

Behrens, Kristian; Brown, Mark; Bougna, Théophile

Conference Paper

The Determinants of Agglomeration Redux: Transportation, Trade, and Input-Output Linkages

54th Congress of the European Regional Science Association: "Regional development & globalisation: Best practices", 26-29 August 2014, St. Petersburg, Russia

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Behrens, Kristian; Brown, Mark; Bougna, Théophile (2014) : The Determinants of Agglomeration Redux: Transportation, Trade, and Input-Output Linkages, 54th Congress of the European Regional Science Association: "Regional development & globalisation: Best practices", 26-29 August 2014, St. Petersburg, Russia, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/124195>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The Determinants of Agglomeration Redux

Kristian Behrens* Mark Brown† Théophile Bougna‡

November 7, 2013

– Preliminary and incomplete. Do not quote or circulate. –

Abstract

We study the determinants of agglomeration of Canadian manufacturing industries from 1990 to 2009. In so doing, we revisit the seminal contribution by Rosenthal and Strange (2001, “The determinants of agglomeration”, *J Urban Econ* 50(2), 191–229) using a long panel and continuous measures of localization. We pay particular attention to the role of transportation costs – constructed using extensive Canadian trucking microdata – international trade exposure, and input sharing – constructed using micro-geographic location patterns of plants. We find that between 1990 and 2009, industry localization has persistently fallen. The average degree of localization decreased by 36% within 10km, by 22.6% within 100km, and by 11.3% within 500km. Declining localization is associated with import competition, particularly from low wage countries, increasing transportation costs and the spreading out of upstream input suppliers and downstream demand for intermediate inputs. While we find strong evidence of trade-driven changes in localization, we find less evidence for knowledge spillovers and labour market pooling as drivers in changes in localization.

Keywords: Location patterns; Manufacturing industries; Micro-geographic data; Canada.

JEL classification: R12; R30; L11.

*Canada Research Chair, Département des Sciences Économiques, École des Sciences de Gestion, Université du Québec à Montréal (ESG–UQAM), Canada; CIRPÉE; and CEPR. E-mail: behrens.kristian@uqam.ca

†Statistics Canada, Economic Analysis Division (EAD), Chief-specialist Regional and Urban Economic Analysis. E-mail: Mark.Brown@statcan.gc.ca

‡Département des Sciences Économiques, École des Sciences de Gestion, Université du Québec à Montréal (ESG–UQAM), Canada; and CIRPÉE. E-mail: bougna_lonla.theophile@uqam.ca

1 Introduction

This paper investigates the determinants of the geographical concentration of Canadian manufacturing industries over the past two decades. We revisit and build upon the seminal contribution by Rosenthal and Strange (2001) by using more sophisticated panel of micro-geographic measures of agglomeration, computed from geo-coded plant-level data, and more elaborate spatial proxies of input sharing. We also examine in detail how changes in transportation costs – constructed using extensive Canadian trucking microdata – and the international trading environment affect the spatial distribution of manufacturing industries. Both factors are predicted by theory to influence the geographic concentration of industry. Since contribution by Rosenthal and Strange (2001), two literatures have emerged. The first, based on Duranton and Overman (2005), seeks to measure localization using continuous measures that sidestep many of the issues associated with the modifiable areal unit problem (MAUP) associated with the Ellison and Glaeser (1997) index and provides a means to establish statistical inference. The second is a literature that attempts to identify the factors associated with the geographic concentration of industry, namely the Marshallian trinity – labour market pooling, density of input suppliers and knowledge spillovers (see Rigby and Essletzbichler 2001, Henderson 2003, Baldwin et al. 2008, and Baldwin, Brown, and Rigby 2010).

In line with this literature, this paper attempts to identify the influence of the Marshallian trinity on the localization of industry. However, we also try to push beyond the trinity to look at other factors that might drive localization. We pay particular attention to three aspects. First, we use a much more elaborate measure of transportation costs. Second, the panel nature of our data allows us to consider the impact of international trade on industry location, broken down by imports and exports and source countries. Third, we use a novel and much more detailed micro-geographic measure of input-output linkages than has been used before in the literature. Hence, on top of revisiting key questions in agglomeration economics following Rosenthal and Strange (2001), we provide much more detailed evidence on the impact of transport costs, international trade, and input-output linkages on the spatial structure of the economy.

Our key results can be summarized as follows. Over the twenty-year period between 1990 and 2009 industry localization has persistently fallen. Between 1990 and 2009, the average degree of localization of industries – measured as the proportion of plants pairs within given distances bounds – decreased by 36% for plant pairs within 10km of each other, 22.6% for plant pairs within 100km, and 11.3% for plant pairs within 500km. Declining localization is associated with increases in import competition, particularly from low wage countries, rising transportation costs and the spreading out of upstream input suppliers and downstream demand for intermediate inputs. While we find strong evidence of trade-drive localization, we find less evidence for knowledge spillovers and labour market pooling as drivers of localization. Turning briefly to the literature, many authors have established the important role of spatial concentration of manufacturing activity on productivity

and growth. There is, in particular, a large theoretical literature that develops classes of mechanisms in order to explain the existence of urban agglomeration economies. Marshall's trinity (labor market pooling, input sharing, and knowledge spillovers) are three key sources of agglomerative spillovers emphasized in that literature. Despite the existence of that extensive theoretical literature, the passage from modeling these mechanisms to identifying them empirically is not straightforward. This is mainly due to the Marshallian equivalence – i.e., all mechanisms share the prediction that productivity increases with the scale of an activity at a given location, irrespective of the channels through which the effects percolate (see Duranton and Overman, 2004). Thus, only few academic papers address the empirical question relative to the causes driving the agglomeration of those industries. In general, these studies have relied on the discrete indices.

The empirical evidence suggests that that the Marshallian trinity and a set of other factors all influence localization. Ellison and Glaeser (1999) applied the EG-index to a study of the importance of natural advantage to industrial agglomeration; they found that 20% of measured agglomeration levels arise through the natural advantage channels. Rosenthal and Strange (2001; 2004) regressed the EG-index on a selection of important industry characteristics that are: knowledge spillovers, labor market pooling, input sharing, product shipping costs, and natural advantage. Their results showed strong effects of all of these factors, with labor market pooling exhibiting the strongest correlation with agglomeration. Ellison, Glaeser, and Kerr (2010) test Marshall's theories of industrial agglomeration by examining which industries locate near one another, or coagglomerate. By examining coagglomeration patterns, they found that all three forces of Marshall's theories of agglomeration are supported and are similar in magnitude, with input-output flows being the greater among equals. Specifically, a one standard deviation growth in labor or input-output dependencies increases coagglomeration by around one seventh of a standard deviation. Using Canadian data, Behrens (2013) investigate the role of international trade in the de-concentration trend observes in the Canadian manufacturing industries with emphasis on textiles. His results suggest that growing imports from low-cost countries are associated with firm death and less localization, while growing exports to high-cost countries are associated with firm survival and more localization. Dumais, Ellison, and Glaeser (1997) analyze the impact of entry, exit, and firm growth on the geographic distribution of manufacturing employment in the US between 1972 and 1992. They find that firm entry has a dispersive effect, while firm exit is agglomerative. Devereux, Griffith, and Simpson (2007) examine whether discretionary government grants influence where domestic and multinational firms locate new plants, and how the presence of agglomeration externalities interacts with these policy instruments. Their results suggested that pre-existing agglomerative structures affect firm's location decisions upon industry entrance.

Of course, it is also well known that transportation costs are crucial determinant of industry location. A lower transport cost in a given industry is likely to cause agglomeration. The basic idea is that, in response to the lower transportation cost they face, firms will agglomerate their activities in the centre of markets (e.g., in the corridor between Toronto and Montreal) in order to take

advantage of scale economies. However, there is limited work into the elaboration of good measures of transportation costs, and even less to their application to the analysis of the determinants of agglomeration. For example, Rosenthal and Strange (2001), use the ratio of inventories to sales at the end of the year as a proxy for perishability of outputs, itself a proxy for transportation costs. Lu and Tao (2009) used a similar proxy i.e., the finished good to output ratio, where finished goods are goods that have been finalized but not been sold. We will use measures constructed from the micro-data files on truck shipments within Canada and between Canada and the US that is invariant to the spatial structure of industry and so side-steps the often endogenous nature of standard transportation measures (e.g, transportation margins from input-output accounts).

The remainder of the paper is organized as follows. First, we develop a descriptive analysis of the Duranton-Overman index over the 1990 to 2009 period. Here we ask whether industries are becoming more or less localized through time and whether industry characteristics are associated with changes in localization (Section 2). Second, we move from the descriptive work to a multivariate analysis where we test the effect of the primary factors thought to influence localization. These include our trade related measures, transportation costs, and measures related to input sharing, labour market pooling and knowledge spillovers (Section 3). Third, we move from the main findings to a set of robustness checks. These test whether our results are sensitive to the distance thresholds used to measure localization, the weightings used to construct the Duranton-Overman index (i.e., employment and sales weights as opposed to an unweighted index), and how transportation costs are treated in the model (Section 4). This is followed by the conclusion (Section 5).

2 Trends in localization from 1990 to 2009

We measure the geographical concentration of industries using the Duranton and Overman (2005, 2008) K -density measure of localization, which we henceforth refer to as the Duranton-Overman (DO) index for short (see Appendix C for additional details on these measures). This measure looks at how close establishments are relative to each other by considering the distribution of bilateral distances among them. It provides a very detailed micro-geographic description of location patterns, and allows for statistical testing of whether the patterns may be due to chance or not. We estimate the K -densities at the NAICS 6-digit level, which corresponds to 258 industries. We use NAICS classified geo-coded plant-level data from the ASM to compute great circle distances between pairs of plants, from which the DO indices are estimated.¹ Since the K -density is a distribution function,

¹The survey frame of the ASM has evolved over time. Early in the period, it was relatively stable with, on average, about 32,000 plants per sample year. The sample of plants was restricted to those with total employment (production plus non-production workers) above zero, and plants must have sales in excess of \$30,000. Also, aggregate records were excluded. These records represent multiple (typically small) plants without latitudes and longitudes. In 2000, however, the number of plants in the survey increased substantially as the ASM moved from its own frame to Statistics Canada's centralized Business Register, increasing the sample to an average of 53,000 plants. In 2004, however, the number of plants in the frame was once again restricted, with many of the small plants once again excluded, or

we can also compute its cumulative (CDF) up to some distance d . The CDF at distance d thus tells us what share of establishment pairs is located less than distance d from each other.

Table 9 in Appendix D summarizes the K -density CDF for the most localized industries in 1990, 1999, and 2009, respectively. To understand how to interpret the K -density CDF, take ‘Women’s and Girls’ Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing’ (NAICS 315231) as an example. In 1990, 61 percent of the distances between plants in that industry were less than 50 kilometers. Stated differently, if we pick two firms in that industry at random, there is a probability of 61 percent that these firms are less than 50 kilometers apart. If we, however, pick two firms at random among *all* manufacturing firms, the probability that those firms are less than 50 kilometers apart would only be about 7.6 percent in 1990 (see Table 1 below). Clearly, this large difference suggests that the location patterns of firms in the ‘Women’s and Girls’ Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing’ industry are very different from those of manufacturing in general. Firms in that industry are much closer than they ‘should be’ if they were distributed like overall manufacturing.

Whereas the standard K -densities are computed based on plant counts, we can also compute weighted versions (see Duranton and Overman, 2005). We compute, in particular, K -densities weighted by either plant-level employment or plant-level sales. For these weighted versions, the foregoing interpretations remain true, except that the unit of observation is now the employee or a dollar of sales. We will report results for the weighted measures only as robustness checks, since the qualitative patterns are similar to the ones of the unweighted measures.

2.1 General overview

This paper is motivated by the trends in industrial localization in Canada, and by what they can tell us about the underlying factors that influence the geographic concentration of industries. As a prelude to the econometric analysis to follow, the trends in the localization of Canadian manufacturing industries are tracked over the past 20 years, and related to changes in the underlying characteristics of these industries.

There is growing evidence that the localization of manufacturing industries in Canada has been declining over the first decade of the 2000s (see Behrens, 2013; and Behrens and Bougna, 2013). This can clearly be seen from Table 1 and Figure 1 below. Across all industries, there has been a nearly monotonic decline in the average value of the CDF between 1990 and 2009. For instance, the average CDF at 50 km distance was 7.6 percent in 1990, 6.2 percent in 1999, and 5.6 percent in 2009. This trend of geographical deconcentration holds at all distances, thus suggesting a widespread dispersal

included in aggregate records. With this in place, the sample returned to near previous levels, averaging about 33,000 plants between 2004 and 2009. As will become apparent in the descriptive analysis to follow, the expanded survey scope in the early 2000s had little effect on trends in the CDF, but there was an effect on the number of industries found to be localized/dispersed. The econometric analysis in Section 3 deals with the change in the sample frame through year fixed effects.

Table 1: Duranton-Overman industry CDFs, 1990 to 2009.

Year	Unweighted				Employment weighted				Sales weighted			
	10 km	50 km	100 km	500 km	CDF at distance of				10 km	50 km	100 km	500 km
1990	0.020	0.076	0.139	0.420	0.021	0.083	0.151	0.449	0.022	0.086	0.156	0.453
1991	0.019	0.076	0.139	0.423	0.022	0.083	0.152	0.447	0.023	0.087	0.156	0.453
1992	0.020	0.074	0.135	0.418	0.020	0.079	0.147	0.442	0.022	0.084	0.151	0.448
1993	0.019	0.072	0.132	0.416	0.020	0.079	0.145	0.440	0.021	0.082	0.148	0.446
1994	0.017	0.071	0.131	0.413	0.020	0.077	0.143	0.438	0.021	0.081	0.147	0.443
1995	0.017	0.068	0.126	0.402	0.019	0.076	0.141	0.432	0.020	0.080	0.145	0.438
1996	0.016	0.065	0.122	0.402	0.019	0.073	0.136	0.428	0.020	0.076	0.140	0.435
1997	0.016	0.066	0.123	0.401	0.017	0.072	0.135	0.427	0.019	0.077	0.140	0.433
1998	0.016	0.064	0.120	0.396	0.019	0.074	0.135	0.425	0.019	0.078	0.141	0.433
1999	0.015	0.062	0.118	0.398	0.017	0.072	0.134	0.426	0.018	0.076	0.139	0.434
2000	0.014	0.063	0.120	0.383	0.016	0.073	0.135	0.411	0.016	0.075	0.140	0.421
2001	0.013	0.061	0.118	0.383	0.015	0.072	0.136	0.412	0.016	0.076	0.142	0.421
2002	0.013	0.062	0.119	0.383	0.016	0.073	0.137	0.413	0.017	0.078	0.143	0.422
2003	0.013	0.060	0.117	0.384	0.015	0.072	0.137	0.416	0.016	0.075	0.141	0.422
2004	0.013	0.060	0.115	0.379	0.015	0.070	0.132	0.412	0.017	0.074	0.137	0.418
2005	0.012	0.059	0.113	0.379	0.014	0.068	0.130	0.409	0.016	0.072	0.134	0.415
2006	0.013	0.061	0.116	0.378	0.015	0.069	0.131	0.406	0.015	0.072	0.135	0.412
2007	0.012	0.057	0.110	0.374	0.015	0.064	0.122	0.399	0.017	0.069	0.127	0.406
2008	0.012	0.057	0.110	0.376	0.017	0.067	0.125	0.400	0.017	0.069	0.128	0.405
2009	0.013	0.056	0.107	0.373	0.015	0.063	0.121	0.397	0.017	0.068	0.126	0.403
Average	0.015	0.064	0.121	0.394	0.017	0.073	0.136	0.422	0.019	0.077	0.141	0.428
Change	-0.007	-0.021	-0.031	-0.047	-0.006	-0.019	-0.031	-0.051	-0.005	-0.018	-0.030	-0.050
	-36.0%	-27.1%	-22.6%	-11.3%	-28.7%	-23.3%	-20.3%	-11.4%	-21.5%	-21.2%	-19.3%	-11.0%

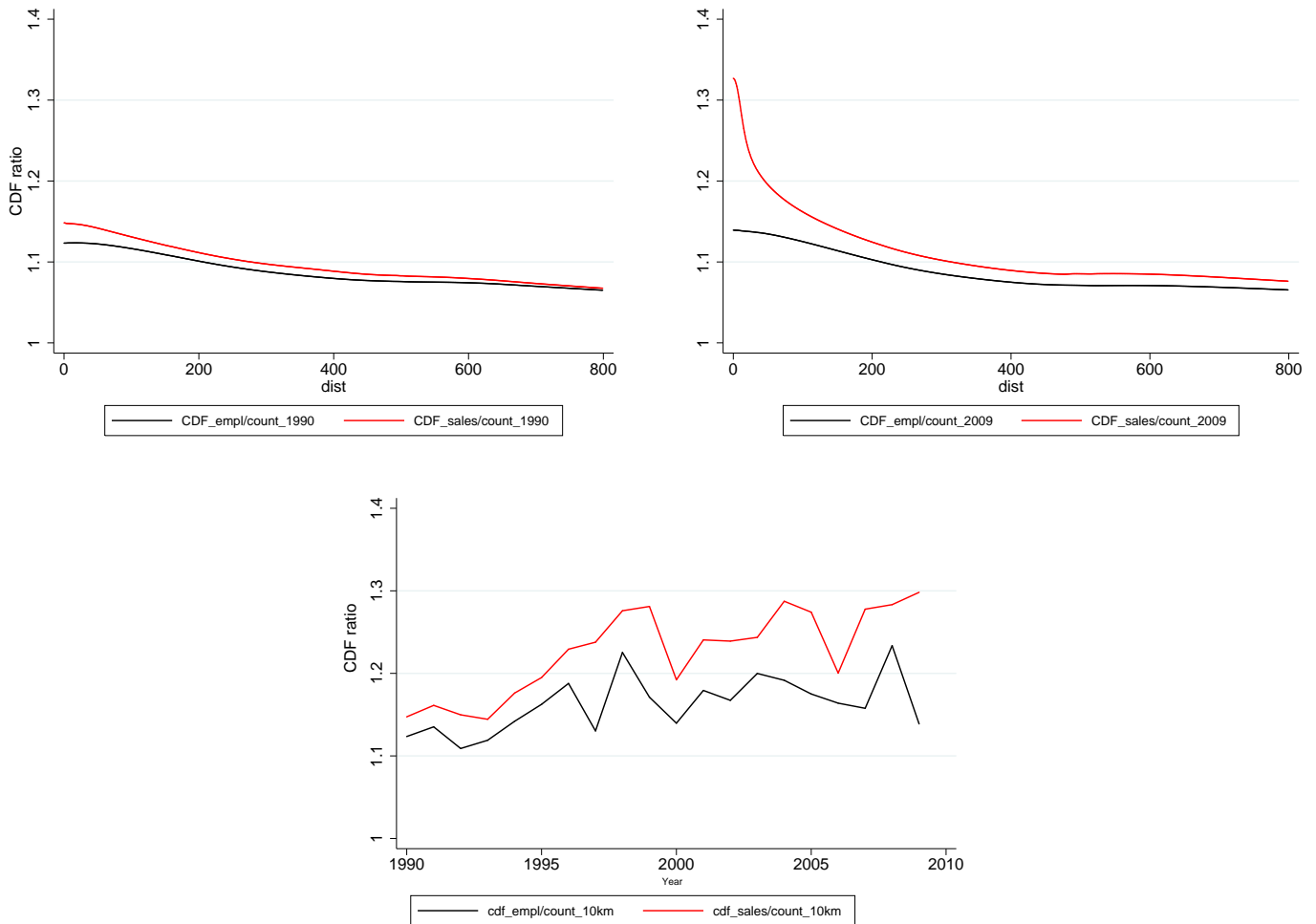
Notes: Authors' computations based on 'Annual Survey of Manufacturers' data, 1990–2009. The means are based on 257 industries and are not weighted.

of manufacturing plants. It also holds true when we consider either the employment weighted or the sales weighted measures of localization (see Table 1). Whereas the CDF of the K -density is easily interpretable and provides a natural measure to track the changing concentration of industries, it cannot tell us anything about whether or not industries are statistically significantly concentrated or not. Table 10 in the Appendix summarizes location patterns by year, based on their statistical significance. As can clearly be seen from that table, the share of statistically significantly localized industries has been decreasing over our study period, thus mimicking the downwards trend in the K -density CDFs. In a nutshell, there is a clear trend towards more dispersion, and that trend is captured by both the CDF and by the statistical tests on the localization of industries.

Whereas concentration has decreased at all distances, the greatest decline, however, was at shorter distances. Firms are dispersing, but less so at longer distances. This suggests that the incentives for firms to locate in very close proximity to each other are, in relative terms, lessening over time. Still, the fact that the CDF continues to fall at 500 km suggests a broader geographic dispersion of manufacturing plants, which is very likely driven by the rising manufacturing output in western Canada.

Comparing the unweighted and the employment or sales weighted K -densities reveals some

Figure 1: Changes in the average count-based CDF to the average employment- or sales-based CDF by distance in 1990 (left panel), in 2009 (right panel), and year-on-year changes at 10 km (bottom panel).



interesting patterns. As can be seen from the two top panels of figure 1, industries are always more concentrated (on average) in terms of employment than in terms of plant counts, and even more concentrated in terms of sales than in terms of employment. This is consistent with the findings of Holmes and Stevens (2002) and others that more localized plants tend to be larger and more productive than less localized plants. Furthermore, although industries in general have become more geographically dispersed according to all three measures, the size of plant pairs in close proximity have tended to increase in (relative) size regardless of whether size is measured in terms of employment or sales. Put differently, the process of dispersion is less pronounced when measured by either employment or sales, thus suggesting that smaller plants drive a substantial part of the dispersion process.

The bottom panel of figure 1 depicts the trend in the year-on-year evolution of the average CDF based on employment relative to the average CDF based on counts, as well as the year-on-year evolution of the average CDF based on sales relative to the average CDF based on counts. As can

be seen from that figure, there is a generalized rise in the ratios through time at 10 km distance, albeit with considerable year-on-year volatility. Between 1990 and 1995, the employment to count and sales to count ratios were – on average – 1.12 and 1.15. In comparison, between 2005 and 2009 their averages were 1.17 and 1.27, respectively.

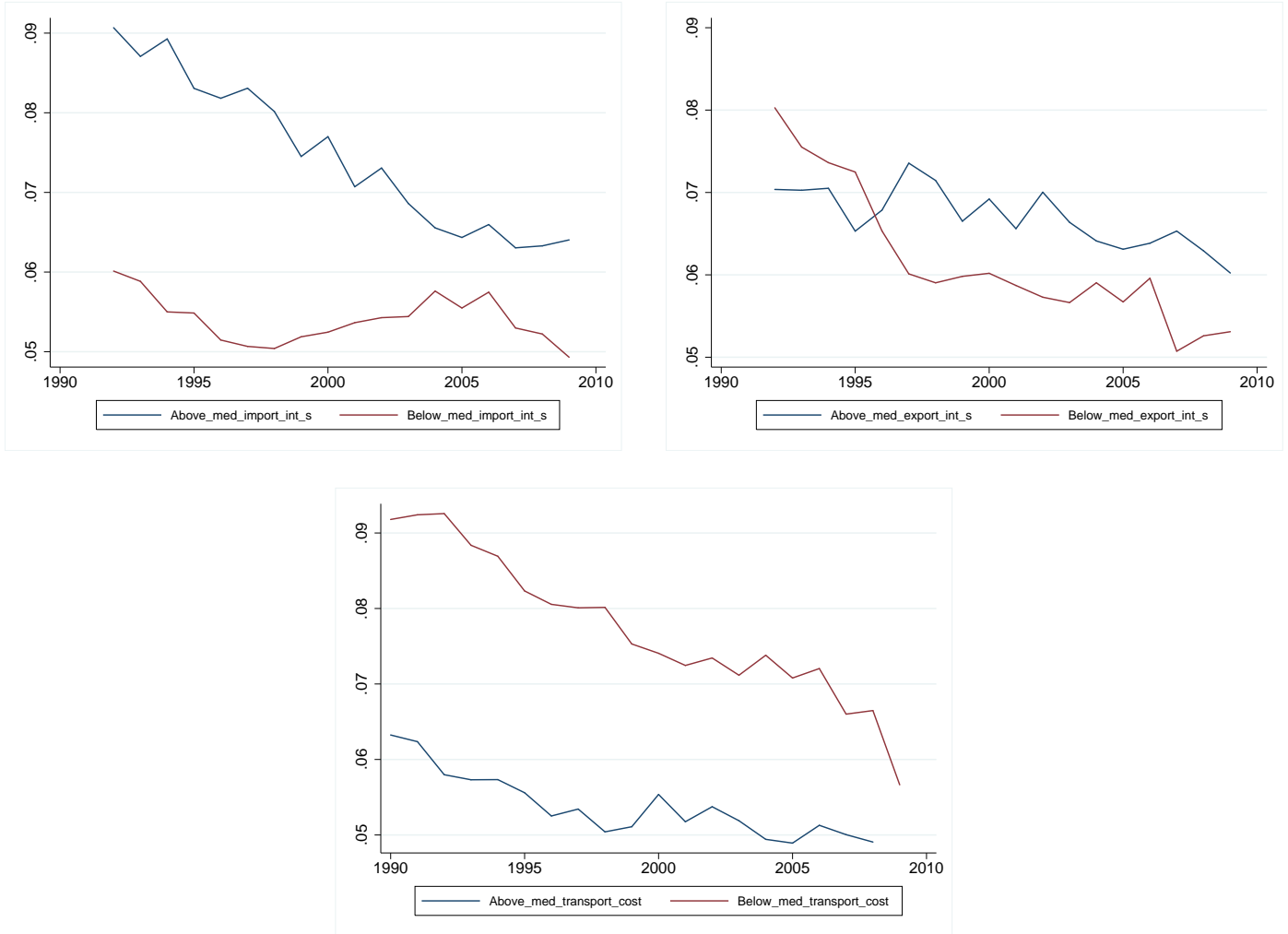
2.2 Trends by industry characteristics

The foregoing figures depict the broad trends yet mask a substantial amount of heterogeneity across industries. To account for that heterogeneity, we now break-down our sample along various dimensions to gain further insights into the broad trends. Of primary interest to us are the relationships between changes in the CDF for industries that differ along various correlates that are conjectured to influence the degree of geographical concentration of industries. The latter correlates can be divided into two broad sets.

The first set, which will be a key focus of our subsequent econometric analysis, can be subsumed under the heading of ‘trade costs’. It includes transportation costs for shipping output within Canada, measures of import and of export exposure, as well as measures of input and output linkages across firms in different industries. All of these measures play central roles in the theories of agglomeration, yet are usually not central to the analysis of the spatial concentration of industries. Falling barriers to trade enhance firms’ export opportunities, but also increase the degree of import competition. These, in turn, may have affected the localization of industry. For instance, growing import competition from low-wage countries may lead to significant exit of firms, which may influence the localization of industry (see, e.g., Dumais, Ellison, and Glaeser, 1997; Behrens, 2013). Similarly, there is a long tradition that links the concentration of industry to changes in transportation costs (notably, in the NEG models; see Krugman, 1991a,b; and Fujita, Krugman and Venables, 1999). The key insight is that high transportation costs are associated with less geographic concentration of industry, although that relationship is non-linear in models relying on urban costs. In a nutshell, we expect to see the largest impact of transport costs on the spatial concentration of industries for intermediate values of transport costs.

The two top panels of figure 2 depict the trends of the average CDF for industries with above and below median import and export intensities, respectively. Import and export intensity are measured by the industry ratio of imports or exports to sales. Import intensity captures the degree to which an industry is exposed to competition from abroad. As can be seen from figure 2, industries with above median import intensity tend to be more geographically concentrated, yet their degree of localization has fallen at a much more rapid pace than that of industries with below median levels of import exposure. This finding suggests that import competition reduces industrial localization, i.e., is a dispersion force, and that industries with high imports (i.e., which use a lot of intermediates) are geographically more concentrated (see Krugman and Livas Elizondo, 1996). Turning to exports, industries with above median export intensities are less geographically

Figure 2: Year-on-year changes of count-based CDF for industries with above and below median import intensities (left panel), export intensities (right panel), and transportation costs (bottom panel).

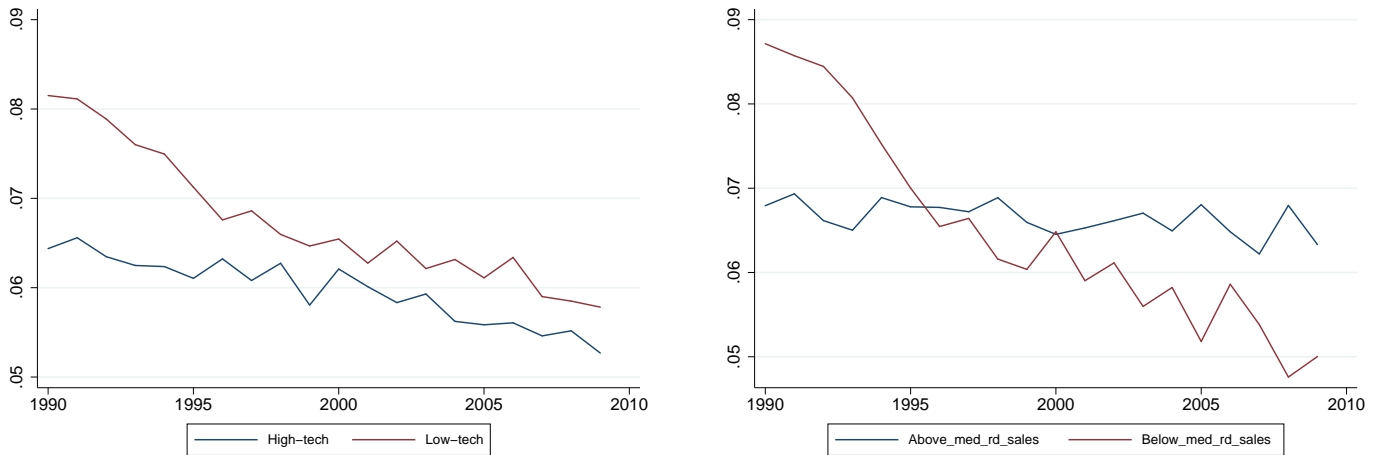


concentrated, but they tend to disperse less quickly than industries with below median export intensities. Hence, over time, industries facing less import competition and being more export oriented tend to experience a smaller decline in their degree of localization. Finally, the bottom panel of figure 2 presents the same information for industries that fall above and below the median transportation cost, where industry transportation costs are based on an ad valorem basis for a 500 km shipment (see Appendix A for a more detailed description of our transportation cost measure). As can be seen from the bottom panel, industries with transportation costs below the median are more geographically concentrated, but tend to experience a more rapid decline in their degree of localization over time.

The second set of industry characteristics we are interested in are those associated with standard agglomeration forces: natural advantage, input sharing, labour market pooling, and knowledge spillovers. We anticipate that industries that are more R&D intensive, with shorter input/output

distances, that are high-tech, and that rely on spatially concentrated natural resources will also be more localized.

Figure 3: Evolution of the industry average CDF by year for high-technology (left panel) and above and below the median R&D intensity (right panel) over the period.

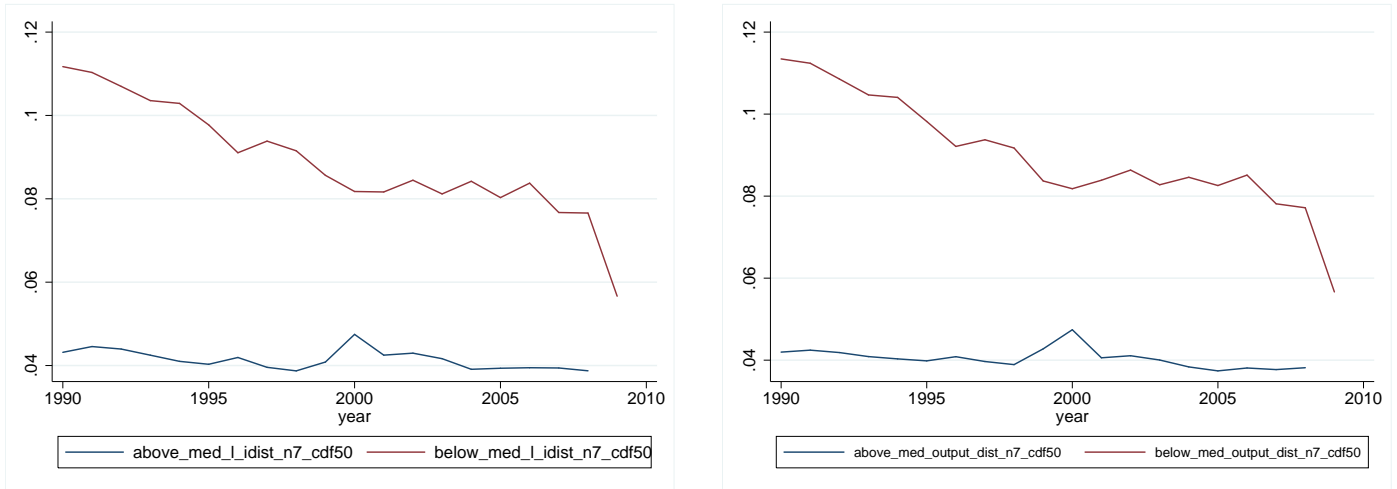


The left panel of figure 3 depicts the changes in the CDFs for industries classified as low- and as high technology.² Contrary to the standard view that high-tech industries are more geographically concentrated – because of their reliance on high-skilled labour and knowledge spillovers – we find that low-tech industries tend to be more highly localized in Canada. This finding is not that surprising given that the most localized industries in Canada tend to be in the clothing and textile sectors (see Table 9 in the Appendix; see also Behrens, 2013). As can further be seen, low-tech industries experienced a much faster decline in localization, leading to convergence through time. The right panel of figure 3 breaks down industries based on their R&D intensity. We expect a priori that industries which invest more in the generation of knowledge will also benefit more from factors such as localized knowledge spillovers. We measure R&D intensity of the industry by the ratio of industry in-house R&D to sales. At the beginning of our period, industries with below median R&D intensity were the most localized, yet this pattern has been reversed by the end of the period. Furthermore, industries with below-median levels of R&D intensity saw their degree of localization fall much more over our study period, whereas those with above-median levels of R&D intensity experienced only modest declines.

As anticipated industries with longer input or output distances were less localized, while those with shorter input or output distances were more localized. Nevertheless, industries with lower

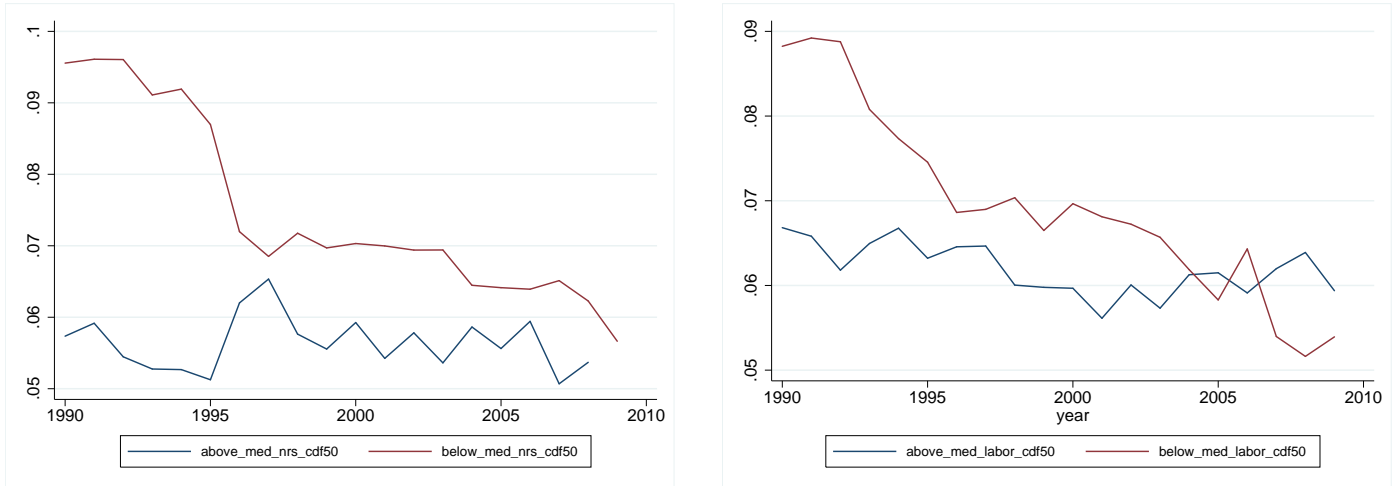
²These definitions are based on the US Bureau of Labor Statistics classification by Hecker (2005). His definition of high-tech industries is 'input based'. An industry is 'high-tech' if it employs a high proportion of scientists, engineers or technicians. As shown by Hecker (2005), these industries are also usually associated with a high R&D-to-sales ratio, and they also largely – but not always – produce goods that are classified as 'high-tech' by the Bureau of Economic Analysis.

Figure 4: Evolution of the industry average CDF by year for the above and below the median: input (left panel) and output (right panel) linkages over the period.



input/output distances experienced a greater decline in their degree of localization over the period (see figure 4).

Figure 5: Evolution of the industry average CDF by year for the above and below the median: Natural resources (left panel) and labor intensity (right panel) over the period.



The left panel of figure 5 depicts the changes in the CDFs for industries that rely on natural resources. Over our period, industries with below median natural resources intensity were the most localized and their degree of localization fell much more over time, while those with above-median levels of natural resources intensity remain stable. The right panel of figure 5 include our proxy for labor market pooling – measured by the ratio of non-production workers to production workers. At the beginning of our period, industries relying less on specialized workers i.e, below the median, were the most localized. But this pattern was reversed by the end of the period and industries

relying on less specialized workers saw their degree of localization fall much more over our study period than those relying on more specialized workers experienced.

To summarize, all evidence points to a significant decrease in the localization of manufacturing industries in Canada over the last 20 years. The pace of decline, however, differs across industries in systematic ways. High-tech industries and those that are more R&D intensive saw their degree of localization fall less. Localization also fell more for industries exposed to more import competition, and less for industries that export more intensively. Industries with lower transportation costs experienced less dispersion relative to those with high transportation costs. While these broad patterns are instructive, they are not sufficient to identify the factors that influence the localization of industry. We therefore now turn to multivariate analysis to identify the sources of changes in the geographical concentration of industries in Canada.

3 Econometric analysis

In what follows, we exploit the panel nature of our data. More precisely, we estimate the following baseline specification:

$$\gamma_{m,t} = \beta X_{m,t} + \alpha_t + \mu_m + \epsilon_{m,t} \quad (1)$$

where :

- The dependent variable $\gamma_{m,t}$ is the continuous localization measure, based on the CDF of the Duranton and Overman (2005) K -density for industry m in year t ;
- $X_{m,t}$ is a vector of industry characteristics – including industry level controls, transportation costs, and the other Marshallian determinants of agglomeration – with associated coefficient vector β ;
- α_t are time fixed effect;
- μ_m are industry fixed effect;
- $\epsilon_{m,t}$ is the error term, which is assumed to be independently and identically distributed with the usual properties for consistency of OLS.

Table 2 below summarizes our main variables and provides some descriptives. Additional information regarding data sources and construction can be found in Appendix A.

Before describing in more details our variables, it is worth discussing identification. First, it is known that some industry characteristics can affect the propensity of industries to agglomerate, whereas agglomeration can at the same time affect these industries characteristics. The three main problems for the identification of agglomeration effects are: simultaneity bias; unobserved heterogeneity and omitted variables; and multiple sources of externalities. Most studies in the literature

are based on cross-sectional data and suffer from these identification problems (see for example Rosenthal and Strange, 2001). Our approach allows us to deal with unobserved characteristics by including industry fixed effects. By including industry trade and transportation cost measures, as well as all the other Marshallian determinants and controls for natural advantage, we are also able to reduce the risk of omitted variable bias. Last, our panel data allows us to time-lag variables we suspect to be endogenous in order to reduce the risk of simultaneity bias.³ Furthermore, by construction, our measures of transportation costs and of input output linkages are ‘reasonably exogenous’ to the spatial structure of the economy, thus alleviating the problem of reverse causality.

Turning to the channels through which agglomeration effects percolate, it is known that they are all observationally equivalent. Our aim is not to disentangle them fully, but rather to investigate in more depth how important transportation- and trade-related variables are in explaining changes in the spatial patterns of the manufacturing industry in Canada.

Before proceeding, one comment is in order. Given the use of panel data, we will identify the effects of the different determinants of agglomeration on the spatial concentration of industries from year-to-year time variation. This is potentially problematic if some variables – like the degree of agglomeration or transportation costs or the educational structure of the workforce – vary only little across two successive years.

3.1 Transportation costs, international trade, and input-output linkages

Transportation costs loom large in the theoretical literature on industry location and performance. Industries with high transportation costs are expected to agglomerate production in locations close to their markets to minimize those costs. Despite their dominant theoretical role, limited work has gone thus far into the elaboration of good measures of transportation costs, and even less to their application to the analysis of the determinants of agglomeration. Rosenthal and Strange (2001) for example, use the ratio of inventories to sales at the end of the year as a proxy for ‘perishability of output’, itself a proxy for transportation costs. Our work aims to improve our understanding of how transportation costs influence the geographical concentration of industries. To this end, we use direct measures of transportation costs constructed from detailed micro-data files on shipments within Canada and between Canada and the US. Since industries tend to localize when their shipping costs are either high (market access) or low (when they are footlose, they seek to agglomerate to exploit other sources of agglomeration economies), we expect transportation costs to have a non-linear and negative effect on the degree of industrial agglomeration, especially for industries with intermediate values of transport costs. Note that, by construction, our measures of transport costs are exogenous to the location patterns of industries (see Appendix A for details). This is worth emphasizing, since using observed transportation costs may significantly downward

³We leave this for a future version of the paper.

bias the estimated coefficients if industries indeed seek to localize such to minimize those costs.

Whereas transportation costs capture the ‘domestic’ part of trade in our model, we also control finely for the role of international trade in the location of industries. It is indeed well known that trade has an impact on the spatial structure of industries (Krugman and Livas Elizondo, 1996; Brühlhart, Carrère, and Trionfetti, 2011). We use detailed yearly data on imports and exports by industry and country of origin (destination) to control for industries’ import and export exposure (the ratio of imports/exports to sales). To capture different effects, we break these measures down by origin: low-cost (Asian) countries; OECD countries; and NAFTA countries.

Finally, another important trade-related source of agglomeration are input and output linkages. Despite their importance, their empirical treatment has been rather rudimentary until now. Lu and Tao (2009), e.g., used the export-intensity of a sector as a proxy for input sharing. The rationale for this proxy is that, when compared to others industries, export industries strongly rely on inputs and information sharing like the information on procedures and international markets where they sell their products. Another approach to modelling input sharing is the one by Duranton and Overman (2005, 2008) and by Ellison, Glaeser, and Kerr (2010), who use the 1987 Benchmark Input-Output Accounts – published by the Bureau of Economic Analysis (BEA) – to measure the extent that industries buy and sell from one another. They find that this customer-supplier relationship (input sharing) is the most important mechanism to explain the colocation of industries, which is suggestive of the importance of input-output linkages for the geographical concentration of industries.

Our measures of input-output linkages are very different and make use of the micro-geographic nature of our data. Consider a firm ℓ active in sector $\Omega(\ell)$. Let Ω denote the set of *sectors* and Ω_s the set of *firms* in sector s . Let $k_s(i, \ell)$ denote the i th closest sector- s firm to the firm ℓ . Our micro-geographic measures of input- and output linkages are constructed as weighted averages as follows:

$$\mathcal{I}\text{dist}(\ell) = \sum_{s \in \Omega \setminus \Omega(\ell)} \omega_{\Omega(\ell),s}^{\text{in}} \times \frac{1}{N} \sum_{i=1}^N d(\ell, k_s(i, \ell)), \quad (2)$$

and

$$\mathcal{O}\text{dist}(\ell) = \sum_{s \in \Omega \setminus \Omega(\ell)} \omega_{\Omega(\ell),s}^{\text{out}} \times \frac{1}{N} \sum_{i=1}^N d(\ell, k_s(i, \ell)), \quad (3)$$

where $d(\cdot, \cdot)$ is the great circle distance between the firms’ postal code centroids, and where $\omega_{\Omega(\ell),s}^{\text{in}}$ and $\omega_{\Omega(\ell),s}^{\text{out}}$ are sectoral input- and output shares (see Appendix C for additional details). Since the weights sum, by construction, to one, $\mathcal{I}\text{dist}(\ell)$ can be interpreted as the average minimum distance of firm ℓ to a dollar of inputs from its N closest suppliers. Analogously, $\mathcal{O}\text{dist}(\ell)$ is the average minimum distance the firm has to ship a dollar of outputs to its N closest customers. The larger are $\mathcal{I}\text{dist}(\ell)$ or $\mathcal{O}\text{dist}(\ell)$ for firm ℓ , the worse are its input or output linkages – it is, on average, further away from a dollar of intermediate inputs or a dollar of demand emanating from the other industries.

Note that our measures are based on plant-level location information, but on national weights for input and output shares. The latter is due to the fact that we do not directly observe input-output linkages at the plant level. Yet, given this, our procedure has the advantage to sidestep problems of endogeneity of those measures — basic economic theory tells us that plants would substitute towards lower cost inputs from closer plants. In some sense, our measure is a bit reminiscent of ‘Bartik-type’ instruments.

We compute the measures (2) and (3) for all years and for all plants, averaging the linkages across the 3, 7, and 10 nearest plants in each industry. We then average them across plants in each industry and each year to get an industry-year specific measure of both input and output linkages. As expected, these measures are strongly correlated. Yet, despite that strong correlation we can include them simultaneously into our regressions and identify their effects.

3.2 Labor market pooling, knowledge spillovers, and other controls

The literature on industrial localization identifies many important sources of externalities that cause spatial concentration, such as knowledge spillovers, labor market pooling, input sharing, transportation costs, scale economies, trade and natural advantages. We will build several in-dependent variables that proxy for the above agglomeration factors in the panel econometric analysis. Marshall’s trinity (knowledge spillovers, labor market pooling, and input sharing), which are the main sources of localization economies, suggests that firms using similar technologies, inputs, and types of workers usually trying to locate near each another in order to reduce their transportation and transactions costs. Usually, except in specific cases, these externalities are not measurable directly.

The first Marshallian force we consider is the knowledge spillovers. By their nature, knowledge spillovers are very hard to measure directly. We assume that if knowledge spillovers exist between plants then they contribute to render industries more innovative and more concentrated. Information between industries have been proxied in the literature using patent citation data i.e., patents in industry i that are cited in patent of industry j . Our approach to proxy the knowledge spillovers is to used industry’s research and development (R&D) intensity, i.e., the ratio of R&D expenditure to the total output of that sector. We expect this variable to have a positive effect on agglomeration (see Table 2 for descriptive statistics).

Along with the knowledge spillover, labor market pooling is another important source of agglomeration. To identify a good proxy for labor pooling in an industry, its important to identify industry characteristics that are related to the specialization of the industry’s labor force – see Rosenthal and Strange (2001) and Lu and Tao (2009). The literature suggests that agglomeration occurs because workers are able to move across firms and industries. Firms agglomerate in order to take advantage of scale economies associated with a large labor pool that allows industries to use the same type of workers. Since it is difficult to identify these characteristics, we will employ a proxy related to the workers’ occupations by using the non-production to production worker

ratio. We, therefore, expect this proxy to have a positive effect on industrial localization. We also attempted to measure labor market pooling with the educational characteristics of workers, which are available at an aggregate industry level (L-level) from Statistics Canada KLEMS database (e.g., the ratio of hours worked by all workers with post-secondary and university degree and the labor productivity index). This measure, however, proved not to provide significant results, perhaps because it is measure at a more aggregate level than the non-production to production worker ratio, which is measured at the 6-digit NAICS level and is derived from the same database as our measure of localization. We also control for the importance of natural advantages use in the agglomeration process. This importance have been pointed out by Kim (1995) and Ellison and Glaeser (1997). We will use the share of inputs from natural resource-based industries and the sectoral energy inputs as a share of total sector output as proxies for the natural advantages. We also control for basic industry structure and scale effects by using the following proxies: total industry employment, mean plant size, Herfindahl index of plant concentration (employment based), share of plants controlled by multi-plant firms and share of plants controlled by foreign controlled plants (see Table 2).

Table 2: Description of key variables and summary statistics.

Variable name and Description	Industry detail	Mean	Standard deviation		
			Overall	Between	Within
<i>Trade, transportation, and IO variables</i>					
Share of industry imports from Asian countries (excluding OECD members)	NAICS6	0.12	0.23	0.17	0.06
Share of imports from OECD member countries (excluding U.S. and Mexico)	NAICS6	0.16	0.18	0.13	0.05
Share of imports from NAFTA countries (U.S. and Mexico)	NAICS6	0.67	0.33	0.26	0.07
Share of industry exports from Asian countries (excluding OECD members)	NAICS6	0.03	0.08	0.05	0.03
Share of exports from OECD member countries (excluding U.S. and Mexico)	NAICS6	0.09	0.13	0.08	0.05
Share of exports from NAFTA countries (U.S. and Mexico)	NAICS6	0.84	0.25	0.18	0.07
Ad valorem trucking costs for an avg. load shipped 500km as a share of goods shipped	L-level	0.03	0.03	0.03	0.01
Industry mean of the avg. dist.(km) to a dollar of inputs from the 7 nearest plants	NAICS6	608	811	267	544
Industry mean of the avg. dist.(km) to ship a dollar of output to the 7 nearest plants	NAICS6	619	861	290	571
<i>Labor market pooling, knowledge spillovers, and other controls</i>					
Share of input from natural resource-based industries	L-level	0.11	0.2	0.17	0.03
Sectoral energy inputs as a share of total sector output	L-level	0.03	0.06	0.05	0.01
Total industry employment	NAICS6	7124	9728	7859	1869
Herfindahl index of plant-level employment concentration	NAICS6	0.07	0.08	0.06	0.02
Mean plant size	NAICS6	74	181	139	42
Share of plants controlled by multi-plant firms	NAICS6	0.21	0.24	0.18	0.06
Share of foreign controlled plants	NAICS6	0.15	0.2	0.14	0.06
Share of hours worked by all persons with post-secondary education	NAICS6	0.40	0.11	0.07	0.04
Ratio of non-production workers to production workers	NAICS6	0.33	0.23	0.14	0.09
Intramural research and development expenditures as a share of industry sales	NAICS6	0.01	0.04	0.03	0.01

Notes: All descriptive statistics are based on the sample used in the regression analysis, which includes 4306 observations covering 256 industries over 18 years. Missing observations result in the reduced sample. The standard deviation is decomposed into between and within components, which measure cross-sectional and time series variation, respectively.

3.3 Estimation results

The baseline dependent variable for the main results is the cumulative sum of the K -density (CDF) at 50km. We run different regressions depending on the inclusion of industry characteristics and others controls in the model. In the first step, we run a baseline (standard) model, where we include time and industry fixed effects, basic industry structure and size characteristics – plant size, industry structure (Herfindahl index, mean plant size, share of plants controlled by multi-plant firms and share of foreign controlled plants) – and natural advantages (share of inputs from natural resource-based industries and energy inputs as a share of total output). In the second step, we run the previous model where we added transportation costs and trade variables. In a third model, we include the input-output sharing proxies – the industry mean of the average distance to a dollar of inputs (output) from (to) the seven nearest plants. In a final model we added labor market pooling proxy (ratio of non-production workers to production workers) and knowledge spillover proxy (intramural research and development expenditures as a share of industry sales). Throughout the analysis, we employ both a fixed-effects (within) and between estimators.

In choosing the fixed effects estimator we performed the Hausman test to confirm that the appropriate estimator is a fixed-effects, instead of random-effects. The result of the test strongly (1% level) indicating that the fixed-effects estimator is preferred. We report the results for the between estimator, which only take into account cross-sectional variability (based on the means of the variables over the period), because it has the advantage of capturing the high level of cross-sectional variability in some of the variables, which is not seen in the time series. For instance, there is much more variability in the R&D expenditures as a share of industry sales across industries than through time (see Table 2). We fully recognize, of course, that the between estimator does not clear out the unobserved variable bias (on characteristics fixed through time) that the fixed-effects estimator takes into account.

The two main sets of results based on the fixed-effects and between estimators are presented in Tables 3 and 4, respectively. Results in both tables are based on the unweighted (plant count) CDF at 50km as the dependent variable. Since all variables – except for trade and the share of plants controlled by multi-plant and foreign firms – enter as natural logs in the regression, their coefficients can be interpreted as elasticities. Using the full model (Table 3 , Model 4), the overall fit of the standard fixed-effects model indicate that about 32% of the time-series variation in localization can be explained by the variability of the included explanatory variables. This overall fit is higher in the between model indicating about 66% of the cross-industry variance (Table 4, Model 4).

In the baseline fixed-effects model that only includes controls, only industry employment and the share of foreign controlled firms are statistically significant; that is, industries that are growing and those with a higher share of foreign controlled firms tend to be more localized. This finding is consistent with the literature, which finds growing industries tend to become more localized (Dumais, Ellison and Glaeser 1997 and Behrens 2013) and foreign firms tend to chose to locate

within existing clusters (see Head et al. 1995 and Guimaraes et al. 2000). With the full model (model 4), the natural resource share of inputs becomes positive and significant. This is consistent with the view that firms will tend to localize in order to minimize the cost of access to natural resource-based inputs (e.g., fish processors).

The between estimator provides a wider set of significant results on the controls, but these do not hold when the trade-related and Marshallian correlates are added to the model. The exception is the Herfindahl index of plant concentration. Industries whose employment tends to be concentrated in fewer plants tend to be less localized, after taking into account average plant size.

Returning to the fixed-effects estimates, the remainder of the models progress incrementally, with the trade- and transportation cost-related variables added first (model 2), followed by measures of input sharing (model 3), and finally variables intended to capture labour market pooling and knowledge spillover effect (model 4). Across the import measures, a rising share of imports is associated with falling localization. The (non-OECD) Asian share of imports, which we use as a proxy for low-wage countries, has the highest coefficient (in absolute value) and is the most statistically significant. This is a somewhat surprising result, because we would expect the productivity enhancing effects of localization to shelter firms from low wage competition. The export share of industry output is insignificant across all country classes and models.

Regarding transportation costs, the broad expectation is that higher transportation costs will result in the dispersal of industry as firms, *ceteris paribus*, attempt to minimize these costs. That said, in practice we expect there to be an (quasi) inverted U-shape to the effect of transportation costs on location. At very low transportation costs, their effect is expected to disappear such that as they fall industries shift from concentrating to dispersing to a more random pattern of location. At the other extreme, as transportation costs rise to very high levels, we expect its marginal effect to go to zero as the industry achieves a completely dispersed location pattern. Hence, our expectation is to have no or positive effect of transportation costs at very low levels, a negative effect at moderate levels and no effect at very high levels. To accommodate these effects, we estimate the effect of transportation costs with a spline, allowing the coefficients to vary between ad valorem rates of 0 to 0.05% (low), 0.05 to 15% (moderate), and 15% or greater (high). These are admittedly arbitrary categories, but ones that we believe make intuitive sense. We also estimate the model without the spline, but these results are left to the next section. The results are, by and large, consistent with this expectations. At low levels, the effect of transportation costs is positive or insignificant. At moderate levels, the coefficient is negative and significant, and at high levels it is coefficient is negative, but not significant.

Model 3 adds the input-output sharing variables. They are found to be highly statistically significant at the 1% level of confidence with negative coefficients. While we view this result at this stage as correlative, it suggests that the geographic dispersal (or concentration) of upstream supply and downstream demand goes hand-in-hand with the falling (increasing) localization. Unlike input-output sharing, our proxies for labor market pooling and knowledge spillovers are found to

be insignificant (see model 4). We find this to be particularly surprising for the R&D share of sales, which in the descriptive analyses appeared to be closely associated with localization, or at least its maintenance. The between estimator of the same model yields different results (see Table4, Model 4). In terms of the trade variables, none of the import variables are significant, while on the export-side the OECD share of exports is negative and significant. The transportation cost variable provides results that are broadly similar to the fixed effects model and so does the input sharing variables. Labor market pooling, proxied by the nonproduction to production worker ratio (in its quadratic form), is significant in its quadratic form at the 10% level with the expected positive sign. The R&D share of sales remains insignificant.

Table 3: Fixed-effects estimates of equation (1).

Variables	Dependent variable is the CDF at 50 kilometers			
	(Model 1)	(Model 2)	(Model 3)	(Model 4)
Industry employment	0.21*** (0.0666)	0.173*** (0.0651)	0.241*** (0.0549)	0.246*** (0.0554)
Industry Herfindahl index	-0.0483 (0.0377)	-0.058 (0.0363)	-0.0413 (0.0289)	-0.0436 (0.0296)
Average plant size	-0.0203 (0.0823)	-0.0144 (0.0815)	-0.107 (0.0666)	-0.108 (0.0665)
Share of multiplant firms	-0.235 (0.174)	-0.167 (0.167)	-0.0663 (0.13)	-0.0752 (0.13)
Share of foreign controlled firms	0.524** (0.206)	0.6** (0.204)	0.422*** (0.151)	0.417*** (0.15)
Natural resource share of inputs	0.032 (0.0215)	0.0327 (0.0203)	0.0397** (0.0183)	0.0391** (0.0182)
Energy share of inputs	0.102 (0.563)	-0.161 (0.542)	-0.364 (0.45)	-0.353 (0.456)
Asian share of imports		-1.222*** (0.367)	-1.131*** (0.343)	-1.153*** (0.345)
OECD share of imports		-0.686** (0.345)	-0.508* (0.297)	-0.516* (0.299)
NAFTA share of imports		-0.792** (0.32)	-0.567** (0.287)	-0.584** (0.287)
Asian share of exports		-0.0381 (0.234)	0.0367 (0.212)	0.0384 (0.213)
OECD share of exports		0.268 (0.204)	0.289 (0.183)	0.292 (0.184)
NAFTA share of exports		0.141 (0.201)	0.144 (0.187)	0.145 (0.188)
Ad valorem trucking costs (0 to 0.05%)		0.197* (0.104)	0.00357 (0.102)	0.00786 (0.103)
Ad valorem trucking costs (0.05 to 15%)		-0.213** (0.0909)	-0.174** (0.0739)	-0.177** (0.073)
Ad valorem trucking costs (15% or greater)		-0.147 (0.363)	-0.452 (0.293)	-0.539 (0.345)
Input distance			-0.321*** (0.0452)	-0.32*** (0.0447)
Output distance			-0.137*** (0.0331)	-0.139*** (0.033)
Share of hours worked by all persons with post-secondary education				0.0518 (0.0819)
R&D as a share of sales				0.01 (0.0101)
Constant	-4.651*** (0.446)	-2.256** (0.767)	-0.795 (0.733)	-0.634 (0.745)
Year dummies	Yes	Yes	Yes	Yes
Industry dummies	Yes	Yes	Yes	Yes
Observations	4,306	4,306	4,306	4,306
R-squared	0.113	0.143	0.322	0.323
Number of industries	256	256	256	256

Notes: ***, **, * indicate significance at 1%, 5% and 10% levels, respectively. Cluster-robust standard errors are given in parentheses.

Table 4: Between estimates of equation (1).

Variables	Dependent variable is the CDF at 50 kilometers			
	(Model 1)	Model (2)	Model (3)	Model (4)
Industry employment	-0.127 (0.0798)	-0.0838 (0.0798)	0.00869 (0.0591)	0.00172 (0.0603)
Industry Herfindahl index	-0.186** (0.0843)	-0.165* (0.0839)	-0.162*** (0.0617)	-0.168*** (0.0623)
Average plant size	0.255** (0.0984)	0.184* (0.0951)	0.065 (0.0706)	0.0629 (0.0752)
Share of multiplant firms	-1.235*** (0.374)	-0.636* (0.365)	-0.222 (0.267)	-0.207 (0.27)
Share of foreign controlled firms	-0.075 (0.41)	-0.112 (0.402)	0.338 (0.294)	0.295 (0.302)
Natural resource share of inputs	-0.0964*** (0.0228)	-0.0738*** (0.0227)	0.0186 (0.0177)	0.00467 (0.0191)
Energy share of inputs	0.27 (0.404)	0.772* (0.401)	-0.204 (0.302)	-0.0135 (0.322)
Asian share of imports		1.039** (0.449)	0.442 (0.329)	0.51 (0.335)
OECD share of imports		0.714 (0.457)	0.43 (0.333)	0.49 (0.336)
NAFTA share of imports		0.543 (0.351)	0.443* (0.255)	0.471* (0.256)
Asian share of exports		-0.481 (0.93)	0.0844 (0.679)	0.153 (0.682)
OECD share of exports		-1.583*** (0.575)	-0.947** (0.421)	-0.965** (0.427)
NAFTA share of exports		-0.193 (0.413)	-0.206 (0.3)	-0.223 (0.303)
Ad valorem trucking costs (0 to 0.05%)		0.237 (0.17)	0.264** (0.124)	0.294** (0.132)
Ad valorem trucking costs (0.05 to 15%)		-0.212*** (0.0517)	-0.111*** (0.0385)	-0.106*** (0.0388)
Ad valorem trucking costs (15% or greater)		-0.723 (1.298)	(-0.856) (0.95)	-0.888 (0.96)
Input distance			-0.269** (0.114)	-0.277** (0.117)
Output distance			-0.606*** (0.106)	-0.595*** (0.108)
Non-production to production worker ratio				0.514* (0.262)
Non-production to production worker ratio squared				0.215** (0.107)
R&D as a share of sales				-0.00966 (0.0227)
Constant	-3.697*** (0.413)	-2.738*** (1.000)	3.14*** (0.837)	3.276*** (0.975)
Observations	No	No	No	No
R-squared	4306	4306	4306	4306
Number of naics	0.242	0.365	0.668	0.674
	256	256	256	256

Notes: ***, **, * indicate significance at 1%, 5% and 10% levels respectively. Standards errors are in parentheses.

4 Robustness checks

We now provide some evidence on the robustness of our key findings.

4.1 Results by distance and weighting

In this section, we run the fixed-effects and the between-effects estimations of model 4 with different CDF measures as a dependent variable: unweighted and weighted (by sales and by employment) at 50km, 200km and 500km. As can be seen from Tables 5 through 7 below, the results are robust to these different weighting schemes. Continuing with the unweighted results (Table 5), there is a general tendency for the values and significance of variables to attenuate as the CDF increases in distance. The exception is the transport cost variables whose coefficient does not fall and in fact rises in significance, which is consistent with firms economising on access to output markets, which will tend to be more geographically dispersed. The employment weighted and sales weighted measures provide broadly similar results (see tables 6 and 7). The effect of import competition, however, tends to be more limited to import competition from Asia and its coefficient tends to be smaller. This suggests much of the adaptation of import competition, particularly from low wage countries which are responsible for the bulk of exit in Canadian manufacturing (Behrens, 2013), is likely occurring in smaller plants/firms. The transportation cost variable tends to be weaker in the weighted models, but as will be seen in the next section, if we use a simpler form the negative effect of transportation cost on localization remains, regardless of the weighting scheme.

4.2 Transportation costs

The final robustness check is on the transportation cost variable. Rather than entering it into model 4 using spline. We also estimate the model without allowing its effect to vary by its level. Table 8 presents the coefficients and level of significance on ad valorem transportation costs by distance and CDF weighting. Focussing on the fixed effects results, the general pattern is for the effect of transportation costs to be negative and significant with the influence and statistical significance of transportation costs increasing as the distance threshold of the CDF increases, particularly for the employment and sales weighted cases. The between estimator yields generally insignificant results, with the exception of the unweighted measure of localization.

5 Concluding remarks

In this paper, we have revisited the seminal contribution by Rosenthal and Strange (2001) on the determinants of agglomeration. To this end, and to improve on previous work that has essentially looked at cross-sectional data and used the Ellison and Glaeser (1997) measure of localization, we have used a long panel of Canadian data and continuous measures of localization following

Table 5: Fixed-effects and between estimates of equation (1).

Variables	Fixed-effects estimator of model (4)			Between estimator of model (4)		
	Dependent variable			Dependent variable		
	CDF 50km	CDF 200km	CDF 500km	CDF 50km	CDF 200km	CDF 500km
Industry employment	0.246*** (0.0554)	0.151*** (0.0485)	0.0488 (0.0399)	0.00172 (0.0603)	-0.0189 (0.0498)	-0.0328 (0.0378)
Herfindahl index	-0.0436 (0.0296)	-0.0418* (0.0235)	-0.0265 (0.0196)	-0.168** (0.0623)	-0.127** (0.0515)	-0.0757* (0.0391)
Mean industry employment	-0.108 (0.0665)	-0.038 (0.0575)	0.028 (0.0471)	0.0629 (0.0752)	0.0774 (0.062)	0.0806* (0.0471)
Share of multinational firms	-0.0752 (0.13)	-0.148 (0.112)	-0.188** (0.087(9))	-0.207 (0.27)	-0.21 (0.223)	-0.101 (0.169)
Share of foreign controlled firms	0.417** (0.15)	0.38*** (0.114)	0.311*** (0.0921)	0.295 (0.302)	0.227 (0.25)	0.101 (0.19)
Natural resource share of inputs	0.0391* (0.0182)	0.03* (0.0135)	0.0165 (0.0103)	0.00467 (0.0191)	0.00109 (0.0158)	0.00582 (0.012)
Energy share of inputs	-0.353 (0.456)	-0.316 (0.408)	-0.234 (0.327)	-0.0135 (0.322)	0.0192 (0.266)	-0.0565 (0.202)
Asian share of imports	-1.153*** (0.345)	-0.725*** (0.213)	-0.404** (0.188)	0.51 (0.335)	0.482* (0.276)	0.391* (0.21)
OECD share of imports	-0.516* (0.299)	-0.262 (0.224)	-0.0932 (0.227)	0.49 (0.336)	0.537* (0.277)	0.446** (0.211)
NAFTA share of imports	-0.584* (0.287)	-0.315 (0.198)	-0.126 (0.18(7))	0.471* (0.256)	0.48** (0.212)	0.397** (0.161)
Asian share of exports	0.0384 (0.213)	0.0375 (0.181)	0.0158 (0.121)	0.153 (0.682)	-0.0426 (0.563)	-0.21 (0.428)
OECD share of exports	0.292 (0.184)	0.281* (0.154)	0.241** (0.113)	-0.965* (0.427)	-0.929*** (0.352)	-0.701*** (0.267)
NAFTA share of exports	0.145 (0.188)	0.167 (0.154)	0.113 (0.104)	-0.223 (0.303)	-0.258 (0.25)	-0.25 (0.19)
Ad valorem trucking costs (0 to 0.05%)	0.00786 (0.103)	-0.0292 (0.0925)	-0.0583 (0.0672)	0.294* (0.132)	0.225** (0.109)	0.149* (0.083)
Ad valorem trucking costs (0.05 to 15%)	-0.177** (0.073)	-0.172*** (0.0648)	-0.177*** (0.0513)	-0.106*** (0.0388)	-0.0778** (0.0321)	-0.0463* (0.0244)
Ad valorem trucking costs (15% or greater)	-0.539 (0.345)	-0.405 (0.29)	-0.293 (0.211)	-0.888 (0.96)	-0.866 (0.793)	-0.726 (0.602)
Input distance	-0.32*** (0.0447)	-0.262*** (0.0363)	-0.209*** (0.0283)	-0.277** (0.117)	-0.204** (0.0967)	-0.139* (0.0734)
Output distance	-0.139*** (0.033)	-0.115*** (0.0273)	-0.0786*** (0.0206)	-0.595*** (0.108)	-0.489*** (0.0895)	-0.361*** (0.068)
Non-production-production worker ratio	0.0518 (0.0819)	0.0606 (0.0635)	0.0439 (0.0465)	0.514* (0.262)	0.457** (0.216)	0.306* (0.164)
Non-production-production worker ratio squared	0.0193 (0.0322)	0.0199 (0.0255)	0.0144 (0.0191)	0.215** (0.107)	0.204** (0.0887)	0.145** (0.0674)
R&D share of sales	0.01 (0.0101)	0.013 (0.00796)	0.0119* (0.00634)	-0.00966 (0.0227)	0.00118 (0.0188)	0.012 (0.0143)
Constant	-0.634 (0.745)	0.155 (0.6460)	0.545 (0.492)	3.276*** (0.975)	3.376*** (0.805)	2.872*** (0.611)
Year dummies	Yes	Yes	Yes	No	No	No
Industry dummies	Yes	Yes	Yes	No	No	No
Observations	4,306	4,306	4,306	4306	4,306	4,306
R-squared	0.323	0.328	0.335	0.674	0.661	0.629
Number of industries	256	256	256	256	256	256

Notes: ***, **, * indicate significance at 1%, 5% and 10% levels respectively. Cluster-robust standard errors in parentheses for the fixed-effects estimators, normal standard errors for the between estimator.

Table 6: Fixed-effects and between estimates of equation (1).

Variables	Fixed-effects estimator of model (4)			Between estimator of model (4)		
	Dependent variable (empl. weighted)			Dependent variable (empl. weighted)		
	CDF 50km	CDF 200km	CDF 500km	CDF 50km	CDF 200km	CDF 500km
Industry employment	0.225*** (0.0548)	0.146*** (0.0467)	0.0589 (0.0371)	0.101 (0.0692)	0.0597 (0.0559)	0.0269 (0.0415)
Herfindahl index	-0.0463 (0.0349)	-0.0373 (0.0306)	-0.00696 (0.0247)	-0.0571 (0.0715)	-0.0363 (0.0577)	-0.004 (0.0428)
Mean industry employment	-0.0678 (0.0673)	-0.0173 (0.0589)	0.0223 (0.0468)	-0.04 (0.0861)	-0.00832 (0.0696)	0.00899 (0.0516)
Share of multinational firms	-0.106 (0.14)	-0.186 (0.119)	-0.214** (0.0903)	-0.347 (0.309)	-0.337 (0.25)	-0.199 (0.185)
Share of foreign controlled firms	0.284** (0.136)	0.318*** (0.117)	0.305*** (0.0929)	0.71** (0.346)	0.624** (0.28)	0.426** (0.208)
Natural resource share of inputs	0.0262* (0.0152)	0.0243** (0.0122)	0.0157* (0.0093)	0.00453 (0.0219)	-0.00248 (0.0177)	0.00295 (0.0131)
Energy share of inputs	-0.363 (0.54)	-0.382 (0.482)	-0.347 (0.378)	-0.226 (0.369)	-0.158 (0.298)	-0.177 (0.221)
Asian share of imports	-0.544** (.256)	-0.43** (0.217)	-0.289 (0.192)	0.413 (0.384)	0.401 (0.31)	0.305 (0.23)
OECD share of imports	-0.27 (0.229)	-0.162 (0.225)	-0.0666 (0.223)	0.527 (0.385)	0.517* (0.311)	0.399* (0.231)
NAFTA share of imports	-0.225 (0.214)	-0.155 (0.197)	-0.0764 (0.186)	0.504* (0.294)	0.463* (0.237)	0.352** (0.176)
Asian share of exports	-0.0946 (0.248)	-0.0298 (0.191)	-0.0129 (0.136)	-0.167 (0.782)	-0.192 (0.632)	-0.201 (0.469)
OECD share of exports	0.0967 (0.188)	0.147 (0.139)	0.133 (0.101)	-0.973** (0.489)	-0.979** (0.395)	-0.747** (0.293)
NAFTA share of exports	0.0424 (0.191)	0.098 (0.13)	0.0623 (0.0864)	-0.222 (0.347)	-0.224 (0.281)	-0.197 (0.208)
Ad valorem trucking costs (0 to 0.05%)	0.0482 (0.0717)	0.0163 (0.0594)	-0.0718 (0.0473)	0.301** (0.152)	0.223* (0.123)	0.124 (0.0909)
Ad valorem trucking costs (0.05 to 15%)	-0.128 (0.0799)	-0.128* (0.0716)	-0.134** (0.0598)	-0.084* (0.0445)	-0.0591 (0.036)	-0.0346 (0.0267)
Ad valorem trucking costs (15% or greater)	-0.177 (0.271)	-0.116 (0.233)	-0.0599 (0.176)	-0.785 (1.1)	-0.785 (0.889)	-0.668 (0.66)
Input distance	-0.224*** (0.0482)	-0.189*** (0.0393)	-0.163*** (0.0282)	-0.253* (0.134)	-0.186* (0.108)	-0.12 (0.0805)
Output distance	-0.115*** (0.0344)	-0.0875*** (0.0278)	-0.0526*** (0.0184)	-0.673*** (0.124)	-0.52*** (0.1)	-0.384*** (0.0745)
Non-production-production worker ratio	-0.171 (0.181)	-0.08 (0.147)	-0.0208 (0.119)	0.442 (0.3)	0.433* (0.242)	0.3* (0.18)
Non-production-production worker ratio squared	-0.0607 (0.0625)	-0.0317 (0.0504)	-0.00869 (0.0405)	0.193 (0.123)	0.206** (0.0995)	0.157** (0.0739)
R&D share of sales	0.00773 (0.00992)	0.00924 (0.00772)	0.00827 (0.00574)	7.83E-05 (0.0261)	0.0114 (0.0211)	0.0198 (0.0156)
Constant	-1.173* (0.5979)	-0.091 (0.4959)	0.295 (0.3972)	3.781*** (1.118)	3.615*** (0.903)	2.912*** (0.67)
Year dummies	Yes	Yes	Yes	No	No	No
Industry dummies	Yes	Yes	Yes	No	No	No
Observations	4,306	4,306	4,306	4,306	4,306	4,306
R-squared	0.216	0.227	0.242	0.633	0.623	0.601
Number of industries	256	256	256	256	256	256

Notes: ***, **, * indicate significance at 1%, 5% and 10% levels respectively. Cluster-robust standard errors in parentheses for the fixed-effects estimators, normal standard errors for the between estimator.

Table 7: Fixed-effects and between estimates of equation (1).

Variables	fixed-effects estimator of model (4)			Between estimator of model (4)		
	Dependent variable (sales weighted)			Dependent variable (sales weighted)		
	CDF 50km	CDF 200km	CDF 500km	CDF 50km	CDF 200km	CDF 500km
Industry employment	0.251***	0.17***	0.0763**	0.0818	0.0489	0.0234
	0.0554	(0.0468)	(0.0368)	(0.0736)	(0.0579)	(0.0425)
Herfindahl index	-0.0235	-0.0174	0.00927	-0.0752	-0.0463	-0.00654
	0.0368	(0.0316)	(0.0249)	(0.076)	(0.0598)	(0.0439)
Mean industry employment	-0.0929	-0.0382	0.00632	-0.0165	0.00864	0.0193
	0.0666	(0.0586)	(0.0469)	(0.0916)	(0.0721)	(0.053)
Share of multinational firms	-0.0956	-0.187	-0.225**	-0.391	-0.364	-0.214
	0.14	(0.119)	(0.0888)	(0.329)	(0.259)	(0.19)
Share of foreign controlled firms	0.283*	0.309**	0.275***	0.615*	0.553*	0.379*
	0.146	(0.124)	(0.0951)	(0.369)	(0.29)	(0.213)
Natural resource share of inputs	0.0157	0.0161	0.00829	0.0103	0.00113	0.00616
	0.0157	(0.0127)	(0.00936)	(0.0233)	(0.0183)	(0.0135)
Energy share of inputs	-0.291	-0.32	-0.327	-0.353	-0.272	-0.275
	0.586	(0.523)	(0.393)	(0.392)	(0.308)	(0.227)
Asian share of imports	-0.679**	-0.584**	-0.395*	0.413	0.412	0.304
	0.291	(0.237)	(0.205)	(0.408)	(0.321)	(0.236)
OECD share of imports	-0.239	-0.166	-0.0588	0.476	0.504	0.398*
	0.258	(0.244)	(0.231)	(0.409)	(0.322)	(0.237)
NAFTA share of imports	-0.228	-0.189	-0.0868	0.48	0.461*	0.36**
	0.251	(0.223)	(0.2)	(0.312)	(0.246)	(0.181)
Asian share of exports	-0.15	-0.0318	0.03	-0.38	-0.408	-0.409
	0.247	(0.177)	(0.104)	(0.832)	(0.655)	(0.481)
OECD share of exports	0.2	0.251*	0.24**	-0.911*	-0.944**	-0.727**
	0.183	(0.139)	(0.0979)	(0.52)	(0.409)	(0.301)
NAFTA share of exports	0.0921	0.159	0.144*	-0.171	-0.204	-0.188
	0.189	(0.134)	(0.0852)	(0.369)	(0.291)	(0.213)
Ad valorem trucking costs (0 to 0.05%)	0.0669	0.037	-0.0866*	0.283*	0.224*	0.127
	0.067	(0.0557)	(0.0465)	(0.161)	(0.127)	(0.0932)
Ad valorem trucking costs (0.05 to 15%)	-0.0889	-0.0904	-0.103**	-0.0878*	-0.0543	-0.0294
	0.0742	(0.0642)	(0.0523)	(0.0474)	(0.0373)	(0.0274)
Ad valorem trucking costs (15% or greater)	-0.389	-0.332	-0.218	-0.973	-0.992	-0.841
	0.272	(0.231)	(0.179)	(1.171)	(0.921)	(0.677)
Input distance	-0.219***	-0.177***	-0.154***	-0.284**	-0.218*	-0.142*
	0.0479	(0.0393)	(0.0283)	(0.143)	(0.112)	(0.0825)
Output distance	-0.106***	-0.083***	-0.0517***	-0.65***	-0.487***	-0.363***
	0.0352	(0.0282)	(0.0185)	(0.132)	(0.104)	(0.0764)
Non-production-production worker ratio	-0.00764	0.0673	0.0903	0.445	0.431*	0.292
	0.103	(0.0778)	(0.0703)	(0.319)	(0.251)	(0.184)
Non-production-production worker ratio squared	-0.015	0.0115	0.0245	0.178	0.192*	0.146*
	0.0363	(0.0275)	(0.0242)	(0.131)	(0.103)	(0.0757)
R&D share of sales	0.00745	0.00881	0.00823	-0.00499	0.0111	0.0228
	0.0101	(0.00808)	(0.00604)	(0.0277)	(0.0218)	(0.016)
Constant	-1.627***	-0.494	-0.22	3.846***	3.714***	3.045***
	0.601	(0.497)	(0.39)	(1.189)	(0.935)	(0.687)
Year dummies	Yes	Yes	Yes	No	No	No
Industry dummies	Yes	Yes	Yes	No	No	No
Observations	4,306	4,306	4,306	4,306	4,306	4,306
R-squared	0.209	0.223	0.239	0.615	0.611	0.596
Number of industries	256	256	256	256	256	256

Notes: ***, **, * indicate significance at 1%, 5% and 10% levels respectively. Cluster-robust standard errors in parentheses for the fixed-effects estimators, normal standard errors for the between estimator.

Table 8: Coefficient estimates for ad valorem transportation costs.

Distance	Weighting		
	Unweighted	Employment	Sales
	<i>Fixed effects</i>		
CDF 50km	-0.130*	-0.077	-0.049
CDF 200km	-0.134**	-0.085	-0.058
CDF 500km	-0.144***	-0.113**	-0.100**
	<i>Between</i>		
CDF 50km	-0.073**	-0.052	-0.058
CDF 200km	-0.054*	-0.037	-0.034
CDF 500km	-0.032	-0.024	-0.02

Notes: ***, **, * indicate significance at the 1%, 5% and 10% levels, respectively. Cluster-robust standard errors are used for the fixed effects estimator and normal standard errors for the between estimator.

Duranton and Overman (2005). Doing so allows us to control for unobserved heterogeneity, and to identify changes in the degree of localization from time changes in the different covariates. We have paid particular attention to the role of transportation costs – constructed using extensive Canadian trucking microdata – international trade exposure, and input sharing – constructed using micro-geographic location patterns of plants. The former two have traditionally received little attention in the empirical literature, despite their important theoretical role. Our results generally point to the importance of transportation and trade in explaining changes in the location patterns of industries.

We find that between 1990 and 2009, industry localization has persistently fallen. The average degree of localization decreased by 36% within 10km, by 22.6% within 100km, and by 11.3% within 500km. Declining localization is associated with import competition, particularly from low wage countries, rising transportation costs and the spreading out of upstream input suppliers and downstream demand for intermediate inputs. While we find strong evidence of trade-driven changes in localization, we find less evidence for knowledge spillovers and labour market pooling as drivers in changes in localization.

Acknowledgements. We thank John Baldwin for very helpful comments and suggestions. financial support from Statistics Canada’s ‘Trade Cost Analytical Projects Initiative’ (API), and from the CRC Program of the Social Sciences and Humanities Research Council (SSHRC) of Canada for the funding of the *Canada Research Chair in Regional Impacts of Globalization* is gratefully acknowledged. Bougna gratefully acknowledges financial support from CIRPÉE. This work was carried out while Behrens and Bougna were deemed employees of Statistics Canada. The views expressed in this paper, as well as all remaining errors, are ours.

References

- [1] Baldwin, John R., Brown, W. Mark, and David L. Rigby (2010) Agglomeration Economies: Microdata Panel Estimates From Canadian Manufacturing, *Journal of Regional Science* 50(5), 915–934.
- [2] Baldwin, John R., Beckstead, D., Brown, W. Mark, and David L. Rigby (2008) Agglomeration and the Geography of Localization Economies in Canada, *Regional Studies* 42(1), 117–132.
- [3] Behrens, Kristian (2013) Strength in Numbers? The Weak Effect of Manufacturing Clusters on Canadian Productivity. CD Howe Commentary #377, Toronto.
- [4] Behrens, Kristian (2012) Unweaving Clusters: Manufacturing Localization and International Trade in Canada, 2001–2009. *In progress*, Univ. of Quebec at Montreal.
- [5] Behrens, Kristian, and Bougna Théophile (2013) An Anatomy of the Geographical Concentration of Canadian Manufacturing Industries. CIRPÉE DISCUSSION PAPER #13-27, Univ. of Quebec at Montreal.
- [6] Brühlhart, Marius, Carrre, Cline, and Trionfetti, Federico (2012) How wages and employment adjust to trade liberalization: Quasi-experimental evidence from Austria, *Journal of International Economics* 86(2012), 68–81.
- [7] Devereux, Micheal P., Rachel Griffith, and Helen Simpson (2004) The geographic distribution of production activity in the UK, *Regional Science and Urban Economics* 34(5), 533–564.
- [8] Dumais, Guy, Glenn D. Ellison, and Edward L. Glaeser (2002) Geographic Concentration as a Dynamic Process, *Review of Economics and Statistics* 84(2), 193–204.
- [9] Duranton, Gilles, and Henry G. Overman (2005) Testing for Localisation Using Micro-Geographic Data, *Review of Economic Studies* 72(4), 1077–1106.
- [10] Duranton, Gilles, and Diego Puga (2004) Micro-foundations of urban agglomeration economies. In: J. Vernon Henderson, and Jacques-François Thisse (eds.), *Handbook of Regional and Urban Economics*, vol. 4. North-Holland: Elsevier B.V., pp. 2063–2117.
- [11] Ellison, Glenn D., and Edward L. Glaeser (1997) Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach, *Journal of Political Economy* 105(5), 889–927.
- [12] Ellison, Glenn D., and Edward L. Glaeser (1999) The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?, *The American Economic Review* 89(2), 311–316.

- [13] Ellison, Glenn D., Edward L. Glaeser, and William R. Kerr (2010) What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns, *American Economic Review* 100(3), 1195–1213.
- [14] Fujita, Masahisa, Krugman, Paul R., and Venables J. Anthony (1999) Spatial Economy, cities, Regions and International Trade, *Journal of International Economics* 57(2002), 247–251.
- [15] Guimaraes , Paulo, Octávio Figueiredo, Douglas Woodward (2000) Agglomeration and the Location of Foreign Direct Investment in Portugal, *Journal of Urban Economics* 47(1), 115–135.
- [16] Head, Keith, John Ries, and Deborrah Swenson (1995) Agglomeration Benefits and Location Choice: Evidence from Japanese Manufacturing Investments in the United States, *Journal of International Economics* 38(3-4), 223–247.
- [17] Holmes, Thomas J. and John J. Stevens (2002) Geographic concentration and establishment scale, *Review of Economics and Statistics*. 84(4), 682–690.
- [18] Holmes, Thomas J. and John J. Stevens (2004) Spatial Distribution of Economic Activities in North America. In: J. Vernon Henderson and Jacques-François Thisse (eds.), *Handbook of Regional and Urban Economics*, vol. 4. North-Holland: Elsevier B.V., pp. 2797–2843.
- [19] Kim, Sukkoo (1999) Regions, Resources, and Economic Geography: Sources of U.S. regional comparative advantage, 1880–1987, *Regional Science and Urban Economics* 29(1), 1–32.
- [20] Krugman, Paul R. (1991) Increasing Returns and Economic Geography, *Journal of Political Economy* 99(3), 483–499.
- [21] Krugman, Paul R. (1991) *Geography and Trade*, MIT press, Cambridge
- [22] Krugman, Paul R. and Maurice Obstfeld (1994) *International Economics: Theory and Practice, Third Edition*. New York: HarperCollins College Publishing.
- [23] Krugman, Paul R. and Elizondo, Raul L. (1996) Trade Policy and the Third World Metropolis, *Journal of Development Economics* 49(1), 137–150.
- [24] Lu, Jiangyong and Zhigang Tao (2009) Trends and determinants of China’s industrial agglomeration, *Journal of Urban Economics* 65(2) 167–180.
- [25] Rigby, David L. and J
”urgen Essletzbichler (2002) Agglomeration Economies and Productivity Differences in US Cities, *Journal of Economic Geography* 2(4), 407–432.

- [26] Rosenthal, Stuart S. and William C. Strange (2004) Evidence on the Nature and Sources of Agglomeration Economies. In: J. Vernon Henderson, and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, vol. 4. North-Holland: Elsevier B.V., pp. 2119–2171.
- [27] Rosenthal, Stuart S. and William C. Strange (2001) The Determinants of Agglomeration, *Journal of Urban Economics* 50(2) 191–229.

Appendix

This Appendix is structured as follows. Appendix A describes our datasets and sources. Appendix B provides details on the Duranton-Overman K -density computations. Appendix C describes the construction of our input-output measures. Last, Appendix D contains additional tables and results.

A. Data and data sources

This appendix provides details on the data used in this paper and the sources.

Firm-level data and industries. Our analysis is based on the *Annual Survey of Manufacturers (ASM) Longitudinal Microdata file*. We also use the ASM to construct controls for the labor market pooling variables, and natural advantage proxies. This data cover the years from 1990 to 2010 and our focus is on manufacturing firms only. For every establishment, we have information on its primary 6-digit NAICS code and up to four secondary 6-digit NAICS codes; the year of establishment; its employment; whether or not it is an exporter; and its 6-digit postal code. The latter allows us to effectively geo-locate the firms. We also used others database as:

The confidential L-level input-output tables. We use these table to construct our plant-level proxies for the importance of input and output linkages (see Appendix C for more details).

KLEMS database. This database also contains various industry-level informations useful for constructing proxies for natural advantage (e.g., energy intensity, water usage etc.).

Estimation of ad valorem transportation rates: To estimate ad valorem rates we first use a pricing model to predicted trucking firm revenues for a 500km trip by commodity for the average tonnage using shipment (waybill) data from Statistics Canadas Trucking Commodity Origin-Destination Survey (see Anderson and Brown 2013). These revenues are then converted into ad valorem trucking costs, by estimating the value of each shipment. This value is derived by multiplying the tonnage of the average shipment on a commodity basis by their respective value per

tonne derived from an experiment export trade file produced only in 2008. These ad valorem estimates at the commodity level, in turn, are used to estimate ad valorem rate for L-level industries in 2008 using a set of industry-commodity concordances. Trucking industry price indices and manufacturing industry price indices from Statistics Canadas KLEMS database are then used to project the ad valorem rates backwards and forwards in time creating an industry-specific ad valorem transportation rate time series invariant to the spatial structure of the industry.

Geographical data. To geolocate firms, we used latitude and longitude data of postal code centroids obtained from Statistics Canada’s Postal Code Conversion files (PCCF). These files associate each postal code with different Standard Geographical Classifications (SGC) that are used for reporting census data. We match firm-level postal code information with geographical coordinates from the PCCF, using the postal code data for the next year in order to consider the fact that there is a six months delay in the updating of postal codes.

B. The distance-Based Duranton-Overman approach

Following Duranton and Overman (2005, 2008), hereafter DO, the estimator of the kernel density (probability density function or PDF) at a given distance d of bilateral distances between plants is given by:

$$\hat{K}(d) = \frac{1}{n(n-1)h} \sum_{i=1}^{n-1} \sum_{j=i+1}^n f\left(\frac{d-d_{ij}}{h}\right), \quad (\text{B.1})$$

where h is the optimal bandwidth, and f a Gaussian kernel function. The distance d_{ij} (in kilometers) between firms i and j is computed as:

$$d_{ij} = 6378.39 \cdot \text{acos}[\cos(|\text{lon}_i - \text{lon}_j|) \cos(\text{lat}_i) \cos(\text{lat}_j) + \sin(\text{lat}_i) \sin(\text{lat}_j)]. \quad (\text{B.2})$$

Alternatively, rather than using plant counts as the unit of observation in (B.1), we can characterize the localization of employment or sales at the industry level. This can be accommodated by adding weights to (B.1) :

$$\hat{K}_W(d) = \frac{1}{h \sum_{i=1}^{n-1} \sum_{j=i+1}^n (e_i + e_j)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n (e_i + e_j) f\left(\frac{d-d_{ij}}{h}\right), \quad (\text{B.3})$$

where e_i and e_j are the employment or sales levels of establishment i and j , respectively.⁴ The weighted K -density thus describes the distribution of bilateral distances between employees – or

⁴Contrary to Duranton and Overman (2005), who use a multiplicative weighting scheme, we use an additive one. The additive scheme gives less weight to pairs of large establishments and more weight to pairs of smaller establishments than the multiplicative scheme does. Using a multiplicative scheme would imply that our results may be too strongly driven by a few very large firms in a given industry.

sales – in a given industry, whereas the unweighted K -density describes the distribution of bilateral distances between plants in that industry.

Let us briefly explain how we compute these continuous measure of localization. Using the overall sample of manufacturing plants located in Canada, we randomly draw as many locations as there are plants in industry A . To each of these locations, we assign randomly a plant from industry A , using its observed employment and its observed sales. This procedure ensures that we control for the overall pattern of concentration in the manufacturing sector as a whole, as well as for the within-industry concentration. We then compute the bilateral distances of this hypothetical industry and estimate the K -density of the bilateral distances. finally, for each industry A . We repeat this procedure 1,000 times to obtain a set of 1,000 estimated values of the K -density at each distance d . To assess whether an industry is significantly localized or dispersed, we compare the actual K -density with that of the counterfactual distribution. We consider a range of distances between zero and 800 kilometers.⁵ We then use our bootstrap distribution of K -densities, generated by the counterfactuals, to construct a two-sided confidence interval that contains 90 percent of these estimated values. The upper bound, $\overline{K}(d)$, of this interval is given by the 95th percentile of the generated values, and the lower bounds, $\underline{K}(d)$, by the 5th percentile of these values. Distributions of observed distances that fall into this confidence band could be ‘as good as random’ and are, therefore, not considered to be either localized or dispersed. The bootstrap procedure generates a confidence band, and any deviation from that band indicates localization or dispersion of the industry. If $\widehat{K}(d) > \overline{K}(d)$ for at least one $d \in [0, 800]$, whereas it never lies below $\underline{K}(d)$ for all $d \in [0, 800]$, industry A is defined as globally localized at the 5 percent confidence level. On the other hand, if $\widehat{K}(d) < \underline{K}(d)$ for at least one $d \in [0, 800]$, industry A is defined as globally dispersed.

While the K -density PDF provides a clear picture of localization at every distance d , and while it allows for statistical testing, it is not suited in capturing globally the location patterns of industries up to some distance d . This can, however, be achieved by using the K -density cumulative distribution up to distance d . In all our econometric estimations, we use as dependent variable the CDF of the K -densities. Those are given by:

$$\text{CDF}(d) = \sum_{\delta=1}^d \widehat{K}(\delta) \quad \text{and} \quad \text{CDF}_W(d) = \sum_{\delta=1}^d \widehat{K}_W(\delta). \quad (\text{B.4})$$

Table 9 provides examples for the (unweighted) K -density CDF in 1990, 1999, and 2009, respectively; while Table 10 summarizes the year-on-year location patterns of industries based on the formal

⁵The interactions across ‘neighboring cities’ mostly fall into that range in Canada. In particular, a cutoff distance of 800 kilometers includes interactions within the ‘western cluster’ (Calgary, AB; Edmonton, AB; Saskatoon, SK; and Regina, SK); the ‘plains cluster’ (Winnipeg, MN; Regina, SK; Thunder Bay, ON); the ‘central cluster’ (Toronto, ON; Montréal, QC; Ottawa, ON; and Québec, QC); and the ‘Atlantic cluster’ (Halifax, NS; Fredericton, NB; and Charlottetown, PE). Setting the cutoff distance to 800 kilometers allows us to account for industrial localization at both very small spatial scales, but also at larger interregional scales for which market-mediated input-output and demand linkages, as well as market size, might matter much more.

significance test of Duranton and Overman (2005) that we have described in the foregoing.

C. Input-output measures

We construct detailed measures of input and output linkages at the *plant level* by making use of the microgeographic nature of our data.⁶ We use the L -level national input-output tables from Statistics Canada at buyers' prices. These tables – which constitute the finest sectoral public release – feature 42 sectors that are somewhere in between the NAICS 3- and NAICS 4-digit levels. For each industry, i , we allocate total inputs purchased or outputs sold in the L -level matrix to the corresponding NAICS 6-digit sectors. We allocate total sales to each subsector in proportion to that sector's sales in the total sales to obtain a 259×259 matrix of NAICS 6-digit inputs and outputs, which we use in constructing the linkages. From that table, we compute the share α_{ij} that sector i sells to sector j . We also compute the share β_{ij} that sector i buys from sector j . We systematically exclude within-sector transactions where $i = j$, as those may be capturing all sorts of intra-sectoral agglomeration economies that are conducive to clustering but not correlated with input-output linkages. Thus, the weights we use in equations (2) and (3) are given by

$$\omega_{\Omega(\ell),s}^{\text{in}} \equiv \alpha_{\Omega(\ell),s} \quad \text{and} \quad \omega_{\Omega(\ell),s}^{\text{out}} \equiv \beta_{\Omega(\ell),s}. \quad (\text{C.1})$$

Using the L -level matrix provides smoother series of input-output linkages than those obtained using the confidential W -level national input-output tables (which are directly in the 259×259 industries format). Yet, we also compute all our measures using the W -level tables. Results are robust to the choice of input-output matrix.

D. Additional Tables and Results

Table 9 presents the unweighted CDFs for the ten most geographically localized industries for the years 1990, 1999, and 2009. The ranking is based on industries with more than 10 plants in that year and in 10 or more years of the 20-year study period. Table 10 summarizes the location patterns by year and by statistical significance following the methodology developed by Duranton and Overman (2005). Finally, Tables 10 contain the information on the percentage of industries with random, localized, and dispersed point patterns between 1990 and 2009.

⁶Previous work by, e.g., Duranton and Overman (2005, 2008) or Ellison, Glaeser and Kerr (2010) has essentially focused on using input-output tables to identify the most strongly linked industries. To the best of our knowledge, the microgeographic nature of the data has not yet been explored for input-output linkages.

Table 9: Ten most localized NAICS 6-digit industries according to the DO measure (plant counts).

NAICS	Industry description	CDF
1990		
315231	Women's and Girls' Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing	0.62
315233	Women's and Girls' Cut and Sew Dress Manufacturing	0.55
313240	Knit Fabric Mills	0.53
315292	Fur and Leather Clothing Manufacturing	0.42
315291	Infants' Cut and Sew Clothing Manufacturing	0.32
315210	Cut and Sew Clothing Contracting	0.30
337214	Office Furniture (except Wood) Manufacturing	0.21
332720	Turned Product and Screw, Nut and Bolt Manufacturing	0.21
313110	fibre, Yarn and Thread Mills	0.19
333511	Industrial Mould Manufacturing	0.18
1999		
315231	Women's and Girls' Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing	0.63
313240	Knit Fabric Mills	0.47
315210	Cut and Sew Clothing Contracting	0.22
333220	Rubber and Plastics Industry Machinery Manufacturing	0.20
336370	Motor Vehicle Metal Stamping	0.18
332720	Turned Product and Screw, Nut and Bolt Manufacturing	0.18
336330	Motor Vehicle Steering and Suspension Components (except Spring) Manufacturing	0.17
333519	Other Metalworking Machinery Manufacturing	0.16
337214	Office Furniture (except Wood) Manufacturing	0.15
315291	Infants' Cut and Sew Clothing Manufacturing	0.14
2009		
315231	Women's and Girls' Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing	0.61
322299	All Other Converted Paper Product Manufacturing	0.29
337214	Office Furniture (except Wood) Manufacturing	0.17
336370	Motor Vehicle Metal Stamping	0.17
332720	Turned Product and Screw, Nut and Bolt Manufacturing	0.16
337215	Showcase, Partition, Shelving and Locker Manufacturing	0.15
321112	Shingle and Shake Mills	0.14
331420	Copper Rolling, Drawing, Extruding and Alloying	0.13
336360	Motor Vehicle Seating and Interior Trim Manufacturing	0.13
315110	Hosiery and Sock Mills	0.13

Notes: The CDF at distance d is the cumulative sum of the K -densities up to distance d . Results in this table are reported for a distance $d = 50$ kilometers.

Table 10: Percentage of industries with random, localized, and dispersed point patterns, 1990 to 2009.

Year	Plant Counts			Employment weighted			Sales weighted		
	Random	Localized	Dispersed	Random	Localized	Dispersed	Random	Localized	Dispersed
1990	52.53	34.63	12.84	52.53	36.96	10.51	54.86	37.35	7.78
1991	51.36	36.19	12.45	52.92	38.52	8.56	55.25	36.19	8.56
1992	53.70	36.19	10.12	56.42	35.02	8.56	58.37	33.46	8.17
1993	53.70	34.24	12.06	58.37	33.46	8.17	59.53	31.52	8.95
1994	49.81	36.96	13.23	57.20	33.07	9.73	60.70	30.74	8.56
1995	55.25	33.46	11.28	58.37	33.07	8.56	59.53	32.30	8.17
1996	54.09	35.41	10.51	56.03	35.41	8.56	59.53	33.46	7.00
1997	55.25	35.41	9.34	60.70	32.30	7.00	61.09	32.68	6.23
1998	55.64	34.24	10.12	58.37	35.02	6.61	61.87	32.68	5.45
1999	55.25	34.63	10.12	58.75	35.41	5.84	61.48	32.30	6.23
2000	47.86	37.74	14.40	51.75	40.47	7.78	53.31	40.47	6.23
2001	43.58	41.25	15.18	52.92	40.86	6.23	50.58	42.41	7.00
2002	45.91	39.69	14.40	50.97	41.63	7.39	54.86	37.35	7.78
2003	47.47	36.58	15.95	50.58	40.86	8.56	55.64	35.41	8.95
2004	60.31	30.35	9.34	60.31	33.07	6.61	60.70	32.30	7.00
2005	58.75	33.46	7.78	62.65	31.13	6.23	64.20	31.52	4.28
2006	60.31	30.35	9.34	60.31	33.46	6.23	62.26	33.85	3.89
2007	57.59	33.46	8.95	60.70	33.85	5.45	62.65	32.30	5.06
2008	56.03	34.24	9.73	61.48	31.91	6.61	64.59	29.96	5.45
2009	59.53	33.07	7.39	63.04	31.52	5.45	63.04	31.13	5.84

Source: Authors' computations using Annual Survey of Manufacturers data.