Heterogeneous Agglomeration*

Giulia Faggio
Westminster Business School,
University of Westminster,
35 Marylebone Road, London, NW1 5LS
and Spatial Economics Research Centre,
London School of Economics,
Houghton Street, London, WC2A 2AE, UK
Tel.: +44 (0)20 7852 3592
G.Faggio1@lse.ac.uk

Olmo Silva
Department of Geography and Environment,
and Spatial Economics Research Centre,
London School of Economics,
Houghton Street, London, WC2A 2AE, UK
Tel.: +44 (0)20 7955 6036
O.Silva@lse.ac.uk (corrsponding author)

William C. Strange
Rotman School of Management,
University of Toronto,
105 St. George St.,
Toronto, ON M5S 3E6, Canada
Tel.: +1 (416)978-1949
wstrange@rotman.utoronto.ca

February 29, 2016

Web Appendix: Supplementary tables and figures

Table W1: Characterizing the sectoral breakdown

Sector pair is:	N. of Obs./Pairs	Mean γ^C	Mean LP	Mean IO	Mean KS (IOM)	Mean of Cut-off Variable	Mean E-G Localization Index
New	12972/1081	0.000	0.203	0.006	0.007	1972	0.039
Mixed	26508/2209	0.000	0.234	0.008	0.016	1958	0.027
Old	12972/1081	0.000	0.277	0.014	0.022	1945	0.014
Dynamic	12972/1081	0.000	0.219	0.007	0.012	0.131	0.032
Mixed	26508/2209	0.000	0.234	0.008	0.015	0.106	0.027
Steady	12972/1081	0.000	0.262	0.012	0.020	0.082	0.021
High tech.	7140/595	0.000	0.412	0.014	0.029	--	0.009
Mix tech.	24780/2065	0.000	0.221	0.007	0.016	--	0.023
Low tech.	20532/1711	0.000	0.195	0.010	0.011	--	0.037
High education	12972/1081	0.000	0.328	0.009	0.022	0.148	0.015
Mix education	26508/2209	0.000	0.219	0.008	0.014	0.097	0.027
Low education	12972/1081	0.000	0.184	0.011	0.012	0.047	0.038
Small entrants	12972/1081	0.000	0.248	0.005	0.016	5.676	0.018
Mix entrants	26508/2209	0.000	0.233	0.010	0.014	10.08	0.027
Large entrants	12972/1081	0.000	0.234	0.012	0.018	14.48	0.035
Small incumbents	12972/1081	0.000	0.240	0.005	0.019	10.54	0.017
Mix incumbents	26508/2209	0.000	0.233	0.009	0.014	22.11	0.027
Large entrants	12972/1081	0.000	0.243	0.012	0.016	33.69	0.036

Note: Number of pairs refers to unique (non-repeated) sector combinations. High-tech and low-tech industries are categorized according to the OECD classification (1997). High-education and low-education industries are classified according to the share of college graduates above and below the median across all years (the median is 0.0783). Education level calculated using the UK LFS 1995-1999 data. New/old industry pairs consist of industries where the first year of opening is above/below the median across all years (the median is 1967). Dynamic/steady industry pairs consist of industries where the share of entrants is above/below the median across all years (the median is 0.100). Small/large entrants refer to industry pairs where the average size of entrants is below/above the median across all years (the median is 8.59). Small/large incumbents refer to industry pairs where the average size of incumbents is below/above the median across all years (the median is 18.95). Mixed pairs consist of one old/big entrants/big incumbents/steady industry and one new/small entrants/small incumbents/dynamic industry. Variables in the penultimate column refer to the cut-off variables (first year of opening, entry share, size of entrants and size of incumbents) averaged across industry pairs. Mean Ellison-Glaeser localization index across all industries: 0.027 (std. dev.: 0.048).

Table W2: Additional regressions of coagglomeration measure γ^C on Marshallian forces

	KS measured as Industry of Manufacture (IOM)				*KS measured as Sector of Use (SOU)*				
	(1)	(2)	(3)	(4)					
Dependent variable/	γ^C	γ^C	γ^C	γ^C	γ^C	γ^C	γ^C	γ^C	γ^C
Timing is:	1997-2008	1997	2002	2008	1997-2008.	1997-2008	1997	2002	2008
Labor pooling (LP)	0.166	0.188	0.166	0.146	0.152	0.154	0.170	0.153	0.138
	(0.020)***	(0.024)***	(0.022)***	(0.020)***	(0.019)***	(0.019)***	(0.023)***	(0.021)***	(0.019)***
Input-output sharing (IO)	--	0.083	0.099	0.071	0.071	--	0.067	0.088	0.063
		(0.026)***	(0.028)***	(0.025)***	(0.024)***		(0.024)***	(0.027)***	(0.024)***
Knowledge spill. – IOM (KS)	0.026	0.028	0.029	0.022					
	(0.013)*	(0.014)**	(0.015)*	(0.013)					
Knowledge spill. – SOU (KS)					0.075	0.077	0.101	0.080	0.057
					(0.023)***	(0.022)***	(0.026)***	(0.025)***	(0.022)***
Input sharing	0.057	--	--	--		0.049			
	(0.028)**					(0.027)*			
Output sharing	0.025	--	--	--		0.021			
	(0.031)					(0.030)			

Note: See note to Table 1 for details on variable definitions and samples. Regressions include all Marshallian forces at the same time and control for dissimilarity in use of resources. Variables are standardized to have zero mean and unit standard deviation. Robust standard errors clustered on industry pairs are reported in parentheses.

Table W3: Further robustness checks and extensions

	(1)	(2)	(3)	(4)	(5)
	Staggered Marshallian forces	γ^C excluding London	Control for popul. density	Control for empl. density	Control for Herfind. Index
Labor pooling	0.160	0.139	0.142	0.158	0.165
(LP)	(0.019)***	(0.022)***	(0.018)***	(0.020)***	(0.020)***
Input-output sharing	0.081	0.119	0.091	0.088	0.083
(IO)	(0.026)***	(0.033)***	(0.026)***	(0.026)***	(0.026)***
Knowledge spillovers	0.023	0.017	0.026	0.023	0.026
– IOM (KS)	(0.014)*	(0.014)	(0.014)*	(0.014)*	(0.014)*

Note: See note to Table 1 for details on variable definitions. All regressions control for dissimilarity in use of resources. Variables are standardized to have zero mean and unit standard deviation. Robust standard errors clustered on industry pairs are reported in parentheses. Knowledge spillovers measure is based on probabilistic mapping – Industry of manufacturing (KS – IOM). Column (1) considers γ^C for years from 2000 and Marshallian forces calculated up to 1999. Column (2) excludes London from the calculations of γ^C. Column (3) controls for the average (un-weighted) population density of the TTWAs in which the two sectors are operating, averaged across industry pairs. Column (4) controls for the average (un-weighted) employment density of the TTWAs in which the two sectors are operating, average across industry pairs. Column (5) controls for the Herfindahl index of the two sectors, average across industry pairs.

Table W4: Fifteen most co-agglomerated industry pairs – based on coagglomeration measure γC in 1997

Rank	Industry 1	Industry 2	γ^C	1st TTWA	2nd TTWA	3rd TTWA
1	Ceramic goods other than construction	Ceramic tiles & flags	0.105	Stoke-on-Trent	Exeter	London
2	Knitted & crocheted fabrics	Knitted & crocheted articles	0.086	Leicester	Nottingham	Derby
3	Publishing	Jewellery & related articles	0.054	London	Birmingham	Sheffield
4	Spinning of textiles	Textile weaving	0.054	Bradford	Huddersfield	Leeds
5	Publishing	Printing & reproduction of recorded media	0.037	London	Manchester	Birmingham
6	Finishing of textiles	Knitted & crocheted articles	0.037	Leicester	Manchester	Nottingham
7	Finishing of textiles	Knitted & crocheted fabrics	0.035	Leicester	Nottingham	Manchester
8	Ceramic goods other than construction	Construction products in baked clay	0.033	Stoke-on-Trent	Crawley	Peterborough
9	Basic iron & steel & ferro-alloys	Cutlery, tools & general hardware	0.033	Sheffield	Birmingham	Wolverhampton
10	Basic iron & steel & ferro-alloys	Other first processing of iron & steel	0.033	Sheffield	Dudley	Swansea
11	Other first processing of iron &steel	Forging, pressing, stamping & roll forming of metal	0.031	Dudley	Birmingham	Sheffield
12	Tanning & dressing of leather	Footwear	0.031	Northampton	Hull	Glasgow
13	Knitted & crocheted articles	Footwear	0.030	Leicester	Northampton	Blackburn
14	Iron & steel tubes	Other first processing of iron & steel	0.029	Dudley	Birmingham	Sheffield
15	Spinning of textiles	Finishing of textiles	0.028	Bradford	Manchester	Huddersfield

Note: γ^C is coagglomeration at TTWA level and based total employment. See note to Table 1 for more details.

Table W5: The relationship between coagglomeration γC, Marshallian forces and non-Marshallian mechanisms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Additional control details:	Year of Opening	Entry Share	Tech.	Education	Size of Entrants	Size of Incumbents	Joint Controls
Labor pooling (LP)	0.166	0.165	0.188	0.176	0.164	0.161	0.189
	(0.020)***	(0.020)***	(0.022)***	(0.021)***	(0.020)***	(0.019)***	(0.023)***
Input-output sharing (IO)	0.085	0.082	0.075	0.079	0.080	0.081	0.073
	(0.026)***	(0.025)***	(0.025)***	(0.025)***	(0.025)***	(0.025)***	(0.025)**
Knowledge spill. – IOM (KS)	0.028	0.024	0.027	0.029	0.023	0.024	0.036
	(0.013)**	(0.014)*	(0.014)**	(0.014)**	(0.014)*	(0.014)*	(0.014)***
Year of opening	0.046						0.048
	(0.014)***						(0.014)***
Entry share		0.003					0.003
		(0.012)					(0.014)
High tech			-0.227				-0.223
			(0.055)***				(0.065)***
Mix tech			-0.087				-0.083
			(0.035)***				(0.039)**
Share college graduates				-0.059			-0.045
				(0.018)***			(0.021)**
Size of entrants					0.028		0.009
					(0.014)*		(0.020)
Size of incumbents						0.032	0.058
						(0.016)*	(0.025)**

Note: See note to Table 1 for details on variable definitions. Specifications as in Column (4) of Table 2 plus the additional control variables listed in the leftmost column. See Appendix Table W1 and Section II.C for more detail.

Table W6: The heterogeneous relationship between coagglomeration γ^C and Marshallian forces – Regional level of aggregation

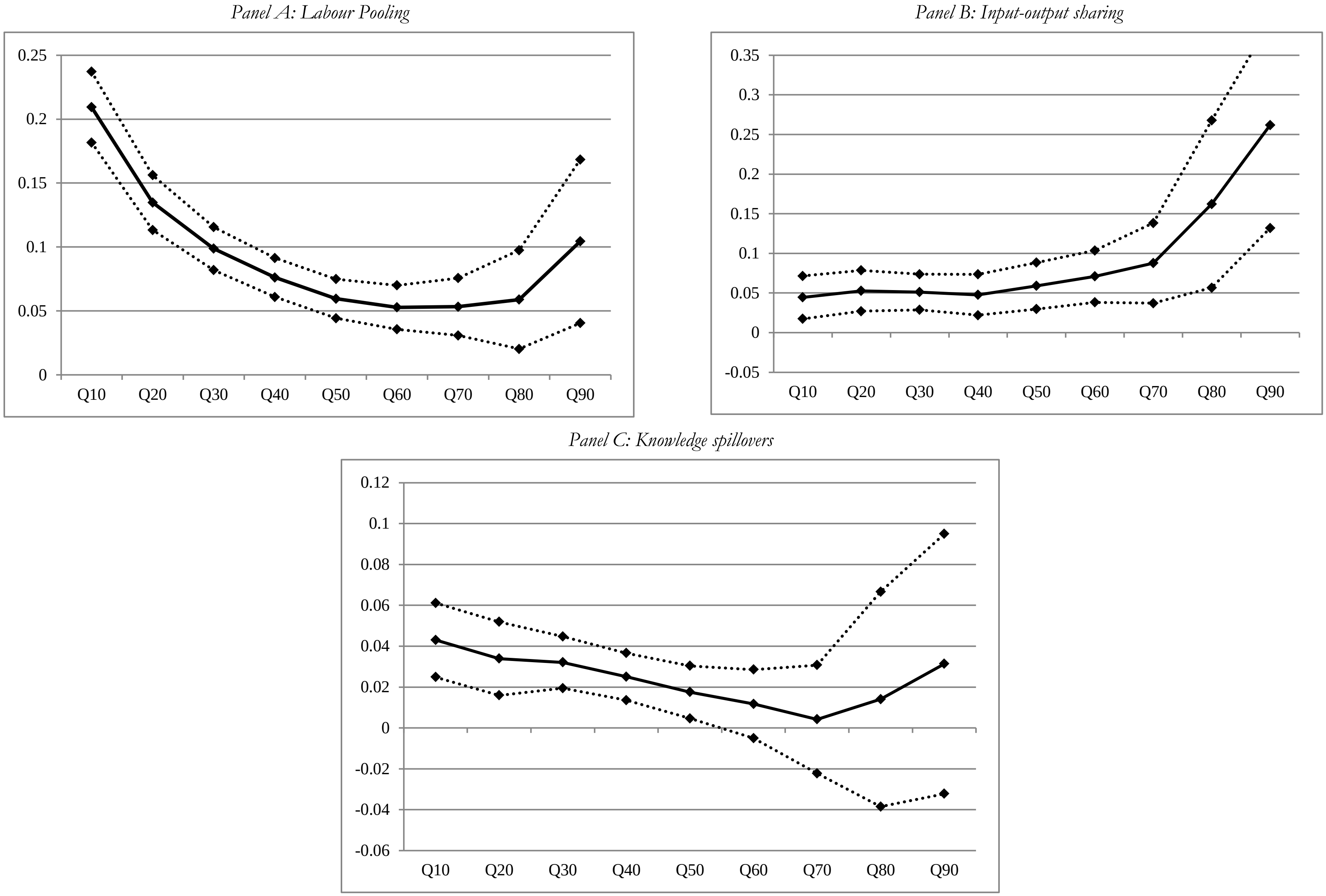
Panel A: Adaptation	Dynamic	Mixed	Steady
Labor pooling (LP)	0.134	0.122	0.161
	(0.048)***	(0.025)***	(0.028)***
Input-output sharing (IO)	0.184	0.127	0.059
	(0.106)*	(0.038)***	(0.023)**
Knowledge spillovers – IOM (KS)	0.209	0.030	-0.018
	(0.071)***	(0.016)*	(0.015)
N of. Observations/Pairs	12972/1081	26508/2209	12972/1081

Panel B: Technology	High-tech	Mixed-tech	Low-tech
Labor pooling (LP)	0.096	0.127	0.197
	(0.024)***	(0.021)***	(0.040)***
Input-output sharing (IO)	0.009	0.096	0.135
	(0.016)	(0.028)***	(0.051)**
Knowledge spillovers – IOM (KS)	0.074	0.022	0.039
	(0.023)***	(0.016)	(0.037)
N of. Observations/Pairs	7140/595	24780/2065	20532/1711

Panel C: Organization	Small incumbents	Mixed incumbents	Large incumbents
Labor pooling	0.102	0.161	0.162
	(0.021)**	(0.025)***	(0.049)***
Input-output sharing	0.138	0.127	0.069
	(0.042)***	(0.033)***	(0.054)
Knowledge spillovers – IOM	0.062	0.007	0.035
	(0.022)***	(0.018)	(0.028)
N of. Observations/Pairs	12972/1081	26508/2209	12972/1081

Note: See note to Table 1 and Tables 5 to 7 for details on variable definitions. Number of pairs refers to unique (non-repeated) sector combinations. All regressions control for dissimilarity in use of resources. Regressions further control for the following variables averaged across sector pairs: entry share (top panel); size of entrants (bottom panel). Variables are standardized to have zero mean and unit standard deviation. Robust standard errors clustered on industry pairs are reported in parentheses.

Figure W1: The effect of Marshallian forces at difference quantiles of γ^C – Regional level of aggregation



Note: See note to Table 1 for details on variable definitions. Variables standardized to have zero mean and unit standard deviation. The figures plot regression coefficients (solid lines) and 95% confidence intervals from quantile regressions that simultaneously include all three Marshallian forces. Confidence intervals from bootstrapped standard errors clustered on industry pairs. Regressions control for dissimilarity in use of resources.

Web Appendix: Data construction and robustness

Data construction

<u>The Business Structure Database (BSD)</u>

Our measures of coagglomeration of UK manufacturing sectors are constructed aggregating micro-level data from the Business Structure Database (BSD) covering the period 1997 to 2008. The data is an annual snapshot (taken in April at the closing of the fiscal year) of the Inter-Departmental Business Register (IDBR), which consists of constantly-updated administrative data collected for taxation purposes. Any business liable for value-added taxation (VAT) and/or with at least one employee registered for tax collection appears on the IDBR. Estimates produced by the Office for National Statistics (ONS) in 2004 show that the businesses listed on the IDBR account for almost 99 per cent of economic activity in the UK.

The data are structured into enterprises and local units. An enterprise is the overall business organization. The local unit can be thought of as a plant or establishment. In the majority of cases (70 per cent), enterprises only have one local unit, while the remaining 30 per cent of the cases represent enterprises with multiple local units. In our work, we make use of data at the local unit level including plants belonging both to single- and multi-plant enterprises and located in England, Wales and Scotland. We neglect Northern Ireland because of poor data coverage.

The initial raw data includes approximately three million local units every year. However, we carry out a series of checks and drop a number of units. First, we investigate the consistency of opening and closing dates of BSD units with their actual existence in the dataset and drop a number of anomalous cases where we identify establishments opening/closing in a specific year, disappearing/reappearing in a subsequent year only to open/close again in a subsequent wave. Stated differently, we only count firms' birth and death once. Second, we check the consistency of units' postcodes and sectors of activity over the years, and drop cases with missing or anomalous information. [1] For example, when we observe two or more plants operating in the same 3-digit industry, sharing the same postcode and being part of the same enterprise, we believe this being a reporting error and drop them. Similarly, we observe a number of same-postcode same-three-digit industry combinations representing anomalous concentration of identical activities at a single address. We believe this is another coding error and drop the plants that belong to the top 5% of the distribution of the number of plants sharing same three-digit industry and the same postcode. Finally, we drop active units with zero employment since this figure includes the owners/managers of the establishment, so it cannot be zero for an active unit, as well as units with an unusually large size (i.e., total employment above the 99th percentile of the distribution for each three-digit industry sector). After applying these restrictions, our dataset still comprises of more than two million plants annually over 12 years (1997-2008).

[1] A UK postcode usually corresponds to a very limited number of addresses or a single large delivery point. While it might not always be a geographically accurate description of where a property is located, it is generally a good approximation. For instance, a building which contains several businesses, but only one external door will only have the external door listed as a delivery point. This example shows that UK postcodes are geographically accurate up to the level of a front door in a particular street.

In terms of industrial classification, we focus on manufacturing and adopt the three-digit Standard Industry Classification (SIC) 1992. The UK SIC is a system for classifying industries by a five-digit code similar to the US SIC used prior to the introduction of the six-digit North American Industry Classification System (NAICS code) in 1997. We also apply a number of restrictions and re-combine a number of sectors to avoid having a limited or erratic evolution in the number of plants and employment during the sample period. Specifically, we exclude Tobacco (SIC160) because the number of plants in this sector tends to be small throughout the sample period (e.g. 43 in 1997). In addition, we combine Leather (SIC181) and Fur Clothes (SIC183) with Other Wearing Apparel (SIC182) to avoid small sample size problems in SIC181 and SIC183. For similar reasons, we also combine the following industries: Manufacture of Vegetal and Animal Oils and Fats (SIC154) with Other Food Products (SIC158); Reproduction of Recorded Media (SIC223) with Printing (SIC222); Coke Oven Products (SIC231) and Processing of Nuclear Fuel (SIC233) with Refined Petroleum Products (SIC232); Man-made Fabrics (SIC246) with Other Chemical Products (SIC247); Articles of Concrete, Plaster and Cement (SIC266) with Manufacture of Cement, Lime and Plaster (SIC265). Our final sample consists of 94 manufacturing 3-digit sectors for a total of 4,371 unique pairwise correlations a year for twelve years (1997-2008).

In terms of geography, our unit of aggregation is the Travel-to-Work Area (TTWA). These are entities constructed 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. TTWAs were devised to delineate areas that can be considered as self-contained labor markets and economically relevant aggregates. In 2007, there were 243 TTWAs within the United Kingdom. In most of our work, we mainly focus on 84 urban TTWAs with population in excess of 100,000. The reason why we focus on TTWAs with more than 100,000 inhabitants is that they cover urban areas where most of the productive activities take place. Rural areas in the UK refer to fairly sparsely populated areas such as the Scottish Highlands, the Welsh Mountains and the Peak and Lake Districts. Very few productive activities are concentrated in these areas – with the exception of tourism and related services. These industries are not covered by our analysis which focuses on solely manufacturing. However, in some robustness checks, we extend our analysis to include all TTWAs – urban and rural, irrespective of their population. The correlation between the coagglomeration metric γ^C that we use in the paper and the γ^C measure for all TTWAs (urban and rural) is very high, at 0.993. This backs our intuition that urban areas drive manufacturing coagglomeration patterns. Note that we aggregate the individual areas of Clacton, Colchester, Lincoln, Grantham, Torquay, and Paignton-Totes into the following urban TTWAs: *(1)* Clacton & Colchester; *(2)* Lincoln & Grantham; and *(3)* Torquay & Paignton-Totes. Even before the aggregation, the areas of Colchester, Lincoln and Torquay each had a population above the 100,000 threshold.

UK Labour Force Survey (LFS)

The UK Labour Force Survey (LFS) is a quarterly representative survey of households living at private addresses in the UK and is conducted by the Office for National Statistics (ONS) to collect information about individuals' labor market experiences. In our analysis, we use the years between 1995 and 1999 which allow for a consistent coding of the industrial and occupational classification of workers' jobs.

Each quarter of the LFS contains between 64,000 (earlier years) and 52,000 (later years) households, equivalent to about 120,000-150,000 individuals. In our analysis, we focus on 16-59 aged women and 16-64 aged men, and on individuals either working as employees or as self-employed. Excluding self-employed individuals does not affect our analysis.

In order to assign each individual to a TTWA, we retain workers living in England, Scotland and Wales (LFS data for Northern Ireland has poor coverage), and with a valid geographical identifier, namely the ward of residence (roughly equivalent to a US census tract). Additionally, we select individuals with non-missing information on: *(i)* gender and age; *(ii)* educational qualifications; *(iii)* industry and occupation. We exclude people working for the armed forces.

These restrictions leave us with a set of approximately 200,000 individuals each year for a total of 1.03m, of which 820,000 and 210,000 live in urban and rural areas, respectively. Next, we select individuals living in urban areas and working in manufacturing only. The final sample consists of about 35,000 workers a year for a total of 166,000 individuals. We use the UK Standard Industrial Classification (SIC) 1992 and the UK Standard Occupational Classification (SOC) 1990 at the three-digit level for these individuals' jobs to construct a proxy for the extent of labor pooling occurring between manufacturing sectors.

UK Input-Output Tables

To capture the flow of goods between industry pairs, we use the ONS Input-Output Analytical Tables for 1995 to 1999. For each industry, we calculate the shares of inputs bought from/sold to other industries as a fraction of the total intermediate inputs/outputs. Note that we exclude direct sales to consumers.

The sector classification in the I-O Tables is more aggregated than the three-digit SIC classification we use and only includes 77 manufacturing industries. In order to assign I-O shares to a SIC three-digit sector belonging to the same I-O sector, we use an apportioning procedure based on their employment share within the group averaged over 1995-1999. These shares are obtained using the relevant years of the BSD (our main dataset).

The EPO-CESPRI Dataset

The main data source for our analysis of patent citation flows is the EPO-CESPRI data provided by Bocconi University. This database provides cleaned and consistently coded information extracted from the European Patent Office (EPO) data for the period 1977 and 2009. Approximately 144,000 patents were filed by 160,000 UK inventors (multiple-investors can be recorded for each patent). These generate a stream of more than 77,000 citations of UK patents over the observed time-window.

In order to construct knowledge spillover measures we impose a number of restrictions. First, we exclude self-citations from the same inventor or the applying company at which he/she is based. Second, we exclude citing patents filed after 2000 and before 1981, and cited patents filed after 1997. The aim of these restrictions is twofold: *(a)* we want to guarantee that on average cited patents are at least three years older than citing ones; *(b)* we want to guarantee that our knowledge-spillover measures are constructed for the initial years of our sample (i.e. up to 2000) so that they are measured at a similar time as the labor-pooling and input-output sharing metrics. Expanding the sample to include all years does not affect the results.

It should be noted that while the US Patent and Trademark Office (USPTO) requires patent applicants to declare all relevant references and citations, the EPO does not apply this rule and all citations come directly from the patent examiners. As a result, the average number of EPO patent citations is much smaller than the corresponding figure for USPTO patents, and EPO numbers do not suffer from UPSTO-type "citation inflation" (see Hall et al. 2000). According to Breschi and Lissoni (2004), USPTO patents cited approximately 13 other patents and received on average 10.2 citations. The corresponding numbers for EPO patents are much lower at 4 and 2.8, respectively.

Patents in the EPO dataset (as in any other patent database) are categorized using technological classes rather than a standard industrial classification. To create a mapping industrial sectors and technological classification, we follow the literature and adopt two approaches: *(1)* a probabilistic mapping based on the Industry of Manufacture (IOM); and *(2)* an alternative probabilistic mapping based on the Sector of Use (SOU). These are based on correspondences developed by Silverman (2002) who studied approximately 150,000 patents filed at the Canadian Patent Office between 1990 and 1993. More information is available from Silverman's website:
http://www-2.rotman.utoronto.ca/~silverman/ipcsic/documentation_IPC-SIC_concordance.htm

The OECD Technology Classification

Based on the intensity of both direct R&D (i.e. R&D expenditure) and indirect R&D (i.e. embodied technology flows) in the output of manufacturing sectors across 10 OECD countries over the period 1980 to 1996, the OECD classifies as high tech or medium-high tech the following manufacturing industries: SIC241 "Manufacturing of basic chemicals" to SIC246 "Manufacturing of other chemicals & man-made fibres"; SIC291 "Manufacturing of other machinery for production/use of mechanical power N.E.C." to SIC297 "Manufacturing of domestic appliances"; SIC300 "Manufacturing of office machinery & computers"; SIC311 "Manufacturing of electric motors, generators & transformers" to SIC316 "Manufacturing of electrical equipment N.E.C."; SIC231 "Manufacturing of electronic valves, tubes & electronic components" to SIC323 "Manufacturing of TV/radio receivers & sound/video recording/reproducing"; SIC331 "Manufacturing of medical, surgical & orthopedic equipment" to SIC335 "Manufacture of watches & clocks; SIC341 "Manufacturing of motor vehicles" to SIC343 "Manufacturing of parts & accessories for vehicles/engines"; and SIC352 "Manufacturing of railway/tramway locomotives/rolling stock" to SIC355 "Manufacturing of other transport equipment N.E.C.". See OECD (1997) for more details.

US Data for Instrumental Variables

In order to address potential endogeneity issues, we follow Ellison et al. (2010) and instrument our UK-based proxies for the three Marshallian forces using almost identical measures obtained from US data.

Starting with labor pooling, we create a measure of the similarity in the occupational inputs of two industries using the National Industrial-Occupation Employment Matrix (NIOEM) published by the Bureau of Labor Statistics. Following the approach we have taken for the UK data, we construct the shares of different types of workers used in each manufacturing sector and then correlate the percentage of different types of occupations across industry pairs to obtain a proxy for labor sharing. In order to link this proxy to our data, we map US NIOEM industry codes to UK SIC codes. Since the US manufacturing classification is less detailed than the one that we adopt (79 vs. 94 sectors), we attribute the same US industry-occupation shares to multiple UK sectors. Note also that we construct the US labor correlation measure using all available data spanning the period 1983 to 1998. Restricting the calculations of this instrument to the period 1995-1998 does not affect our IV results.

We construct an instrument for input-output sharing following a similar approach. To begin with, we use the 1987 Benchmark Input-Output Accounts published by the Bureau of Economic Analysis (BEA) to measure the flows of intermediate inputs exchanged between US industries at the same level of aggregation as used in Ellison et al. (2010) and map these values from the 140 US manufacturing sectors to the 94 UK industries. In our regression analysis we focus on the maximum between the inputs and outputs that two industries are sharing irrespective of the direction of the flow (given that our data treats industry pairs symmetrically). Consistently, we use US data to construct a proxy for the maximum of the input-output linkages between industries and use this as an instrument.

Finally, we construct our instrument for knowledge spillovers using the NBER Patent Data initially assembled by Hall et al. (2001). The data cover patents granted by the US Patent and Trademark Office (USPTO) between 1975 and 1999 and record citation flows across patents. Following our main approach, we use a probabilistic mapping based on the industry of manufacture (IOM) to map technology to industrial classes and convert citations across US sectors to our UK classification based on 94 industries. Note that this instrument is different from the one adopted by Ellison et al. (2010) who used UK patents registered at the USPTO to instrument for US patents registered at the same office. Conversely, we use information coming from the USPTO about flows of citations among US patents to instrument for citations among UK patents registered at the European Patent Office (EPO). Part of the mechanical problems discussed by Ellison et al. (2010) in relation to this instrument is thus by-passed.

Robustness

The paper discusses a range of robustness tests. We now detail an additional set of checks. The broad pattern of finding persists throughout. The results in Tables W2 and W6 and Figure W1 are discussed in the body of the paper.

To start with, note that in our analysis in Section III, we follow Ellison et al. (2010) and do not correct γ^C for differences in the variance of area-industry employment shares. We assess the robustness of our findings against this issue by including in our specification industry i and industry j dummies. Results are very similar for labor-pooling, slightly smaller, but still significant for input-output, and larger and more precisely estimated for knowledge spillovers.

Next, Table W3 presents additional robustness checks on our main specification (discussed in Section III). Column (1) mitigates concerns with reverse causation by staggering our regressions and considering the effect of the three Marshallian forces measured up to 1999 on coagglomeration γ^C for the years 2000-2008. This check confirms our previous results. Columns (2) to (5) investigate whether any correlation between agglomeration and coagglomeration has the potential to bias our findings. To begin with, we exclude London – the biggest agglomeration in the UK – from the calculations of γ^C and re-estimate our empirical models. Although we find that the effect of KS is attenuated, our broad conclusions are unaffected. In Columns (3) and (4) of the table we include proxies for the extent of agglomeration of the areas where the two sectors in the pair are operating. In particular, we include: (i) the mean population density of all the TTWAs in which the sectors are operating, averaged across the pair (Column 3); and (ii) the mean employment density of all the areas where the sectors are operating, averaged across the pair (Column 4).[2] Employment density is calculated as total employment across all sectors in a TTWA divided by the area size expressed in square kilometers, so this proxy captures general urbanization economies – much as population density – stemming from operating in a larger market.[3] Adding these controls to our regressions has little effect on our estimates. Finally, in Column (5) we add to our specification the average Herfindahl index across the sector pair to check whether industrial concentration (as opposed to urbanization economies) affects our findings. Once again, we find no evidence that our results are sensitive to these considerations and confirm our previous conclusions.

Table W5 presents results of models where, instead of partitioning industries into groups as in Tables 4-6 in the main text, we include average values across the industries in the pair of the partitioning variables used in Tables 4-6. These are: firm age (year of opening); entry share; high/mixed tech (low tech omitted category); share of college graduates; size of entrants; and size of incumbents. First, we note that results are comparable to those presented and discussed elsewhere for the three proxies for the Marshallian forces. We also note that do not adopt this as our preferred specification because of its loose link to theory, as discussed in the text. We therefore only comment briefly on the results for the non-Marshallian variables. We find a positive and significant coefficient on year-of-opening (the inverse of age) in Column (1) and an insignificant coefficient on entry share in Column (2). The former result is consistent with the nursery city/unplanned interactions ideas discussed in the text. The latter is weakly supportive. The dummies for high- and mixed-technology pairs in Column (3) are instead both negative and significant. Controlling for Marshallian forces, we see more coagglomeration of low-technology industries. This result is the opposite of what one might expect to find based on the predictions of a nursery city model. In Column (4), the coefficient of average college share is significant and negative. Given the strength of human capital effects in other models (e.g., Rauch, 1993, or Rosenthal and Strange, 2008), this is unexpected.[4] Finally, entrant size has a positive and marginally significant coefficient in Column (5), as does incumbent size in Column (6). Controlling for Marshallian forces, we do not find much of a small firm effect. As shown in Column (7), jointly controlling for all the non-Marshallian factors does not affect our conclusions. All coefficients retain their sign, size and significance – with the exception of the proxy for the size of entrants. This is not surprising given its high

[2]Controlling for the dissimilarity of employment density, instead of the mean, does not affect our results.

[3]The correlation between the two urbanization proxies and coagglomeration is small and negative at -0.148 for population density and -0.082 for employment density. These numbers shrink to zero and 0.007 if we exclude London.

[4]These results hold if we exclude London. Conversely, dropping the three Marshallian proxies from the specifications yields insignificant estimates of the effect of either human capital or technology on coagglomeration.

correlation (0.712) with the proxy for the size of incumbents. In sum, simply including controls for non-Marshallian forces using industry-pair averages in a coagglomeration/microfoundation model, while failing to allow for heterogeneity, generates weak and sometimes puzzling results.

Finally, we investigate the robustness of our results in Table 4 in the main text by studying whether new/old pairs respond differently to Marshallian linkages when they are measured closer/further in time relative to coagglomeration. To do this, we run separate regressions for 1997, 2002 and 2008. The patterns presented in Table 4 are confirmed with no evidence of additional significant heterogeneity.

References

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

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

Hall, B., A. Jaffe and M. Trajtenberg (2000), "Market Value and Patent Citations: A First Look," NBER Working Paper 7441.

Hall, B., A. Jaffe and M. Trajtenberg (2001), "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tool", NBER Working Paper 8498.

OECD (1997), "Revision of High-Technology Sector and Product Classification", STI Working Paper 1997/2.