabrics	SIC363	Mfg. of musical instruments
(LP)					SIC171	Preparation/spinning of textile fibres	SIC267	Cutting, shaping & finishing of stone
					SIC173	Finishing of textiles	SIC154	Mfg. of vegetable and animal oils and fats
					SIC172	Textile weaving	SIC365	Mfg. of games & toys
Input-Output	0.0556	0.0172	0.2510	2.9055	SIC 262	Mfg. of ceramic goods other than for construction	SIC232	Mfg. of refined petroleum products
(IO)					SIC263	Mfg. of ceramic tiles & flags	SIC181	Mfg. of leather clothes
					SIC231	Mfg. of coke oven products	SIC183	Dressing/dyeing of fur; mfg. of fur articles
					SIC286	Mfg. of cutlery, tools & general hardware	SIC157	Mfg. of prepared animal feeds
Knowledge spillovers	0.0182	0.0106	0.0888	-0.3387	SIC262	Mfg. of ceramic goods other than for construction	SIC233	Processing of nuclear fuel
(KS)					SIC265	Mfg. of cement, lime and plaster	SIC181	Mfg. of leather clothes
					SIC300	Mfg. of office machinery & computers	SIC183	Dressing/dyeing of fur; mfg. of fur articles
					SIC171	Preparation/spinning of textile fibres	SIC362	Mfg. of jewellery & related articles

Note: The table presents descriptive statistics of estimates of the effect of labour pooling (LP), input-output (IO) and knowledge spillovers (KS) on industrial coagglomeration. Coefficients come from single-industry regressions that exploit the variation in the coagglomeration of the industry in question with other industries (mutually exclusive pairs only). Our dataset comprises of 97 industries. The full set of estimates is presented in Appendix Table 1.

Table 4: Relationship between estimated Marshallian forces strength and sectoral characteristics – Univariate regression results; weighted by inverse of standard error

		Labour pooling		Input sharing		Knowledge spillovers	
		(1)	(2)	(3)	(4)	(5)	(6)
EGI – Agglomeration Index-	continuous	.1505*** (.051)	.	.0006 (.020)	.	.0181** (.009)	.
	top	.	.2262*** (.056)	.	-.0007 (.047)	.	.0458* (.024)
	bottom	.	-.0337*** (.013)	.	-.0021*** (.006)	.	-.0090 (.006)
Entry Share	continuous	.0131* (.008)	.	.0003 (.004)	.	.0037 (.004)	.
	top	.	.0071 (.046)	.	-.0083 (.010)	.	-.0006 (.011)
	bottom	.	-.0329** (.013)	.	.0031 (.007)	.	-.0015 (.009)
Share Highly Educated	continuous	.0053 (.007)	.	.0059 (.004)	.	.0077* (.004)	.
	top	.	.0080 (.010)	.	.0213*** (.008)	.	.0213* (.012)
	bottom	.	.0501* (.025)	.	-.0084 (.013)	.	-.0041 (.007)
Incumbent Employment Size	continuous	-.0002 (.027)	.	.0032 (.013)	.	.0023 (.009)	.
	top	.	.0024 (.054)	.	-.0203** (.009)	.	-.0009 (.010)
	bottom	.	-.0258* (.013)	.	-.0136 (.011)	.	-.0171*** (.006)

Note: robust standard errors in parenthesis. Number of observations 97 except in the panel focusing on Incumbent Employment Size where SIC233 (Processing of nuclear fuel, an outlier with 399 employees) is excluded. Results using the continuous version of the variables listed in the first column are reported in Columns (1), (3) and (5). Results using dummies identifying industries in the top 10% and bottom 10% of these variables are reported in Columns (2), (4) and (6).

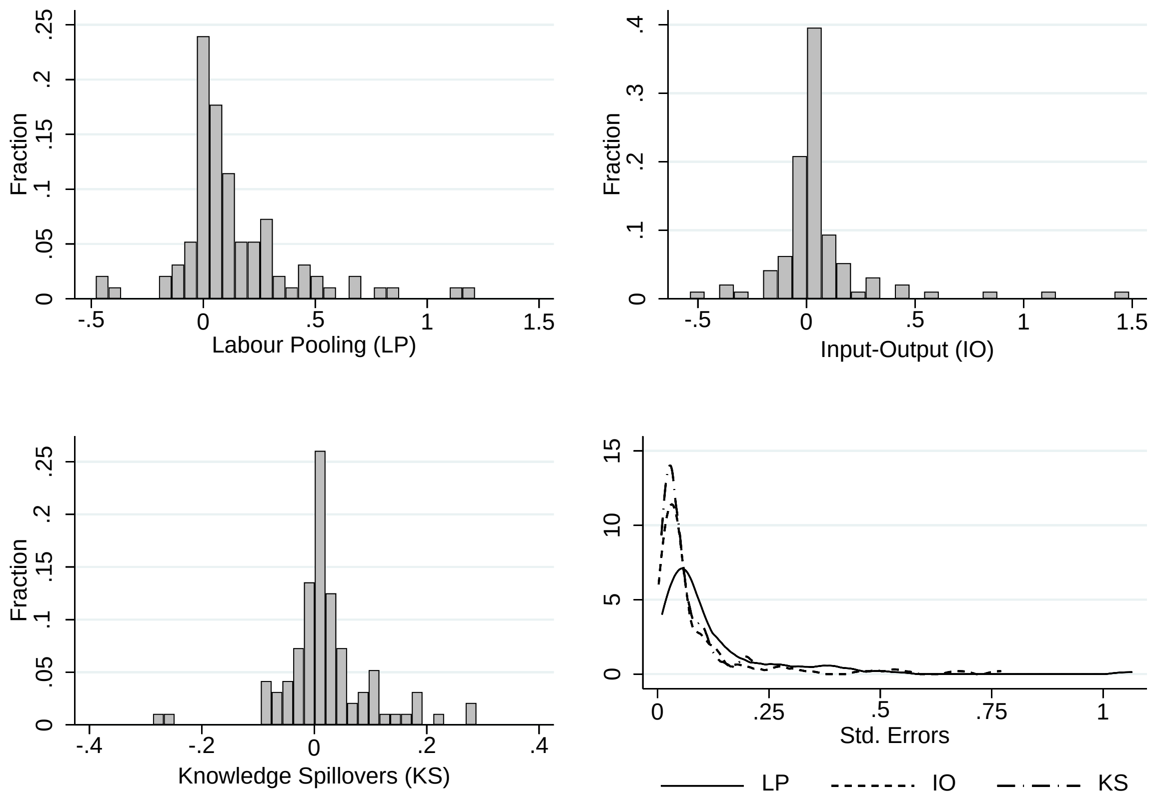
Table 5: Relationship between estimated Marshallian forces strength and sectoral characteristics – Multivariate regression results; weighted by inverse of standard error

		Labour pooling		Input sharing		Knowledge spillovers	
		(1)	(2)	(3)	(4)	(5)	(6)
EGI – Agglomeration Index	continuous	.1464*** (.055)	.	.0071 (.020)	.	.0171** (.008)	.
	top	.	.2200*** (.064)	.	.0017 (.045)	.	.0478* (.024)
	bottom	.	-.0272* (.017)	.	-.0208*** (.007)	.	-.0063 (.006)
Entry Share	continuous	.0051 (.009)	.	-.0007 (.003)	.	.0024 (.003)	.
	top	.	-.0006 (.050)	.	-.0080 (.010)	.	-.0002 (.014)
	bottom	.	-.0103 (.018)	.	.0140* (.008)	.	.0079 (.005)
Share Highly Educated	continuous	-.0101 (.007)	.	.0061 (.004)	.	.0078** (.004)	.
	top	.	-.0083 (.012)	.	.0077 (.008)	.	.0216* (.012)
	bottom	.	.0465* (.027)	.	-.0113 (.017)	.	.0028 (.008)
Incumbent Employment Size	continuous	-.0061 (.023)	.	.0016 (.011)	.	-.0043 (.008)	.
	top	.	-.0115 (.054)	.	-.0138 (.014)	.	.0014 (.012)
	bottom	.	-.0249 (.021)	.	-.0174 (.013)	.	-.0148 (.009)

Note: See Table 4.

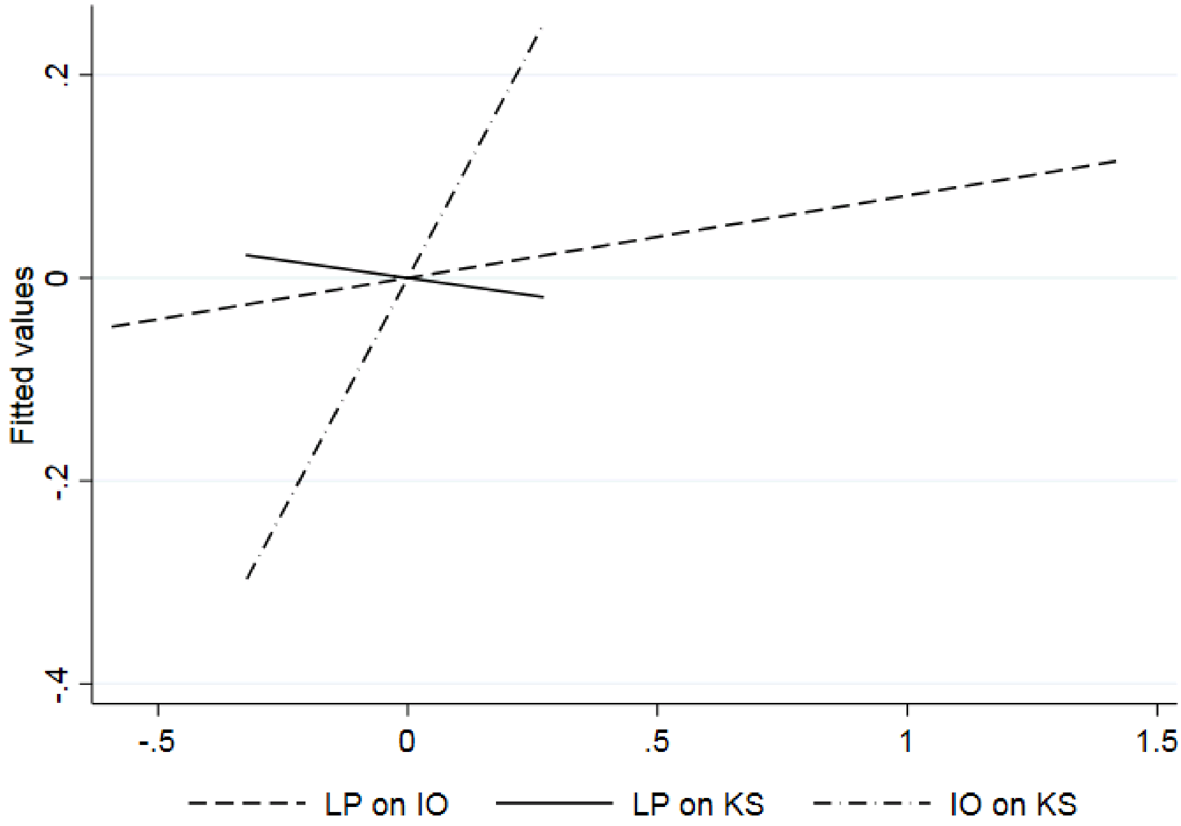
Figures

Figure 1: Distribution of the estimated strength of the Marshallian forces



Note: The top two plots and the bottom left plot present histograms for the distribution of the effect of labour pooling (LP), input-output (IO) and knowledge spillovers (KS) on industrial coagglomeration. These coefficients come from single-industry regressions that exploit the variation in the coagglomeration of the industry in question with other industries (mutually exclusive pairs only) over twelve years. Our dataset comprises of 97 industries. The bottom right plot presents the distribution of standard errors of the three set of Marshallian forces estimates (LP, IO and KS). Standard errors clustered at the industry pairs. The full set of estimates is presented in Appendix Table 1.

Figure 2: Associations (linear fit) between Marshallian forces across industrial sectors



Note: The plots presents linear-fit lines obtained from regressing one Marshallian force on another Marshallian force as detailed in the legend. When pairing up Marshallian force, the one with the smallest amount of variation was used as right-hand side variable to make sure the predictions plotted in the graph cover the actual variation in the variable on the right-hand side (and do not 'predict' out-of-sample). The original Marshallian forces were normalized to have zero mean so that all plots cross at the axis origins. Regression coefficients (standard errors) are follows. LP on IO: 0.0811 (0.0696); LP on KS: -0.0694 (0.3810); IO on KS: 0.9210 (0.5021).

Appendix Tables and Figures

Appendix Table A1.1: Single-Industry estimates of the impact of Labour Pooling, Input-Output and Knowledge Spillovers on Industrial Coagglomeration

		Labour Pooling			*Input-Output*			*Knowledge Spillovers*		
Industrial code	Sector Description	Estimate	Std. Err.	Signif.	Estimate	Std. Err.	Signif.	Estimate	Std. Err.	Signif.
151	Production/processing/preserving of meat	0.1144	0.0901		-0.0037	0.0179		-0.0178	0.0288	
152	Processing/preserving of fish	0.1161	0.2072		0.0094	0.1136		0.0077	0.0612	
153	Processing/preserving of fruit & vegetables	-0.0058	0.0506		0.0098	0.0397		0.0115	0.0144	
154	Mfg. of vegetable and animal oils and fats	-0.3919	0.4087		-0.1383	0.5135		0.0915	0.1468	
155	Mfg. of dairy products	-0.0213	0.0209		0.0398	0.0358		0.0073	0.0121	
156	Mfg. of grain mill products, starches and starch	0.0518	0.0726		-0.0355	0.0364		0.0160	0.0270	
157	Mfg. of prepared animal feeds	0.1328	0.0908		-0.2696	0.1181	**	0.0353	0.0287	
158	Mfg. of other food products	-0.0720	0.0642		0.0222	0.0332		0.0391	0.0305	
159	Mfg. of beverages	-0.0335	0.0542		0.0727	0.0426	*	-0.0164	0.0133	
171	Preparation/spinning of textile fibres	1.1082	0.3623	**	0.0950	0.0355	**	0.1873	0.0979	*
172	Textile weaving	0.7900	0.2244	**	-0.0889	0.1052		0.1676	0.1068	
173	Finishing of textiles	0.8650	0.1601	**	-0.0565	0.0421		0.0215	0.0640	
174	Mfg. of textile articles, except apparel	0.0289	0.0238		0.1698	0.0316	**	0.0049	0.0191	
175	Mfg. of other textiles	0.4365	0.1477	**	-0.0302	0.1193		0.0637	0.0869	
176	Mfg. of knitted & crocheted fabrics	1.2135	0.5229	**	0.0611	0.0781		0.0451	0.1029	
177	Mfg. of knitted & crocheted articles	0.4989	0.3553		-0.0143	0.1072		0.1101	0.0740	
181	Mfg. of leather clothes	0.5175	0.4088		-0.3740	0.0830	**	-0.2870	0.2129	
182	Mfg. of wearing apparel and accessories	0.4702	0.2682	*	-0.0665	0.1132		0.0439	0.1103	
183	Dressing/dyeing of fur; mfg. of fur articles	0.4631	0.4687		-0.3656	0.1123	**	-0.2629	0.2050	
191	Tanning/dressing of leather	-0.0078	0.3913		0.0628	0.0364	*	0.0127	0.0583	
192	Mfg. of luggage, handbags & similar	0.1993	0.1238		-0.0419	0.0262		-0.0615	0.0380	
193	Mfg. of footwear	0.1070	0.1403		0.1030	0.2553		0.0385	0.0392	
201	Sawmilling/planning/impregnation of wood	0.1074	0.0927		-0.0146	0.0513		0.0529	0.0414	
202	Mfg. of veneer sheets, plywood, laminboard & other panels/boards	0.0131	0.0504		0.0298	0.0275		-0.0103	0.0117	

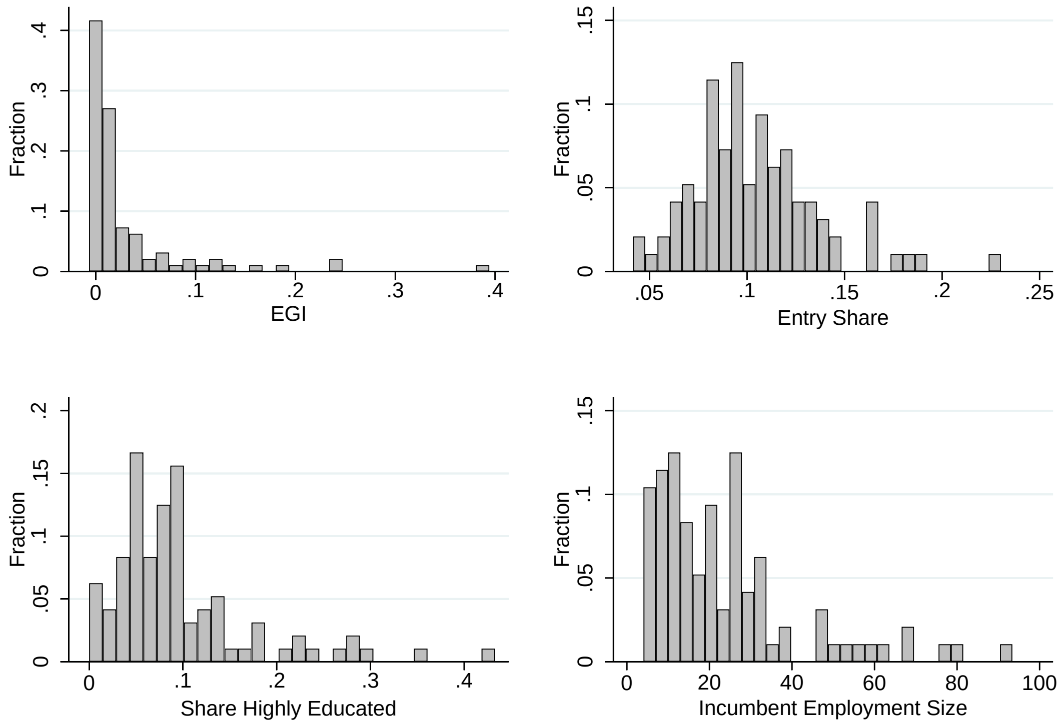
203	Mfg. of builders' carpentry & joinery	0.0427	0.0518		0.0063	0.0224		-0.0077	0.0132	
204	Mfg. of wooden containers	0.1214	0.1030		-0.0350	0.0404		0.0194	0.0097	**
205	Mfg. of other products of wood, cork & straw	0.0058	0.0238		0.0128	0.0116		-0.0129	0.0094	
211	Mfg. of pulp, paper & paperboard	0.1179	0.0821		-0.0049	0.0263		-0.0218	0.0350	
212	Mfg. of articles of paper & paperboard	-0.0099	0.0602		0.0222	0.0175		0.0101	0.0237	
221	Publishing	0.3071	1.0654		0.3239	0.2916		0.0941	0.2019	
222	Printing and service activities related to printing	-0.1370	0.3212		0.4208	0.3346		0.0415	0.1362	
231	Mfg. of coke oven products	0.0698	0.1178		0.8325	0.7713		-0.0071	0.1150	
232	Mfg. of refined petroleum products	-0.1138	0.1474		-0.5365	1.1437		-0.0458	0.0525	
233	Processing of nuclear fuel	0.2348	0.1448		-0.2691	0.6758		-0.3038	0.1909	
241	Mfg. of basic chemicals	0.0165	0.0532		0.0439	0.0238	*	0.1018	0.0517	*
242	Mfg. of pesticides & agro-chemical products	-0.0068	0.1019		0.0045	0.0320		0.1182	0.0562	**
243	Mfg. of paints, varnishes & similar	0.0376	0.0467		0.0338	0.0335		-0.0744	0.0927	
244	Mfg. of pharma., medicinal chemicals & botanical products	0.0128	0.0614		-0.1268	0.1177		0.0228	0.0402	
245	Mfg. of soap, detergents, perfumes & cleaning/polishing preparations	0.0077	0.0371		0.0869	0.0410	**	0.0369	0.0270	
246	Mfg. of other chemical products	0.0668	0.0488		0.0469	0.0202	**	-0.0683	0.0716	
247	Mfg. of man-made fibres	0.5491	0.2419	**	0.1170	0.0539	**	0.1104	0.0928	
251	Mfg. of rubber products	0.1856	0.1853		0.0440	0.0311		0.0209	0.0415	
252	Mfg. of plastic products	0.0955	0.0661		-0.0074	0.0183		0.0091	0.0154	
261	Mfg. of glass & glass products	-0.0181	0.0353		0.0650	0.0238	**	-0.0140	0.0265	
262	Mfg. of ceramic goods other than for construction	0.2598	0.1787		1.4854	0.4645	**	0.2885	0.1142	**
263	Mfg. of ceramic tiles & flags	0.3761	0.1799	**	1.1089	0.5490	**	0.1058	0.0763	
264	Mfg. of bricks, tiles and construction products, in baked clay	0.2820	0.1506	*	-0.1363	0.2789		0.0777	0.0553	
265	Mfg. of cement, lime and plaster	-0.1422	0.0556	**	-0.1543	0.0749	**	0.2793	0.0914	**
266	Mfg. of articles of concrete, plaster, lime & cement	0.0807	0.0921		0.0111	0.0637		0.0068	0.0352	
267	Cutting, shaping & finishing of stone	-0.4656	0.3024		0.1115	0.0521	**	-0.0342	0.0252	
268	Mfg. of other mineral products	0.0494	0.0800		-0.0926	0.0653		0.0538	0.0347	
271	Mfg. of basic iron, steel & of ferro-alloys	0.2807	0.1173	**	0.3323	0.1759	*	0.1774	0.1807	
272	Mfg. of tubes	0.2514	0.1019	**	0.2342	0.1524		-0.0802	0.0370	**

273	Other processing of iron, steel & production of ferro-alloys	0.3292	0.1097	**	0.4186	0.1995	**	-0.0317	0.0817	
274	Mfg. of basic precious & non-ferrous metals	0.3036	0.0558	**	0.0428	0.0342		-0.0385	0.0192	**
275	Casting of metals	0.6708	0.1286	**	-0.0110	0.0668		-0.0232	0.0484	
281	Mfg. of structural metal products	0.0652	0.0299	**	0.1393	0.0405	**	0.0107	0.0182	
282	Mfg. of tanks, reservoirs, radiators & boilers	0.2048	0.0774	**	0.1902	0.0945	**	-0.0530	0.0398	
283	Mfg. of steam generators, except boilers	-0.0003	0.0385		0.0168	0.0607		0.0000	0.0344	
284	Forging, pressing, stamping & rolling metal	0.3519	0.1195	**	0.3274	0.1566	**	-0.0769	0.1015	
285	Treatment/coating of metals	0.0587	0.0312	*	0.1398	0.0383	**	-0.0239	0.0114	**
286	Mfg. of cutlery, tools & general hardware	0.2377	0.0654	**	0.5987	0.2117	**	-0.0392	0.0494	
287	Mfg. of other fabricated metal products	0.1758	0.0389	**	0.0917	0.0430	**	0.0524	0.0297	*
291	Mfg. of other machinery for production/use of mechanical power N.E.C.	0.1179	0.0633	*	0.0697	0.0334	**	-0.0071	0.0413	
292	Mfg. of other general purpose machinery	0.0338	0.0152	**	-0.0012	0.0026		0.0170	0.0144	
293	Mfg. of agricultural/forestry machinery	0.1168	0.0667	*	0.0452	0.0421		-0.0175	0.0587	
295	Mfg. of other special purpose machinery	0.0240	0.0432		0.0914	0.0527	*	0.0221	0.0149	
297	Mfg. of domestic appliances	0.0473	0.0560		0.1435	0.0670	**	0.0082	0.0440	
300	Mfg. of office machinery & computers	-0.0577	0.0811		0.0174	0.0180		0.2150	0.0497	**
311	Mfg. of electric motors, generators & transformers	-0.0189	0.0116		0.0201	0.0187		0.0197	0.0120	
312	Mfg. of electricity distribution & control apparatus	0.0148	0.0103		0.0133	0.0116		0.0017	0.0135	
313	Mfg. of insulated wire & cable	0.0179	0.0442		0.0174	0.0210		0.0358	0.0315	
314	Mfg. of accumulators, cells & batteries	0.0071	0.0619		-0.0389	0.0315		0.1409	0.0951	
315	Mfg. of lighting equipment & lamps	-0.0782	0.0442	*	0.0226	0.0321		0.0031	0.0276	
316	Mfg. of electrical equipment N.E.C.	0.0075	0.0274		0.0298	0.0175	*	0.0043	0.0128	
321	Mfg. of electronic valves, tubes & electronic components	0.0543	0.0369		0.0042	0.0206		-0.0113	0.0284	
322	Mfg. of TV/radio transmitters & telephones/telegraphs	0.0313	0.0338		0.0004	0.0193		0.0392	0.0113	**
323	Mfg. of TV/radio receivers & sound/video recording/reproducing	0.0314	0.0180	*	0.0295	0.0499		0.0385	0.0145	**
331	Mfg. of medical, surgical & orthopaedic equipment	0.0189	0.0201		0.0345	0.0141	**	0.0094	0.0087	
332	Mfg. of instruments for measuring, checking, testing & navigating	0.0320	0.0125	**	0.0365	0.0096	**	0.0197	0.0183	
333	Mfg. of industrial process control equipment	0.0884	0.0439	**	0.0075	0.0270		-0.0277	0.0218	

334	Mfg. of optical & photographic equipment	0.0258	0.0444		0.0336	0.0248		0.0071	0.0445	
335	Mfg. of watches & clocks	-0.0549	0.2471		-0.1312	0.0759	*	0.0168	0.0295	
341	Mfg. of motor vehicles	0.2914	0.0632	**	0.0172	0.0452		0.1005	0.0514	*
342	Mfg. of bodies for vehicles, trailers & semi-trailers	0.1684	0.1061		-0.0034	0.0407		0.0193	0.0234	
343	Mfg. of parts & accessories for vehicles/engines	0.2739	0.0639	**	0.1177	0.0480	**	0.0126	0.0304	
351	Building & repairing of ships/boats	0.2039	0.0773	**	-0.0502	0.1227		0.0084	0.0471	
352	Mfg. of railway/tramway locomotives/rolling stock	-0.1168	0.0512	**	-0.0139	0.0117		-0.0039	0.0143	
353	Mfg. of aircraft/spacecraft	0.1872	0.0666	**	-0.1883	0.0793	**	-0.0347	0.0644	
354	Mfg. of motorcycles/bicycles	0.1558	0.0465	**	0.0088	0.0240		0.0268	0.0161	*
355	Mfg. of other transport equipment N.E.C.	0.0275	0.0661		0.0008	0.0210		-0.0045	0.0188	
361	Mfg. of furniture	0.0002	0.0310		0.0186	0.0136		0.0085	0.0155	
362	Mfg. of jewellery & related articles	0.6844	1.1279		0.0164	0.0366		-0.0918	0.1461	
363	Mfg. of musical instruments	-0.4802	0.2851	*	0.0016	0.0119		0.0700	0.0542	
364	Mfg. of sports goods	0.0334	0.0272		0.0058	0.0176		-0.0802	0.0510	
365	Mfg. of games & toys	-0.1465	0.0933		-0.0686	0.0346	*	0.1817	0.0642	**
366	Miscellaneous manufacturing N.E.C	-0.0107	0.0270		-0.0272	0.0122	**	0.0357	0.0282	

Note: Regression coefficients come from single-industry regressions that exploit the variation in the coagglomeration of the 97 industry in question with other 96 industries (mutually exclusive pairs only). Standard errors clustered at the industry pairs. Sector codes and description taken from the Office for National Statistics SIC 1992 classification list.

Appendix Figure 1: Distribution of industry characteristics



Note: Number of observations 97 except for average continuous employment which excludes SIC233 (Processing of nuclear fuel, an outlier with 399 employees). EGI is the Ellison and Glaeser Index of agglomeration. Descriptive statistics for the four indicators are as follows. EGI: mean (0.0321), SD (0.0603) and median (0.0084). Entry Share: mean (0.1047), SD (0.0327) and median (0.099); Share Highly Educated: mean (0.0986), SD (0.0801) and median (0.0783); Incumbent Employment Size: mean (23.733), SD (18.291) and median (19.221).

Web Appendix – Additional regressions

Table A2.1: Relationship between estimated Marshallian forces strength and sectoral characteristics – Univariate regression results; un-weighted regressions

		Labour pooling		Input sharing		Knowledge spillovers	
		(1)	(2)	(3)	(4)	(5)	(6)
EGI – weighted regressions	continuous	.1065***	.	.0830	.	.0094	.
		(.038)		(.068)		(.020)	
	top	.	.3126***	.	.1130	.	.0076
			(.096)		(.161)		(.059)
	bottom	.	-.1545**	.	-.0413*	.	-.0116
			(.059)		(.025)		(.012)
Entry Share	continuous	-.0118	.	-.0091	.	.0227**	.
		(.021)		(.026)		(.009)	
	top	.	-.1377**	.	-.0436	.	.0590**
			(.069)		(.049)		(.027)
	bottom	.	-.0291	.	.0325	.	-.0383
			(.057)		(.102)		(.029)
Share Highly Educated	continuous	-.0606***	.	-.0352	.	.0051	.
		(.022)		(.022)		(.013)	
	top	.	-.0658	.	-.1342*	.	-.0198
			(.049)		(.075)		(.043)
	bottom	.	.1895	.	-.1478**	.	-.0576
			(.118)		(.058)		(.041)
Incumbent Employment Size	continuous	.0009	.	-.0058	.	.0447**	.
		(.070)		(.060)		(.019)	
	top	.	-.0825	.	.0154	.	.0285
			(.088)		(.106)		(.024)
	bottom	.	-.0804	.	-.1358**	.	-.0687
			(.123)		(.059)		(.044)

Note: robust standard errors in parenthesis. Number of observations 97 except in the panel focusing on Incumbent Employment Size where SIC233 (Processing of nuclear fuel, an outlier with 399 employees) is excluded. Results using the continuous version of the variables listed in the first column are reported in Columns (1), (3) and (5). Results using dummies identifying industries in the top 10% and bottom 10% of these variables are reported in Columns (2), (4) and (6).

Table A2.2: Relationship between estimated Marshallian forces strength and age of sector – Univariate regression results; weighted by inverse of standard error

		Labour pooling		Input sharing		Knowledge spillovers	
		(1)	(2)	(3)	(4)	(5)	(6)
Age of sector	continuous	0.0026		-0.0036		0.0029	
		(0.0066)		(0.0042)		(0.0039)	
	top		0.0070		0.0261*		0.0017
			(0.0169)		(0.0140)		(0.0135)
	bottom		0.0732***		0.0033		-0.0110
			(0.0274)		(0.0085)		(0.0087)

Note: robust standard errors in parenthesis. Number of observations 97. Results using the continuous version of the variables listed in the first column are reported in Columns (1), (3) and (5). Results using dummies identifying industries in the top 10% and bottom 10% of these variables are reported in Columns (2), (4) and (6). Age of the sector identified as the difference between the final year in the data set (2007) and the year of opening of the oldest plant in the sector (e.g., for computer SIC300, this would be 2007 – 1961 = 46).

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1618	Andrew Mountford Jonathan Wadsworth	Trainspotting: ‘Good Jobs’, Training and Skilled Immigration
1617	Marta De Philippis Federico Rossi	Parents, Schools and Human Capital Differences across Countries
1616	Michael Amior	Education and Geographical Mobility: The Role of Wage Rents
1615	Ria Ivandic Tom Kirchmaier Stephen Machin	Jihadi Attacks, Media and Local Hate Crime
1614	Johannes Boehm Swati Dhingra John Morrow	The Comparative Advantage of Firms
1613	Tito Boeri Pietro Garibaldi	A Tale of Comprehensive Labor Market Reforms: Evidence from the Italian Jobs Act
1612	Esteban Aucejo Teresa Romano Eric S. Taylor	Does Evaluation Distort Teacher Effort and Decisions? Quasi-experimental Evidence from a Policy of Retesting Students
1611	Jane K. Dokko Benjamin J. Keys Lindsay E. Relihan	Affordability, Financial Innovation and the Start of the Housing Boom
1610	Antonella Nocco Gianmarco I.P. Ottaviano Matteo Salto	Geography, Competition and Optimal Multilateral Trade Policy
1609	Andrew E. Clark Conchita D’Ambrosio Marta Barazzetta	Childhood Circumstances and Young Adult Outcomes: The Role of Mothers’ Financial Problems

1608	Boris Hirsch Elke J. Jahn Alan Manning Michael Oberfichtner	The Urban Wage Premium in Imperfect Labour Markets
1607	Max Nathan Anna Rosso	Innovative Events
1606	Christopher T. Stanton Catherine Thomas	Missing Trade in Tasks: Employer Outsourcing in the Gig Economy
1605	Jan-Emmanuel De Neve Christian Krekel George Ward	Employee Wellbeing, Productivity and Firm Performance
1604	Stephen Gibbons Cong Peng Cheng Keat Tang	Valuing the Environmental Benefits of Canals Using House Prices
1603	Mary Amiti Stephen J. Redding David Weinstein	The Impact of the 2018 Trade War on U.S. Prices and Welfare
1602	Greer Gosnell Ralf Martin Mirabelle Muûls Quentin Coutellier Goran Strbac Mingyang Sun Simon Tindermans	Making Smart Meters Smarter the Smart Way
1601	Antoine Dechezleprêtre Caterina Gennaioli Ralf Martin Mirabelle Muûls Thomas Stoerk	Searching for Carbon Leaks in Multinational Companies
1600	Jeremiah Dittmar Skipper Seabold	New Media and Competition: Printing and Europe's Transformation after Gutenberg

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