Journal of Economic Geography (2013) **0**, 1–27 doi: 10.1093/jeg/lbaa007

Tales of the city: what do agglomeration cases tell us about agglomeration in general?

Giulia Faggio, Olmo Silva**[,2] and William C. Strange****

*Department of Economics, School of Arts and Social Sciences, City, University of London, Northampton Square, London EC1V 0HB, UK
**Department of Geography and Environment, London School of Economics, Houghton Street, London WC2A 2AE, UK
***Rotman School of Management, University of Toronto, 105 St. George St., Toronto, ON M5S 3E6, Canada
[2]Correspondence to: *email <O.Silva@lse.ac.uk>*

Abstract

This article considers the heterogeneous microfoundations of agglomeration economies. It studies the co-location of industries to look for evidence of labour pooling, input sharing and knowledge spillovers. The novel contribution of the article 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 microfoundation models for 97 manufacturing sectors, including the classic agglomeration cases of automobiles, computers, cutlery and textiles. These four casesÐas with all of the individual industry models we estimateÐclearly show the importance of the Marshallian forces. However, they also highlight how the importance of these forces varies across industriesÐimplying that extrapolation from cases should be viewed with caution. The article concludes with an investigation of the pattern of heterogeneity. The degree of an industry's clustering (localisation), entrepreneurship, incumbent firm size and worker education are shown to contribute to the pattern of heterogeneous microfoundations.

Keywords: Agglomeration, microfoundations, heterogeneity, industrial clusters
JEL classi®cations: R10, R12, L52, L60
Date submitted: 3 June 2019 **Editorial decision:** 17 February 2020 **Date Accepted:** 28 February 2020

1. 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 programme leaves out important specific details. A more qualitative literature on agglomeration considers specific cases. These case studies have considered a range of industriesÐincluding 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 article builds on Ellison *et al.* (2010), who consider the microfoundations of coagglomeration of industry pairs. The key result in Ellison *et al.* (2010) is that Marshall's three forcesÐinput sharing, knowledge spillovers and labour poolingÐare 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 labour 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 characterised by the presence of smaller firms show stronger effects from input linkagesÐa result in the spirit of Chinitz (1961).

The present article 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 spilloversÐallowing for like-for-like cross-industry comparisons on the determinants of agglomeration. Furthermore, this approach allows us to characterise the forces that drive one industry's agglomeration (e.g. the industry is dominated by small firms or high levels of firm entry) rather than the forces that 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 generalise easily and directly to the universe of industriesÐor to other industriesÐand 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-locationÐwith 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 USA). More broadly, our study argues for shifting the attention solely from questions such as `Is labour pooling or input-sharing more important for agglomerations?' to investigations that establish which industries are most responsive to a particular force underlying economic agglomeration.

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.

To carry out our analysis, we employ confidential firm-level data for 97 manufacturing industries from the UK Business Structure Database (BSD) covering the years 1997±2008. We match this information with a range of other data on industry characteristics 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 labour markets. In that sense, they correspond to US 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 labour market pooling are importantÐ while knowledge spillovers are not. The picture is different when considering textilesÐ 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 labour 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 generalise from cutlery to textiles and illustrate more generally the limits of extrapolation.

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

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 (localisation), its entrepreneurship (as proxied by its new firms' creation rate), its incumbent firms' size and its workers' education. In studying these patterns, we are forced to deal with various ambiguous predictions arising from theory. For instance, different theoretical frameworks suggest that the agglomeration of a given industry can be a substitute or a complement for that industry's coagglomeration with other industries.[2] On the one hand, there may be a substitution effect of

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.

localisation if the presence of own industry activity fosters agglomerationÐmaking cross-industry coagglomeration not valuable. On the other hand, industries that benefit from labour pooling, input sharing and knowledge spillovers will seek to enjoy these benefits *within* and *across* sectorsÐthat is, by both agglomerating and coagglomerating. There are similar ambiguities (discussed in Section 5) associated with the likely effect of entrepreneurship, size of incumbent firms and workers' educationÐa standard, but imperfect, proxy for skills.

To (empirically) resolve some of these ambiguities, we regress our estimated coefficients for the Marshallian agglomeration forces on our proxies for own-industry agglomeration, entrepreneurship, incumbent firm size and worker education. Regarding whether agglomeration and coagglomeration are complements or substitutes, we find that for labour pooling and knowledge spillovers, complementarity dominates. However, we find little effects on input sharing. Regarding entrepreneurship, we find that more dynamic industries have larger labour pooling coefficients. Conversely, input sharing coefficients are smaller for the most dynamic industries. We do not find a significant relationship with the knowledge spillover coefficients. Next, we find no impact of incumbent size on labour pooling, while input sharing is less important with large existing firmsÐa Chinitz (1961) effect. For knowledge spillovers, we find instead a sort of anchor effectÐwith smaller firms having smaller effects. Finally, for education, the labour pooling effect is strongest for the less educated workers but input sharing is strongest with a more educated workforceÐa pattern suggestive of the nursery effects discussed by Vernon (1960). We find no significant effect on 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 evidence from individual industry models coupled with the results from our 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 with reasonable accuracy.

The rest of the article is structured as follows. In Section 2, we present the data and describe the main variables we use. In Section 3, we present our findings on the four classic cases of agglomerationÐnamely cutlery, textile, cars and computers. Section 4 discusses the heterogeneity we find when we explore all the manufacturing industries in our sample. Finally, Section 5 presents our attempt at rationalising this heterogeneity. Some concluding remarks are presented in Section 6.

2. Data

2.1. Data and variable construction

The core data we use to carry out our analysis is the UK BSD covering the period 1997±2008. The data are 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 and/or with at least one employee registered for tax collection appear on the IDBR. In 2004, the Office for National Statistics (ONS) estimated that businesses listed on the IDBR accounted for approximately 99% 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 organisation, while the second can be thought of as a plant or establishment. In the majority of cases (70%), 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 include approximately three million local units every year. However, 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. First, `Manufacturing of tobacco products' (SIC160) is dropped because of its limited number of plants throughout the sample period (e.g. 43 in 1997). Secondly, we disregard five industries for which we cannot measure one of our key variables of interestÐnamely, 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 three-digit sectors for a total of 4656 unique pairwise industry combinations for 12 years (1997±2008). The complete dataset thus contains 55,872 industry-pair-by-year observations.[5]

In terms of geographical units of aggregation, we use TTWAsÐwhich 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 labour markets and economically relevant aggregates. In 2007, there were 243 TTWAs within the UK. We focus on Britain (excluding Northern Ireland), split TTWAs into urban and rural ones, and only consider 84 urban TTWAs with population of over 100,000 residents. More detail is provided in the web appendix to our previous work (Faggio *et al.*, 2017).[6]

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:

$$\gamma^{C}{}_{ij} = \frac{\sum_{m=1}^{M} (s_{mi} - x_m)\ (s_{mj} - x_m)}{1 - \sum_{m=1}^{M} (x_m)^2}. \tag{2.1}$$

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

4 For the same reason, our analysis does not consider service industries. As detailed below, we use patent citation counts to proxy for knowledge spillovers. Such information is largely unavailable for non-manufacturing sectors.

5 Note that we checked that our results do not change substantially if we use the 94 sectors considered in our previous work (obtained by re-aggregating sectors with low employment/firm counts; see Faggio *et al.*, 2017).

6 In some extensions, we experimented with keeping rural areas or excluding London from our analysis. Overall, we find similar results.

This measure is related to the covariance of industries across metropolitan areas. To study how this tendency of industries 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 labour pooling, we use the UK Labour Force Survey (LFS) data between 1995 and 1999Ðat the beginning of our observation windowÐand 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 TTWAsÐwhile we drop Northern Ireland (as we did for our main BSD data). Finally, we select individuals living in urban areas and working in manufacturingÐleaving 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 categorises 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 (henceforth, I±O Tables) for 1995±1999. This allows us to measure the extent to which industries buy and sell intermediate inputs from one another. Specifically, we 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.[7]

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±2009.[8] 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. To construct knowledge spillover measures, we exclude self-citations from the same inventor (or the

7 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).

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

company at which he/she is based), as well as citing patents filed after 2000 or before 1981, and cited patents filed after 1997. The latter restrictions guarantee that on average cited patents are at least 3 years older than citing ones and allow us to centre our knowledge spillover measures in the initial years of our sample (i.e. up to 2000)Ðso that they are measured at a similar time as the labour 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 patentsÐwhich are categorised using technological classes. Following the literature, we use a probabilistic mapping based on the Industry of Manufacture (IOM).[9] 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 considers the maximum patent-citation flow between sector i and sector jÐnormalised by total citations in that industry.

Using the various data sources discussed above, we construct an additional set of sector/sector-pair characteristics that we deploy in our analysis. 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 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 I±O Tables) considering their purchases 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 and business services. 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 characterise 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 (localisation) 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 `entrepreneurship'Ðor its dynamismÐby measuring the entry share of new firms in the sector (i.e. the incidence of new firms 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). Fourth, we characterise sectors by measuring the average size of their incumbentsÐthat is, the employment of firms operating both at time t and $t - 1$ (based on the BSD data).

2.2. 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 centred on zero with a standard deviation of 0.008, a minimum of -0.043 and a maximum of 0.175. These figures are

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

Table 1. Descriptive statisticsÐestimation sample for coagglomeration models

	Mean	SD	Min.	Max.
Coagglomeration measures and Marshallian forces				
TTWA total employment coagglomeration (γ^{C})	0.000	0.008	−0.043	0.175
labour 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

Notes: Number of observations: 55,872. The sample contains non-repeated pairwise combination of 97 manufacturing SIC1992 three-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 and ammunition (SIC296), recycling of metal waste and scrap (SIC371) and recycling of non-metal waste and scrap (SIC372) because of missing data on knowledge ¯ows as measured by patent citations.
Sources: The coagglomeration index is computed using the ONS UK BSD 1997±2008. Labour correlation indices are computed from the UK LFS 1995±1999. Input±output measures are calculated using ONS UK I±O 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±1997. Citing patents sampled for the years 1981±2000. Additional control measures are calculated using the UK I±O Tables for 1995±1999.

similar to the patterns in Faggio *et al.* (2017)Ðwhere we used 94 manufacturing industries (instead of 97)Ðand consistent with those of Ellison *et al.* (2010).

The next three rows present descriptive evidence for our proxies for the Marshallian forces. The mean value for labour 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 zeroÐat 0.013 and 0.016, respectivelyÐbut the distributions have a strong right skewÐwith maximum values of 0.782 and 0.420, respectively. This suggests that most industries do not share intermediates or knowledgeÐbut 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 the use of water (0.001). The other three measures have similar mean values at around 0.014±0.020.

The attributes we use to characterise industries are presented graphically in Appendix Figure A1. 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 workersÐwith 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.

3. Four classic agglomeration tales

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:

$$\gamma_{ijt}^{\mathrm{C}} = \alpha + \beta_{\mathrm{LP}}\mathrm{LP}_{ij} + \beta_{\mathrm{IO}}\mathrm{IO}_{ij} + \beta_{\mathrm{KS}}\mathrm{KS}_{ij} + \Sigma_{k=1}{}^{5}\lambda_{\mathrm{k}}\mathrm{Diss}_{ij}^{k} + \varepsilon_{ijt}, \tag{3.1}$$

where $\gamma_{ijt}^{\mathrm{C}}$ 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 labour 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 standardise variables to have unitary standard deviation at the level of the full datasetÐthat is, when considering all manufacturing sectors. This eases comparison of the relative strength of the three Marshallian forces.[10]

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 (3.1) including all manufacturing sectors in our data. In this case, the sample includes 4656 industry pairs repeated over 12 years, giving rise to 55,872 observations. Standard errors are clustered at the industry-pair level.[11] 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 labour pooling has a much stronger effect than input sharing and knowledge spillovers. In particular, the standardised effect of LP is approximately 10%Ðtwo and a half times the impact of IO (at 3.7%) and five times larger than the impact of KS (at 2%).[12]

10 Notice again that we follow most of the literature and focus on manufacturing because we cannot appropriately measure knowledge spillovers using patent-citation flows for the service industries. For an exception, see Kolko (2010) who follows the approach by Ellison *et al.* (2010) to study coagglomeration of service industries. The article only estimates models that include input/output and labour linkages.

11 Note that while there is time-variation in $\gamma_{ijt}^{\mathrm{C}}$, 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 pairÐthat is, to 4656 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 2 years (corresponding to the `Great Recession'). We found broadly comparable results irrespective of the exact years considered.

12 One possible explanation for the weaker KS results is that knowledge spillovers are more difficult to define and measure than other Marshallian forces (see footnote 8).

It is worth emphasising 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 industriesÐwhich is instead our focus here. In our previous paper, we thoroughly assessed the robustness of these findingsÐfor 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. We further probed their causal nature by using an instrumental variable strategy that predicts the extent of labour pooling, input sharing and knowledge spillovers among UK manufacturing industries using proxies for the three Marshallian forces constructed using US data. In the current work, we have carried out a similar set of checks and reached similar conclusions: our findings most likely capture the causal impact of the three Marshallian forces on co-locationÐholding fixed other potential confounders. These results are not reported here for brevity.

We now turn to industry-specific models of our four salient classic tales of agglomerationÐnamely, the textile (SIC171-SIC177), cutlery (SIC286), computer (SIC300) and automobile (SIC341) sectors. When considering microfoundations for these specific sectors, the empirical model in Equation (3.1) is identified by exploiting variation in how one of these industries co-locates with the remaining 96 manufacturing industries. 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 labour pooling effects at 0.367Ðthree 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 itself) 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 than the average manufacturing sector (the EGI index is 0.081 vs. 0.032), and have a less educated workforce than average (the share of a college graduate is 0.050 compared to a manufacturing-wide average of 0.099). The sector also has close-to-average levels of new firms' entry and size of incumbent firms.

The picture is different when considering 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, 172). The results in Table 2 support these intuitions. The coefficient on input linkages is very large and significant (at 0.599Ðsixteen 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 labour 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 generalise from cutlery to textiles: testing whether the strength of IO (and KS) between cutlery and textiles is the same clearly leads us to

Table 2. Coagglomeration and Marshallian forcesÐwhole 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	0.1014*** (0.0144)	0.0366** (0.0149)	0.0199** (0.0092)	0.0321	0.1047	0.0986	27.61
Preparation, weaving and ®nishing of textiles	171-177	0.3672*** (0.1126)	0.0119 (0.0566)	0.1426*** (0.0349)	0.0810	0.1084	0.0505	20.83
Mfg. of cutlery, tools and general hardware	286	0.2377*** (0.0681)	0.5987*** (0.2204)	−0.0392 (0.0514)	0.0379	0.0774	0.0375	11.22
Mfg. of of®ce machinery and computers	300	−0.0577 (0.0844)	0.0174 (0.0187)	0.2150*** (0.0518)	0.0066	0.1658	0.2938	10.69
Mfg. of cars engines and bodies for vehicles	341	0.2914*** (0.0658)	0.0172 (0.0470)	0.1005* (0.0536)	0.0451	0.1664	0.1029	34.58

Notes: Regression coef®cients 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 12 years. Number of observations as follows. All sectors: 55,872. Textiles (SIC171-177): 7812. Cutlery (SIC286), Computers (SIC300) and Cars (SIC241): 1152. Standard errors clustered at the industry pair level reported in parenthesis. The Ellison± Glaeser Agglomeration Index (EGI) reported is an average across industries and years.

reject the nullÐthough from a statistical point of view LP has the same importance in both sectors. In terms of its attributes, cutlery has average agglomeration, relatively low entrepreneurship and low educated workersÐclarifying that, even in terms of underlying organisational 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 firmsÐat 11 employees (40% of the manufacturing average)Ðcoupled 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 discussions 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 0.215Ð10 times larger than for the average manufacturing sector (and statistically different from KS in the rest of manufacturing; p-value on the significance of the difference: 0.000). The input sharing coefficient is also positive but substantially smaller and non-significant (at 0.017Ðhalf 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 labour market pooling (at −0.058). It is worth noting that the latter result does not mean that there is no labour market pooling in this industry. Instead, there could be significant labour pooling taking place within the computer sector itself. Note also that, although the computer industry is similar to cutlery in having small incumbents, the Chinitz-type IO effects are dominated by the importance of knowledge effects. This is possibly due to its highly educated workforce and entrepreneurshipÐfeatures that distinguish this sector from the previous tale. Clearly, this suggests thatÐhowever, appealing it might be to use the computer industry to illustrate agglomeration economies in generalÐthe logic of extrapolating from the computer industry is strained.

The car industry is also highly salient in the agglomeration literature. In the USA, this industry's declining cluster centred 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 averageÐlike the computer sector (though less markedly so)Ðand high entrepreneurship. It is also not markedly more agglomerated than the average manufacturing sector (unlike textiles; computers are instead clearly less agglomerated). The only remarkable difference for computers is that incumbents in the car sector are very large. Despite these broad similarities, the pattern of the regression coefficients differs. Labour 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 organisation 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 nullÐthough the rejection is borderline for the latter (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 97 manufacturing sectors.

4. 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 97 manufacturing sectors covered by our data. Stated differently, we estimate the empirical model in Equation (3.1) industry-by-industryÐthat is considering all 97 sectors and not just the four classic tales. As discussed, these models are identified by the variation in the coagglomeration patterns of one industry with the remaining 96. The most important findings from this exercise are presented in Figure 1 and Table 3. The full set of estimates is presented in Online Appendix Table W1.[13]

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 industries, we find a mean estimated effect of 0.138,[14] but a substantially smaller median impact at 0.059 (see top panel of Table 3). This difference is due to a spreadout distribution of estimates (standard deviation of

13 An editable .xls version of the table can be accessed from the corresponding author's webpage at the following link: http:// personal. lse. ac. uk/ silvao/ DataFromBigTable_ July2018. xlsx.

14 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 industriesÐthat is, for the average sector.

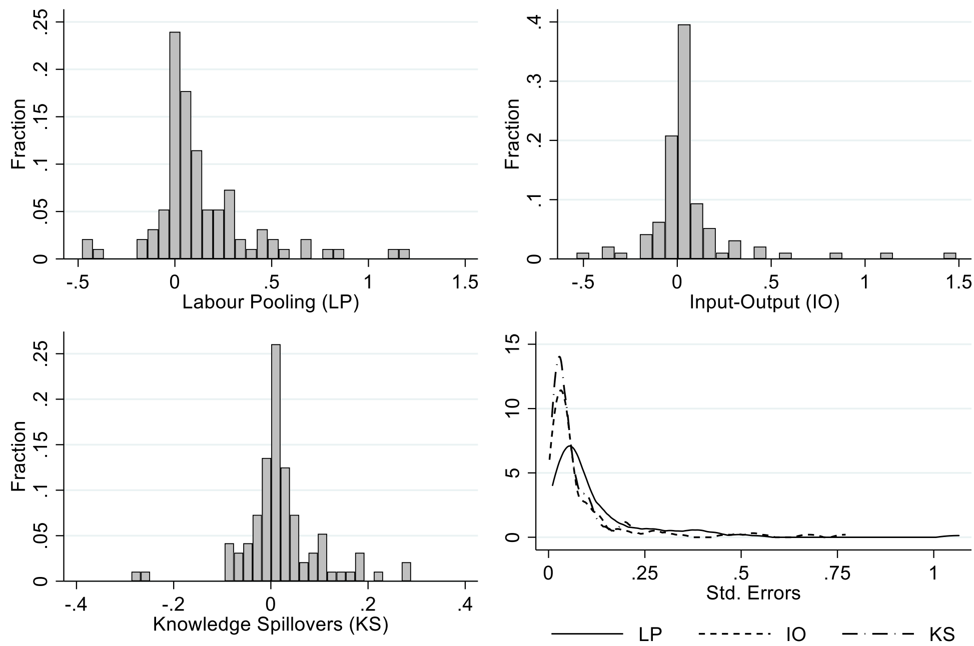

Figure 1. Distribution of the estimated strength of the Marshallian forces.
Notes: The top two plots and the bottom left plot present histograms for the distribution of the effect of labour pooling (LP), input±output (IO) and knowledge spillovers (KS) on industrial coagglomeration. These coefficients 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 12 years. Our dataset comprises of 97 industries. The bottom right plot presents the distribution of standard errors of the three sets of Marshallian forces estimates (LP, IO and KS). Standard errors clustered at the industry pairs. The full set of estimates is presented in Online Appendix Table 1.

0.268) with an evident stretch towards positive values (skew 1.44). The top-left plot of Figure 1 further reveals that the estimated LP effect distribution easily covers the interval (−0.5,0.5) but stretches well above this range on the positive side of the horizontal axisÐreaching values above one (i.e. a unitary standardised effect). However, not all of these estimates are significant. The bottom-right plot of the figure also reveals that the associated standard errorsÐthough mainly concentrated in the interval (0, 0.25)Ðare relatively spread out giving rise to approximately 36% significant estimates. Regarding IO, we find a much smaller average effect at 0.056 and an even smaller median impact at 0.017. This reflects the fact that nearly 40% of the sectors have IO effects very close to zero (Figure 1). On the other hand, the distribution is significantly spread out (standard deviation of 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 linkagesÐbut some industries are very interconnected. We also find relatively stretched out standard errorsÐonce again giving rise to 36% significant estimates. Regarding KS, we find even smaller mean and median values at 0.018 and 0.011, respectively. The KS distribution is less spread out (standard deviation of 0.089) than those of LP and IO, has a small negative skew (−0.339), and is more symmetricÐwith values

Table 3. Summary statistics for the microfoundation forcesÐestimated industry-by-industry

	Mean	Median	SD	Skewness	Top 4	Top 4 industry description	Bottom 4	Bottom 4 industry description
Estimated labour pooling effect (LP)	0.1380	0.0587	0.2677	1.444	SIC176	Mfg. of knitted and crocheted fabrics	SIC363	Mfg. of musical instruments
					SIC171	Preparation/spinning of textile ®bres	SIC267	Cutting, shaping and ®nishing of stone
					SIC173	Finishing of textiles	SIC154	Mfg. of vegetable and animal oils and fats
					SIC172	Textile weaving	SIC365	Mfg. of games and toys
Estimated input±output effect (IO)	0.0556	0.0172	0.2510	2.9055	SIC262	Mfg. of ceramic goods other than for construction	SIC232	Mfg. of re®ned petroleum products
					SIC263	Mfg. of ceramic tiles and ¯ags	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 and general hardware	SIC157	Mfg. of prepared animal feeds
Estimated knowledge	0.0182	0.0106	0.0888	−0.3387	SIC262	Mfg. of ceramic goods other than for construction	SIC233	Processing of nuclear fuel
spillovers effect (KS)					SIC265	Mfg. of cement, lime and plaster	SIC181	Mfg. of leather clothes
					SIC300	Mfg. of of®ce machinery and computers	SIC183	Dressing/dyeing of fur; mfg. of fur articles
					SIC171	Preparation/spinning of textile ®bres	SIC362	Mfg. of jewellery and related articles

Notes: The table presents descriptive statistics of estimates of the effect of labour pooling (LP), input±output (IO) and knowledge spillovers (KS) on industrial coagglomeration. Coef®cients 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 12 years. Our dataset comprises of 97 industries. The full set of estimates is presented in Online Appendix Table 1.

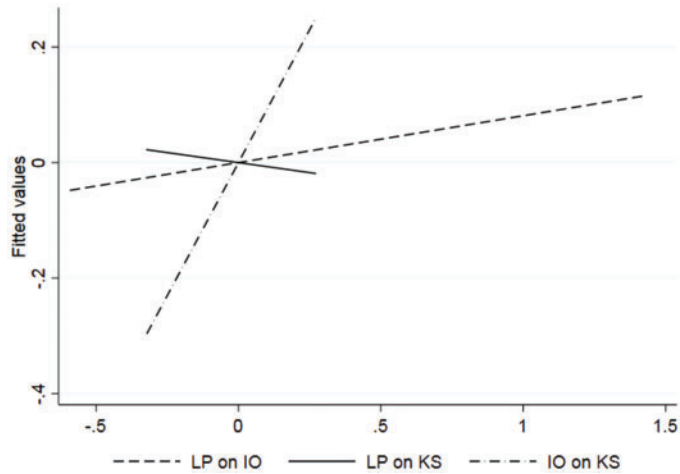

Figure 2. Associations (linear ®t) between Marshallian forces across industrial sectors.
Notes: The plots present linear-fit lines obtained from regressing one Marshallian force on another Marshallian force as detailed in the legend. When pairing up Marshallian force, the one with the smallest amount 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 normalised to have zero mean so that all plots cross at the axis origins. Regression coefficients (standard errors) are as follows. LP on IO: 0.0811 (0.0696); LP on KS: −0.0694 (0.3810); and IO on KS: 0.9210 (0.5021).

concentrated in the interval (−0.2, 0.2) and around 25% of the industries displaying KS effects close to zero. Only approximately 16% of the estimates are significant. Lastly, we investigate the correlations between the three Marshallian forces across our industries. Our findings are presented in Figure 2, which displays linear predictions from univariate regressions of one of the three Marshallian forces on the other twoÐone at the time (e.g. LP on IO and then LP on KS). The figure shows that sectors with high LP also tend to have high IOÐbut these same sectors tend to have low KS. Conversely, we find that the association between IO and KS is positive.[15]

It is also interesting to reflect on the nature of the `top four' and `bottom four' sectors with the highest/lowest microfoundation estimates (Table 3). Starting with LP, all industries in the top four belong to the textile sectorÐone of our tales. The bottom four sectors are instead very different from each other. `Manufacture of musical instruments' (SIC363) and `Cutting, shaping, and finishing of stone' (SIC267) are likely to have highly specialised workersÐa situation where pooling may not be possibleÐwhile `Manufacture of games and toys' (SIC365) and `Manufacturing of vegetable and animal oils and fats' (SIC154) are likely to involve standardised labourÐwhere pooling may not be needed. Turning to IO, one of our talesÐcutleryÐis among the top four sectors. Two of the other three top industries are in the same sectoral divisionÐthat is, the manufacture of ceramics (SIC262 and SIC263)Ðwhich is arguably another classic: Stoke-on-Trent (more generally

15 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 on 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.

Staffordshire) hosts highly concentrated and specialised pottery and ceramic-related productive activities. Conversely, little stands out considering the bottom four IO sectors. When looking at the top four KS sectors, we find that one of the classic talesÐthat is, the computer industryÐclearly ranks very high in terms of its responsiveness to knowledge flows. We also find `Preparation/spinning of textile fibres' (SIC171) among the top four sector and two sectors in the ceramic-related compartment. Interestingly, none of these sector has a high share of college graduates making them different from computers. At the opposite end, the bottom four sectors form a disparate group including a high-skilled sectorÐ`Processing of nuclear fuel' (SIC233; share of graduates 0.317)Ðand a very unskilled oneÐ`Manufacture of clothing and leather' (SIC181; with zero college graduates).

We conclude this section by carrying 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 Online Appendix Table W1 on the basis of the strength of the Marshallian force that best identifies a given taleÐfor 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 taleÐthat is, 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 forcesÐ for example, they are similar to cutlery along LP and KS in terms of both strength and statistical significance (bearing in mind that we are already focusing 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 potential 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: `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 focusing 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 compartmentÐand displays too large a coefficient on KS. 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 are `Manufacture of accumulators, cells and batteries' (SIC314; LP = 0.007, insignificant; IO = −0.039, insignificant; KS = 0.141, insignificant)Ðalthough the impact of KS is not significant despite being the dominant force. Lastly, 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 similar pattern.

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. Nevertheless, 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.

5. Understanding the patterns

5.1. Theoretical foundations and empirical approach

This section systematises the patterns of microfoundations that are present in the individual industry models from Sections 3 and 4. This involves estimating models of the relationship between four key industry characteristicsÐnamely, localization, new firm entry, workforce education and size of incumbent firmsÐand the estimated individual industry coefficients on Marshallian microfoundations. This approach extends 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 is that here we consider individual industriesÐas 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 our analysis, 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 localisation. This question is related to the `old' urbanisation versus localisation debate, where the focus is on whether agglomeration economies depend primarily on the scale of the entire city (urbanisation) or that of an individual industry (localisation). See, for instance, Glaeser *et al.* (1992) and Henderson *et al.* (1995), or the survey by Rosenthal and Strange (2004). On the empirical side, some recent research shows effects that appear to operate within industries (Fallick *et al.*, 2006), while other research finds effects operating between industries (Ellison *et al.*, 2010). Theoretically, it is straightforward to conceive a model where both effects are at workÐwith the agglomeration of an industry being either a substitute or a complement to the coagglomeration of that industry with other sectors. The substitution effect of agglomeration would works as follows: if the presence of own industry activity creates an external increasing return within the industry, then cross-industry coagglomeration might not be as valuable. The complementarity argument would suggests instead that industries that benefit from Marshallian forces 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 argumentsÐand some empirical evidenceÐsuggesting that both complementarity and substitution can be at work. We will consider this issue by relating an industry's Ellison and Glaeser (1997) Index (EGI) of agglomeration to the industry-level coefficients that capture the strength of LP, IO, and KS.

Second, we study how industry dynamism and entrepreneurship relate to the microfoundations of agglomeration. 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 labour market matching, more instability would be reflected in a greater loss associated with poorly matched employers and workers. This, in turn, raises 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 more `entrepreneurial' industry will benefit more from coagglomeration with related industries. However, dynamism might similarly 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 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 entrepreneurship 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 we consider is the relationship of workforce education with agglomeration. It is common to equate education with skills. 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 tradesÐlike Marshall's cutlery workers, discussed earlier in the article. If educated workers have more specialised skills, then labour 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 being especially strong 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 productsÐsuggesting 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 to have knowledge that might spill over, there is ambiguity: Marshall's cutlery workersÐwhile clearly skilledÐalmost certainly did not hold university degrees. To consider these issues empirically, we investigate the relationship between educationÐspecifically the share of graduates in the industry's workforceÐand the industry-by-industry Marshallian coefficients estimated in Section 4.

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 one hand, Chinitz's (1961) classic paper argues that small firms have larger effectsÐin particular, by fostering input sharing linkages. See Rosenthal and Strange (2003, 2010) for empirical results consistent with this idea. On the other hand, other empirical work (Agrawal and Cockburn, 2003; Feldman, 2003) shows `anchor' effects whereby large firms have important externalities. In our analysis, we reassess these questions by studying the relationship between an industry's mean employment of incumbent firmsÐthat is, those already in the marketÐand that industry's Marshallian coefficients.

In sum, theory gives us ambiguous predictions regarding how the estimated LP, IO and KS forces relate to industry characteristics such as localisation, entry share, workforce education and incumbent size. This is a parallel to Boschma (2005) and Caragliu and

Nijkamp (2012), who consider the many dimensions of proximity. These include geographic proximity (incorporated here through the metropolitan area level estimation of Marshallian coefficients) and organisational, institutional, social and cognitive proximityÐ which are at least partly captured by the industry characteristics we consider. It is worth pointing out that Boschma notes that such characteristics may have ambiguous effects on knowledge spillovers or learning (his primary focus). We will attempt to resolve these ambiguities by estimating models relating these industry characteristics to estimated Marshallian coefficients.

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

$$\beta_{ai} = \sum\nolimits_{h} \delta_h X_{hi} + \varepsilon_i, \tag{5.1}$$

where β_{ai} gives the value of the coefficient for Marshallian force aÐwith $a \in$\{LP, IS, IO\}Ðfor industry i (which we estimated using Equation (3.1) sector-by-sector), while X_{hi} represents the value of the industry characteristic hÐwith $h \in$\{EGI, Entry, Education, Incumbent\}Ðfor 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 estimate simple linear models like those indicated by Equation (5.1), as well as non-linear models in which dummies that capture quantiles of the underlying sectoral attributes are used to characterise 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 to look for non-linear relationships.

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-analysisÐwhere 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 equivalentÐgiven 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 (3.1) 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 weigh by the inverse of the standard errorÐwhich 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).[16] We also estimated unweighted models, which are reported in the Online Appendix. While the results from this second approach yield similar intuitions, the estimates are noisier and the patterns are less clear. However, unweighted models do not account for the fact that large estimated Marshallian coefficients need not be statistically significantÐand so should be `discounted' in our analysis. So we consider the unweighted

16 We also experimented with weights inversely proportional to the variance of our estimates. This approach returned similar patterns.

estimates less reliable. Finally, in all specifications we adjust standard errors for heteroskedasticity by using a standard `robust' variance±covariance matrix correction.

5.2. Results

Table 4 presents the results of weighted univariate regressions. For each Marshallian force, the first column reports the results of a continuous model, while the second column gives the 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 that the latter is more likely the case. The labour 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 negativeÐwhich is consistent with agglomeration and coagglomeration being complements.

The second set of results in Table 4 concerns entrepreneurship and industry dynamism as proxied by the entry share. With regard to labour 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 labour pooling can increase productivity in part by reducing unemployment when a city's employers experience labour 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 aboveÐwith `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 labour pooling. In this article, we see that industries with the very least educated workers have the largest coefficients for labour pooling. This presumably reflects labour pooling operating strongly outside of sectors with highly educated workers. This is consistent with the argument above that education is not identical to skills.

Table 4. Relationship between estimated Marshallian forces strength and sectoral characteristicsÐunivariate regression results; weighted by the inverse of standard error

		Labour pooling		Input sharing		Knowledge spillovers	
		(1)	(2)	(3)	(4)	(5)	(6)
EGIÐagglomeration index	Continuous	0.1505*** (0.051)		0.0006 (0.020)		0.0181** (0.009)	
	Top		0.2262*** (0.056)		−0.0007 (0.047)		0.0458* (0.024)
	Bottom		−0.0337*** (0.013)		−0.0021*** (0.006)		−0.0090 (0.006)
Entry share	Continuous	0.0131* (0.008)		0.0003 (0.004)		0.0037 (0.004)	
	Top		0.0071 (0.046)		−0.0083 (0.010)		−0.0006 (0.011)
	Bottom		−0.0329** (0.013)		0.0031 (0.007)		−0.0015 (0.009)
Share highly educated	Continuous	0.0053 (0.007)		0.0059 (0.004)		0.0077* (0.004)	
	Top		0.0080 (0.010)		0.0213*** (0.008)		0.0213* (0.012)
	Bottom		0.0501* (0.025)		−0.0084 (0.013)		−0.0041 (0.007)
Incumbent employment size	Continuous	−0.0002 (0.027)		0.0032 (0.013)		0.0023 (0.009)	
	Top		0.0024 (0.054)		−0.0203** (0.009)		−0.0009 (0.010)
	Bottom		−0.0258* (0.013)		−0.0136 (0.011)		−0.0171*** (0.006)

Notes: Robust standard errors are 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 ®rst 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).

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)Ðthough 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, a result consistent with the results for entry share discussed above. This is in line with the anchor hypothesis offered by Feldman (2003) and Agrawal and Cockburn (2003). Finally, for labour 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

Table 5. Relationship between estimated Marshallian forces strength and sectoral characteristicsÐmultivariate regression results; weighted by the inverse of standard error

		Labour pooling		Input sharing		Knowledge spillovers	
		(1)	(2)	(3)	(4)	(5)	(6)
EGIÐagglomeration Index	Continuous	0.1464***		0.0071		0.0171**	
		(0.055)		(0.020)		(0.008)	
	Top		0.2200***		0.0017		0.0478*
			(0.064)		(0.045)		(0.024)
	Bottom		−0.0272*		−0.0208***		−0.0063
			(0.017)		(0.007)		(0.006)
Entry share	Continuous	0.0051		−0.0007		0.0024	
		(0.009)		(0.003)		(0.003)	
	Top		−0.0006		−0.0080		−0.0002
			(0.050)		(0.010)		(0.014)
	Bottom		−0.0103		0.0140*		0.0079
			(0.018)		(0.008)		(0.005)
Share highly educated	Continuous	−0.0101		0.0061		0.0078**	
		(0.007)		(0.004)		(0.004)	
	Top		−0.0083		0.0077		0.0216*
			(0.012)		(0.008)		(0.012)
	Bottom		0.0465*		−0.0113		0.0028
			(0.027)		(0.017)		(0.008)
Incumbent employment size	Continuous	−0.0061		0.0016		−0.0043	
		(0.023)		(0.011)		(0.008)	
	Top		−0.0115		−0.0138		0.0014
			(0.054)		(0.014)		(0.012)
	Bottom		−0.0249		−0.0174		−0.0148
			(0.021)		(0.013)		(0.009)

Note: See footnote of Table 4.

explained by an organisational dimension of labour pooling: large firms can expand when their rivals are hit with negative shocks.[17]

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 labour 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 the 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.

17 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) but failed to find any striking patterns. Results are presented in the Online Appendix.

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 labour 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 organisation variablesÐnamely, entry share and size of incumbentsÐare 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 LP/IO/KS we observed in Table 4 are somewhat replicated in the multivariate models of Table 5Ðalthough 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 multivariate 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 -0.2116 association (5% significant) with the EGI index.[18] 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. 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 labour market pooling. Finally, regarding industrial organisation, univariate models show that dynamic industries see stronger labour pooling, while industries with large incumbents are less sensitive to input links to other industries. These results are, however, less robust to multivariate specifications.

6. Conclusions

This article 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 labour 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 universalÐsomething which Marshall himself never claimed to be the case. Some industries co-locate with other industries that have similar

18 The figures refer to unweighted 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.

workforce needs. Others instead co-locate with industries to which they are linked via supply chains or in knowledge. While the bulk of the literature has concentrated on determining which of the forces in the classic `Marshallian trinity' matters most for agglomeration, the evidence in our study calls for a change of focusÐwith a stronger emphasis on identifying which industries are more responsive to specific microfoundations.

Besides, our 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 industry pairs. This article, in contrast, provides evidence of heterogeneity in microfoundations at the level of the individual industry. The article 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 and knowledge) will also have a tendency to cluster (which presumably gives additional valuable technological links). The article 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 labour 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 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 similarÐsomething 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 article 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.

Supplementary material

Supplementary data for this paper are available at *Journal of Economic Geography* online.

Acknowledgements

We thank the Economic and Social Research Council (ESRC) UK (Grants ES/G005966/1 and ES/J021342/1) and the Social Sciences and Humanities Research Council (Grant 435-2012-0062) for

®nancial 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, William Kerr, two anonymous referees, Daniel McMillen, Henry Overman, Stuart Rosenthal, Matthew Turner and participants at the 2018 UEA Conference (New York), the 2018 NARSC Meetings (San Antonio), the 2018 Singapore Management University Conference on Urban and Regional Economic and the 2019 UEA Conference (Amsterdam). We are responsible for any errors or omissions. The authors disclose no con¯icts of interest in relation to the material published in this article.

References

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

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

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

Bacolod, M., Blum, B. S., Strange W. C. (2010) Elements of skill: traits, intelligences, and agglomeration. *Journal of Regional Science*, 50: 245±280.

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

Boschma, R. (2005) Proximity and innovation: a critical assessment. *Regional studies*, 39: 61±74.

Breschi, S., Lissoni, F. (2004) Knowledge networks from patent data: methodological issues and research targets. CESPRI Working Paper No. 150.

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

Caragliu, A., Nijkamp, P. (2012) The impact of regional absorptive capacity on spatial knowledge spillovers: the Cohen and Levinthal model revisited. *Applied Economics*, 44: 1363±1374.

Chatterji, A., Glaeser, E., 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., Grif®th R., 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., Puga, D. (2001) Nursery cities. *American Economic Review*, 91: 1454±1477.

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

Ellison, G., 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., Kerr, W. R. (2010) What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, 100: 1195±1213.

Fallick, B., Fleischman, C., Rebitzer J. B. (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: 311±329.

Faggio, G., Silva, O., Strange, W. C. (2017) Heterogeneous agglomeration. *Review of Economics and Statistics*, 99: 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., Kallal, H., Scheinkman, J., Schleifer, A. (1992) Growth in cities. *Journal of Political Economy*, 100: 889±927.
Helsley, R. W., Strange, W. C. (1990) Matching and agglomeration economies in a system of cities. *Regional Science and urban economics*, 20: 189±212.
Helsley, R. W., Strange, W. C. (2002) Innovation and input sharing. *Journal of Urban Economics*, 51: 25±45.
Helsley, R. W., Strange, W. C. (2004) Knowledge barter in cities. *Journal of Urban Economics*, 56: 327±345.
Henderson, J. V., Kuncoro, A., 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.
Kolko, J. (2010) Urbanization, agglomeration, and coagglomeration of service industries. In E. Glaeser (ed.) *NBER Agglomeration Economics*, pp. 151±180. Chicago, IL: University of Chicago Press.
Landes, D. S. (1969) *The Unbound Prometheus*. Cambridge: Cambridge University Press.
Marshall, A. (1890) *Principles of Economics*. MacMillan: London.
Neumark, D., Simpson, H. (2015) Place-based policies. *Handbook of Regional and Urban Economics*, Volume 5, Chapter 18, pp. 1197±1287. Elsevier.
O'Sullivan, A., 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, NY: The Free Press.
Rosenthal, S. S., Strange, W. C. (2003) Geography, industrial organization, and agglomeration. *Review of Economics and Statistics*, 85: 377±393.
Rosenthal, S. S., 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, pp. 2119±2172. Amsterdam: Elsevier.
Rosenthal, S. S., Strange, W. C. (2010) Big establishment/small effects: agglomeration, industrial organization, and entrepreneurship. In E. L. Glaeser (ed.) *Agglomeration Economics - NBER Collections*, pp. 277±302. Chicago, IL: University of Chicago Press.
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 Diversi®cation*. London: Routledge.
Vernon, R. (1960) *Metropolis 1985*. Cambridge, MA: Harvard University Press.ą**Appendix**

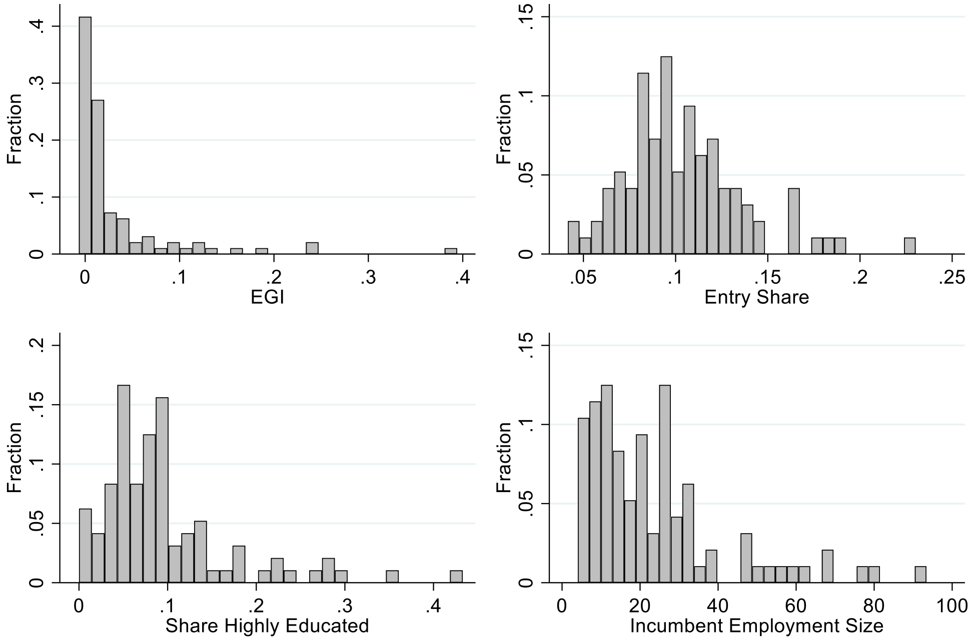

Figure A1. Distribution of industry characteristics.
Notes: Number of observations: 97 except for average continuous employment which excludes SIC233 (processing of nuclear fuel, an outlier with 399 employees). EGI is the Ellison and Glaeser Index of agglomeration. Descriptive statistics for the four 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).