C.I.R.P.É.E

Centre Interuniversitaire sur le Risque,
les Politiques Économiques et l'Emploi

Cahier de recherche/Working Paper **13-27**

An Anatomy of the Geographical Concentration of Canadian Manufacturing Industries

Kristian Behrens
Théophile Bougna

Septembre/September 2013

Behrens: Canada Research Chair, Département des Sciences Économiques, École des Sciences de Gestion, Université du Québec à Montréal (ESG UQAM), Canada; CIRPÉE and CEPR
behrens.kristian@uqam.ca
Bougna: Département des Sciences Économiques, École des Sciences de Gestion, Université du Québec à Montréal (ESG UQAM), Canada
bougna_lonla.theophile@uqam.ca

Abstract:
We document the location patterns of Canadian manufacturing industries – as well as changes in those patterns over the first decade of 2000 – using detailed micro-geographic data. Depending on industry definitions and years, 40 to 60 percent of industries are clustered. According to our measures, manufacturing industries become less geographically concentrated in Canada, i.e., localization is decreasing. Yet, some of the most localized industries are becoming even more localized. We also document the locational trends specific to small firms, young firms, and exporters. We find that their location patterns do not differ significantly from that of the other firms in their industries.

Keywords: Location patterns, manufacturing industries, micro-geographic data, Canada

JEL Classification: R12, L60

> "*Step back and ask, what is the most striking feature of the geography of economic activity? The short answer is surely concentration.*" (Krugman, 1991, p.5)

1 Introduction

One of the most salient features of the economic landscape is the strong geographical concentration of economic activity. That concentration is observed in most countries and at various spatial scales. Famous examples of 'clusters' include the high-technology concentrations of Silicon Valley, Boston's Route 128, and the North Carolina research triangle, as well as concentrations of more mature industries like the automotive cluster in the Detroit-Winsor corridor or the Italian manufacturing 'districts'. In Canada, economic activity – measured by either GDP or employment – is strongly concentrated *across and within* provinces. Ontario and Quebec, for example, are home to about 60 percent of Canadian GDP and 75 percent of manufacturing employment (Statistics Canada, 2013). Within those two provinces, the Toronto metropolitan area, about 0.06 percent of Ontario's provincial surface, concentrates 45 percent of Ontario's GDP; whereas the Montreal metropolitan area concentrates almost 35 percent of Quebec's GDP on about 0.04 percent of Quebec's provincial surface (Institut de Statistiques du Québec, 2013).

The resurgence of spatial analysis in economics has led to a renewed interest in theoretically explaining and empirically analyzing the strong geographical concentration of industries. Clusters and regional development have also often been – and are becoming increasingly more – a matter of concern for policy makers around the world. Quebec's Government, for example, has recently launched the 'Plan Nord', with the aim to invest around $80 billion over the next 25 years to create 20,000 jobs, generate $14 billion in government revenue, and $162 billion for Quebec's GDP. Such huge investment plans – which have a clear regional development component – are unlikely to leave the geography of economic activity unchanged. It is, therefore, important to understand which industries tend to cluster, what location patterns we observe for specific firms that are important targets for economic development (young firms, small firms, exporters), and what the broad trends of geographical concentration are over the last decade. This is the focus of the present paper. A fine analysis of the geographical concentration of industries is a prerequisite to any subsequent analysis dealing with the potential static and dynamic productivity gains – both locally and nationally (see Behrens, 2013, for a recent analysis) – arising from that concentration.

There is a substantial literature dealing with the measurement of industrial localization, i.e., the geographical concentration of industries. Ellison and Glaeser (1997; henceforth EG) have developed an index that has been widely applied to that issue. Despite its numerous advantages and appealing theoretical properties, that index has no strong spatial flavor as it does not take into account the relative positions of the geographical units. We address that issue using two alternative strategies. First, we analyze the geographical concentration in

Canada by explicitly integrating 'neighborhood effects' into the EG index, following recent work by Guimaraes, Figueiredo, and Woodward (2011). Second, we exploit the micro-geographic nature of our data to compute point pattern based continuous measures following Ripley (1976, 1977), Duranton and Overman (2005, 2008; henceforth DO), and Marcon and Puech (2003, 2010). Using continuous measures allows us to sidestep the need for pre-defined administrative units, which give rise to the well-known *modifiable areal unit problem* (henceforth MAUP; Openshaw and Taylor, 1979; Openshaw, 1983).

Using the Ellison-Glaeser and the Duranton-Overman measures of localization, we identify the most concentrated and the most dispersed manufacturing industries in Canada. Consistent with previous findings for the UK, France, and Japan, industries related to textiles and to the extraction of natural resources rank among the most localized industries. We also provide a broad picture of the main trends for the first decade of 2000. Our key findings can be summarized as follows. First, depending on industry definitions and years, 40 to 60 percent of manufacturing industries are clustered, mainly at short distances (less than 150 kilometers), and at distances of about 500 kilometers. These figures suggest that there is less industrial localization in Canada as compared to other developed countries like France or the UK. Second, localization is decreasing, i.e., manufacturing industries become less geographically concentrated in Canada. Third, some of the most strongly localized industries are becoming even more localized. Last, we document the location trends for some specific subgroups of firms: small firms, young firms, and exporters. Understanding those trends is relevant from a policy perspective, since these groups of firms are perceived as being vital for employment, economic growth, and local regional development. Therefore, they are often targets for cluster policy. Our findings suggest that these types of firms are, in general, not more strongly concentrated than all firms in their respective industries. The only exception is for exporters, but their 'excess concentration' tends to get weaker over the first decade of 2000. Hence, targeting those types of firms specifically seems questionable from a policy point-of-view.

What distinguishes our contribution from the existing literature on the geographic concentration of industries is that, to the best of our knowledge, continuous localization measures have until now neither been applied to Canadian data in particular, nor to North American data in general (see Holmes and Stevens, 2004). The empirical literature on localization using micro-geographic data – though growing recently – is still relatively limited. We also use a rich dataset for a recent period that contains a large number of small and young firms – which allows us to examine the location trends specific to those types of firms – and that reports plant-level information on export status – which allows us to look at trends specific to firms involved in international business. Last, our dataset spans a ten year period, which allows us to look at the dynamics of localization. We are not aware of any other study looking at the changes in localization over time using large micro-geographic datasets of firms.

The remainder of the paper is organized as follows. Section 2 provides a snapshot of

manufacturing in Canada. Section 3 presents our empirical results using discrete measures of localization, whereas Section 4 summarizes our empirical results using continuous measures. Finally, Section 5 concludes and places our results more broadly into the policy debate about industry clusters and regional development. We relegate all technicalities and a detailed description of our datasets to an extensive set of appendices.

2 A Snapshot of Canadian Manufacturing, 2001–2009

To set the stage, we first provide a quick overview of the sectoral and geographical structure of manufacturing in Canada from 2001 to 2009. Total salaried employment in Canada in 2001 was 12,978,258 jobs, of which 1,974,636 – or 15.21 percent – were in manufacturing. In 2005, the corresponding numbers were 13,931,343 and 1,837,828 jobs – or 13.19 percent – respectively; whereas they were 14,570,025 and 1,473,472 jobs – or 10.11 percent – in 2009.[1] The downwards trend in manufacturing can also be seen from Table 1, which shows that the number of plants in our data has fallen from 54,379 in 2001 to 46,391 in 2009. This 'de-industrialization' is not specific to Canada and affects most developed countries in a similar way (see, e.g., Duranton, Martin, Mayer, Mayneris, 2011, for the French case). Note that, as can be seen from Table 1, the decrease in the number of plants went hand-in-hand with an increase in average plant size – as measured by employment – except for the Atlantic provinces.

Table 1: Descriptive Statistics by Province.

	2001		2005		2009	
Province	# of plants	Avg. empl.	# of plants	Avg. empl.	# of plants	Avg. empl.
Alberta	3,933	36.100	3,455	44.430	3,581	52.780
British Columbia	6,219	31.930	5,371	33.730	4,991	34.370
Manitoba	1,654	43.330	1,481	55.230	1,263	57.790
New Brunswick	1,395	35.660	1,258	40.080	1,175	36.940
Newfoundland and Labrador	576	43.830	540	44.830	472	42.500
Nova Scotia	1,676	29.930	1,495	37.140	1,296	35.020
Ontario	21,306	45.010	20,966	46.080	19,637	46.760
Prince Edward Island	328	25.350	327	24.410	280	25.430
Quebec	15,939	41.640	14,166	45.690	12,560	49.550
Saskatchewan	1,353	27.360	1,305	32.520	1,091	36.230
Territories	–	–	40	5.940	45	10.140
Total	**54,379**	**36.01**	**50,404**	**37.28**	**46,391**	**38.86**

Source: Authors' computations using Scott's National All Business Directories.

Table 2 shows that our data includes many small manufacturing firms, so that the size distribution of plants is very skewed towards small establishments.[2] On average, only 15 percent of plants have more than 50 employees, whereas the majority of plants – about 70 percent – employ between 1 and 20 workers. Consistent with previous findings in the literature, the size

[1] Source: Statistics Canada, CANSIM.

[2] For simplicity, in what follows we use the terms 'firm' and 'plant' interchangeably.

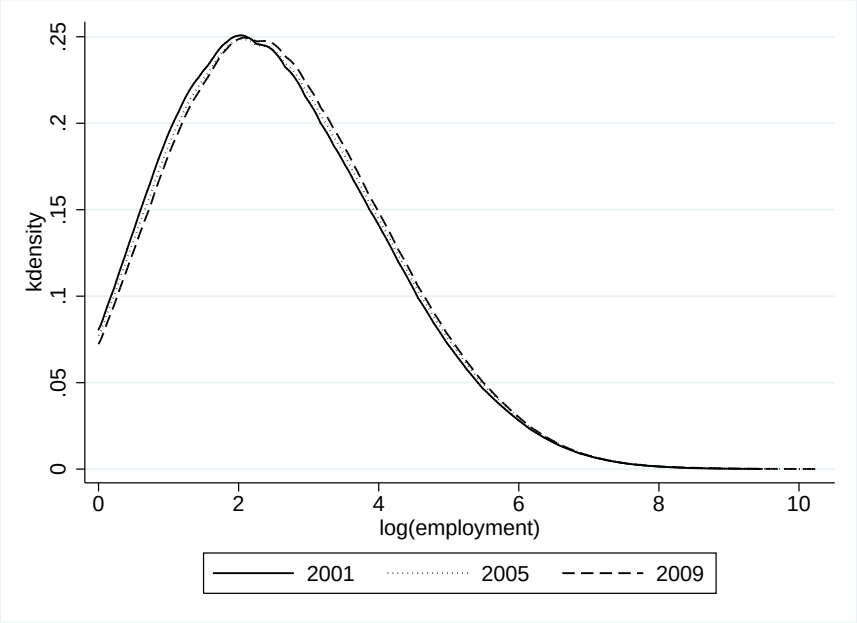

Figure 1: Kernel Densities of (Log) Plant Sizes, Measured by Employment.

distribution of plants is roughly log-normal, and close to Zipf in the upper tail of the distribution (Axtell, 2001; Cabral and Mata, 2003). Figure 1 depicts the kernel densities of the size distributions for 2001, 2005, and 2009, respectively. As can be seen from that figure, the distribution remains fairly stable, despite substantial plant turnover. However, as can also be seen from Figure 1, the distribution gradually shifts to the right – plants become larger on average. Finally, Table 3 summarizes industry-level details of our data, including the average plant size by industry and the number of exporting firms. There is clearly substantial variation across industries, as extensively documented by previous studies (e.g., Bernard and Jensen, 1995).

Table 2: Plant Size Distribution.

	2001		2005		2009	
# of employees	# of plants	Share	# of plants	Share	# of plants	Share
1	2,761	5.277	2,406	4.863	1,978	4.404
2-20	33,249	63.543	31,052	62.763	27,958	62.242
21-50	8,533	16.308	8,231	16.636	7,661	17.056
>50	7,782	14.872	7,787	15.739	7,321	16.299
Total	**52,325**	**100%**	**49,476**	**100%**	**44,918**	**100%**

Source: Authors' computations using Scott's National All Business Directories.

Turning to the spatial dimension, it is well known that population is strongly concentrated geographically in Canada. Indeed, because of historical settlement patterns, the climatic conditions in the north, and access to the large US market to the south, about 90 percent of the Canadian population lives less than 100 miles from the US border. Quite naturally, the overall distribution of manufacturing is thus also strongly concentrated geographically in Canada – namely in Ontario and Quebec and, more generally, along the Canada-US border – as can be

seen from Figure 2. Table 1 provides a more detailed geographical breakdown, and reports the number of plants and the average plant size by province. Observe that Figure 2 displays the 'de-industrialization' trend we mentioned in the foregoing. There is indeed some 'breaking up' of the concentration of manufacturing between 2001 and 2009, especially in the traditional manufacturing corridor running from Quebec City to Winsor, via Montreal and Toronto.

Table 3: Breakdown of Firms by NAICS 3-Digit Industries.

NAICS3	Industry name	# NAICS6	# of firms			Avg. plant size (empl.)			# of exporters		
			2001	2005	2009	2001	2005	2009	2001	2005	2009
311	Food Manufacturing	33	4,807	4,327	3,929	50.114	56.711	62.158	1,667	1,591	1,404
312	Beverage & Tobacco Product Mfg	6	477	426	462	64.522	77.345	64.036	129	134	126
313	Textile Mills	7	539	356	277	51.986	53.858	53.359	246	198	162
314	Textile Product Mills	4	1,413	1,307	1,146	18.340	17.568	17.147	422	488	430
315	Apparel Manufacturing	17	2,364	1,905	1,354	40.631	38.855	36.349	932	819	642
316	Leather & Allied Product Mfg	3	382	308	238	36.728	28.454	29.091	203	163	131
321	Wood Product Manufacturing	14	3,919	3,546	3,127	42.826	48.239	48.557	1,733	1,690	1,436
322	Paper Manufacturing	12	911	854	775	119.594	114.557	115.001	582	588	546
323	Printing & Related Support activ.	6	5,091	4,577	4,089	18.600	22.935	23.964	1,063	1,174	1,041
324	Petroleum & Coal Products Mfg	4	347	318	301	100.009	135.365	130.882	123	115	106
325	Chemical Manufacturing	20	2,183	2,034	1,982	47.907	56.685	63.959	1,231	1,205	1,146
326	Plastics & Rubber Products	14	2,206	2,227	2,084	48.950	57.802	54.252	1,375	1,423	1,334
327	Nonmetallic Mineral Products	12	2,608	2,618	2,473	27.539	27.651	42.394	778	808	766
331	Primary Metal Manufacturing	13	927	820	805	113.145	115.373	106.953	587	534	484
332	Fabricated Metal product Mfg	21	8,018	7,521	7,255	26.504	30.020	31.093	3,014	3,085	2,975
333	Machinery Manufacturing	17	5,237	4,758	4,583	34.210	37.538	41.780	3,160	3,147	2,994
334	Computer & Electronic Products	9	2,130	1,654	1,643	61.658	59.794	63.845	1,433	1,205	1,201
335	Electrical Equip. & Appliances	12	1,193	1,047	1,007	43.489	50.602	47.018	777	749	707
336	Transportation Equipment Mfg	18	2,008	1,907	1,839	116.297	129.609	125.060	990	1,010	918
337	Furniture & Related Product Mfg	10	3,526	3,351	2,869	25.192	29.308	32.065	1,126	1,198	1,001
339	Miscellaneous Manufacturing	7	4,093	4,543	4,153	17.337	16.022	15.934	1,434	1,467	1,353
		259	**54,379**	**50,404**	**46,391**	**52.647**	**57.347**	**57.347**	**23,005**	**22,791**	**20,903**
									42.3%	**45.2%**	**45.1%**

Source: Authors' computations using Scott's National All Business Directories.

Since manufacturing is strongly concentrated geographically in Canada, we will use its overall distribution as the benchmark against which we assess localization in a given sector. Doing so avoids picking up localization patterns that are solely driven by the overall concentration of industries in large metropolitan areas (Combes, Mayer, and Thisse, 2008). We will compute both discrete and continuous measures of localization – for industries in general, but also for certain types of firms like small firms, young firms, and exporters – and analyze their trends over time. When looking at specific types of firms, we will use an even more restrictive benchmark, namely the spatial distribution of all firms in *their industry*. In other words, we will look at the 'excess concentration' of small firms, young firms, and exporters as compared to the concentration of firms in their industry in general. Doing so will provide a very fine picture of the 'state of geography' of manufacturing in Canada, both in terms of industries and in terms of specific types of firms.

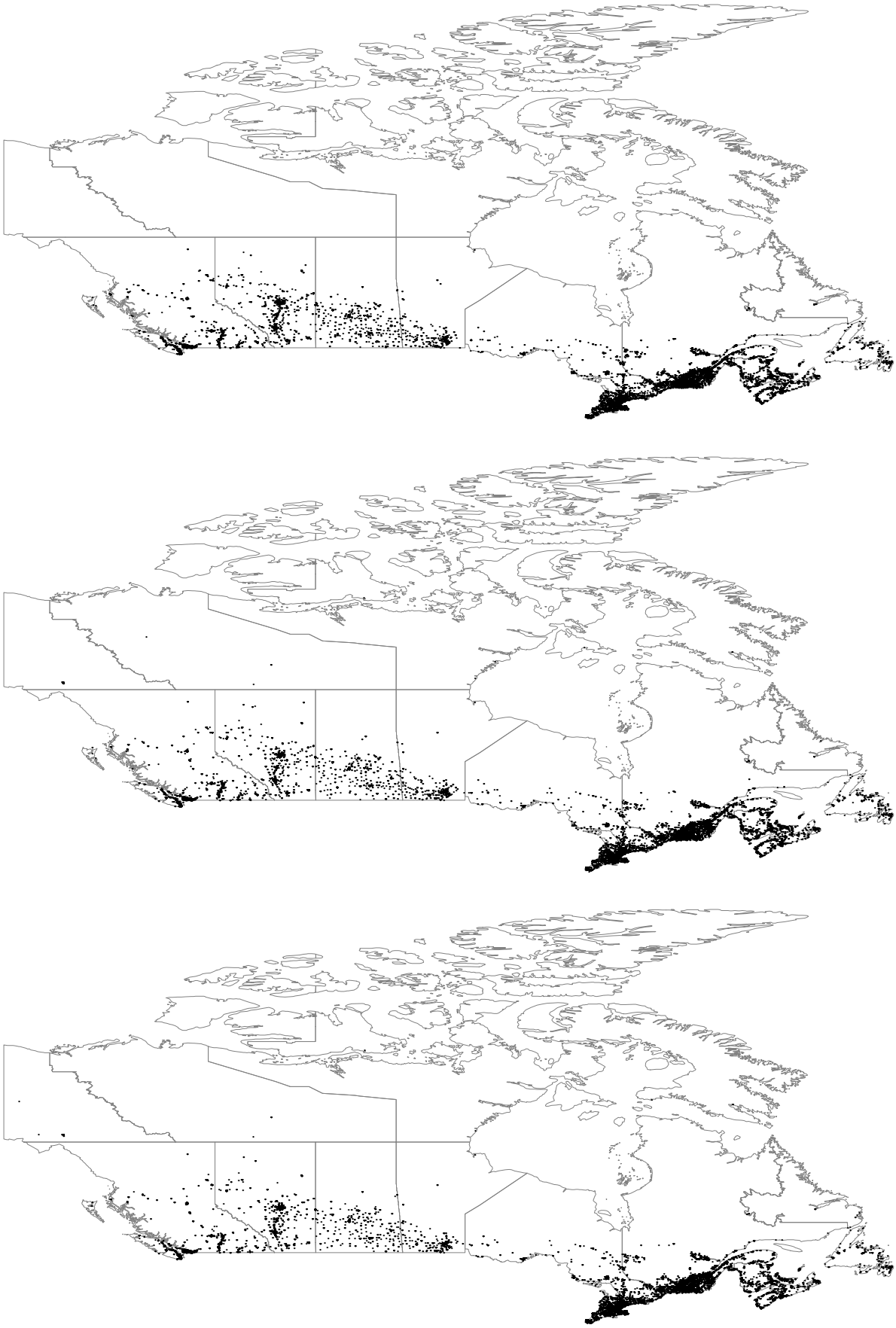

Figure 2: Distribution of Manufacturing Firms in 2001 (Top), 2005 (Middle), and 2009 (Bottom).

3 Discrete Measures: Methodology and Results

We first provide results on the geographical concentration of industries using discrete measures of localization. More precisely, we start by computing the geographical concentration using the Ellison-Glaeser index (Ellison and Glaeser, 1997). This measure, though somewhat sensitive to the way space is subdivided into administrative units, has been widely used in the literature and will allow us to compare our results to existing ones. We also compute a spatially weighted version of the EG index to take into account 'neighborhood effects', i.e., the fact that industry concentrations may stretch across several adjacent administrative units.We later exploit, in Section 4, the micro-geographic nature of our data and provide results using continuous measures of localization to obtain a sharper overall picture of the degree of geographical concentration that is independent of spatial subdivisions.

3.1 Baseline Results

We compute the EG index – and its spatially weighted version (henceforth, EGspat) – for 2001, 2005, and 2009 at the NAICS 6-digit level using three different spatial scales: provinces (PROV), economic regions (ER), and census divisions (CD). Our key findings, shown in Table 4, can be summarized as follows.

First, about 70 to 75 percent of manufacturing industries are localized in Canada. This fraction is lower than the one reported for the US (97 percent), France (95 percent), and the UK (94 percent) in earlier studies by Ellison and Glaeser (1997), Maurel and Sédillot (1999), and Duranton and Overman (2005).[3] Second, the number of localized manufacturing sectors in Canada has decreased between 2001 and 2009. This can be seen in terms of numbers, but also from the decrease in the mean value of the EG index at all spatial scales, safe for the smallest one (CD). We also find that there is a sizeable share of sectors for which the EG index is negative, thus suggesting that dispersion prevails – and increases over time – in some industries. When taken together, all of our foregoing results point to the fact that manufacturing industries have become less geographically concentrated over the first decade of 2000. Third, despite some changes across industries, the EG index is, on average, smaller than its spatially weighted counterpart (see the two bottom panels of Table 4). Put differently, spatial concentration extends over multiple adjacent spatial units, and this fact has to be taken into account when computing the EG index. Note that all our results are fairly robust across years, spatial scales, and to the use of the chosen weighting scheme for computing the EGspat

[3]Duranton and Overman (2005) note that the definition of 'weak localization' by Ellison and Glaeser (1997) picks up manufacturing industries in the UK which have a pattern that is not significantly different from that of spatial randomness. We come back to that point in the next section. Observe that our mean value for the unweighted index at the ER level is very close to the one of 0.034 reported by Duranton and Overman (2005), whereas our median is higher.

index (see the Appendix).

Figure 3 summarizes the distributions of the EG and EGspat indices for the 259 6-digit manufacturing industries in 2001, 2005, and 2009. Observe that these distributions are quite skewed towards zero, i.e., only few industries are highly agglomerated, whereas a majority of them are weakly agglomerated – the EG index is positive but less than 0.05. These results are similar to the ones reported by Maurel and Sédillot (1999) for French industries, and by Ellison and Glaeser (1997) for US industries. We can also see that, despite the general trend towards a decrease in localization between 2001 and 2009, the overall distributions of the EG and EGspat indices have remained fairly stable over time.[4]

Table 4: Mean and Median EG and EGspat Indices at Different Spatial Scales, NAICS 6-Digit Industries.

	2001			2005			2009		
Geography	PROV	ER	CD	PROV	ER	CD	PROV	ER	CD
	Unweighted EG								
Mean	0.074	0.036	0.021	0.073	0.035	0.023	0.060	0.032	0.020
Median	0.023	0.021	0.010	0.023	0.018	0.010	0.019	0.015	0.010
Share < 0	31.660	23.552	26.255	35.521	25.483	25.483	36.154	29.615	29.231
Share $\in (0, 0.05]$	26.255	47.876	58.301	23.552	47.876	59.459	27.692	44.231	56.538
Share > 0.05	42.085	28.571	15.444	40.927	26.641	15.058	36.154	26.154	14.231
	EG weighted by the inverse distance matrix								
Mean	0.080	0.047	0.029	0.086	0.049	0.032	0.077	0.048	0.031
Median	0.025	0.026	0.017	0.028	0.024	0.0157	0.024	0.024	0.016
Share < 0	31.660	17.375	16.602	34.363	18.533	16.602	33.846	20.769	20.000
Share $\in (0, 0.05]$	25.869	47.876	65.251	23.552	47.104	64.479	26.538	45.769	60.000
Share > 0.05	42.471	34.749	18.147	42.085	34.363	18.919	39.615	33.462	20.000
	EG weighted by the common border length								
Mean	0.077	0.051	–	0.093	0.054	–	0.085	0.052	–
Median	0.027	0.030	–	0.026	0.027	–	0.021	0.024	–
Share < 0	32.432	17.761	–	31.274	19.691	–	33.462	23.846	–
Share $\in (0, 0.05]$	26.641	45.946	–	27.027	44.402	–	28.077	41.538	–
Share > 0.05	40.927	36.293	–	41.699	35.907	–	38.462	34.615	–

Notes: Mean and median values for 259 (resp., 260 in 2009) NAICS 6-digit industries. Share < 0 means 'not clustered'. Share $\in (0, 0.05]$ means 'weakly clustered'. Share > 0.05 means 'strongly clustered'. See Ellison and Glaeser (1997) for details.

Table 5 lists the ten most and the ten least localized industries at the NAICS 6-digit level for the years 2001, 2005, and 2009, respectively. As can be seen from that table, various industries related to either textiles or to the extraction and processing of natural resources dominate the group of the most localized industries. The hierarchy of individual industries is almost unchanged when using the EGspat index.[5] Notable exceptions include 'Petrochemical Manufacturing' (NAICS 325110) in 2005. This industry appears dispersed according to the unweighted

[4]The correlation of the EG indices across industries in 2001 and 2009 varies from about 0.83 at the province level to 0.73 at the census division level. The processes generating province-level agglomeration are different from the ones generating agglomeration at the economic region and census division levels (see Rosenthal and Strange, 2003).

[5]The Spearman-rank correlation between the EG and the EGspat indices is 0.96.

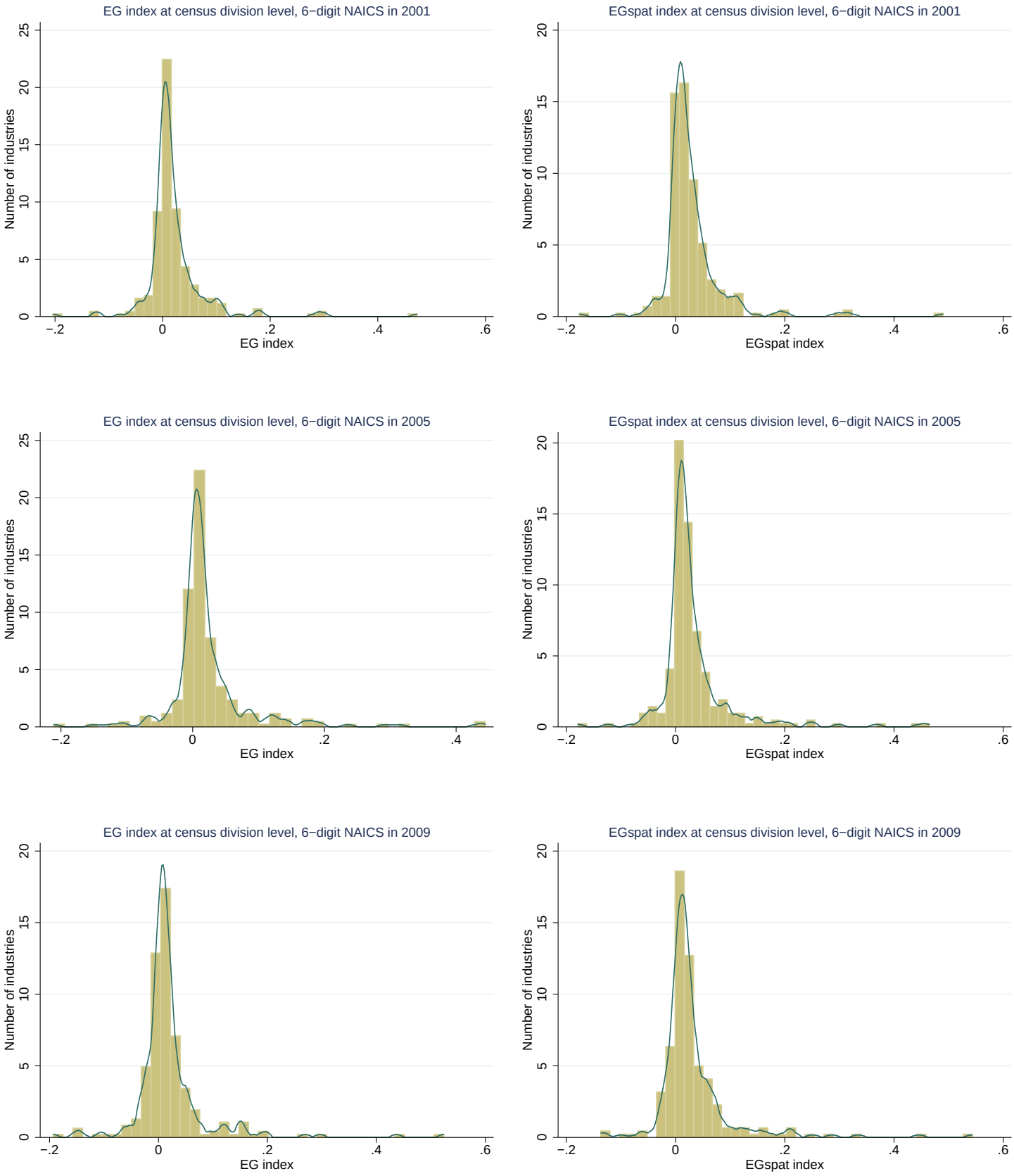

Figure 3: Distribution of the EG Index at the CD Level (NAICS 6-digit), Unweighted EG (Left Panel) and Spatially-weighted EGspat (Right Panel).

EG index, and weakly concentrated using a spatial weighting scheme.

3.2 Sectoral Scope of Localization

Does the level of sectoral aggregation matter for our results? Do NAICS 4-digit industries exhibit comparable location patterns to NAICS 6-digit industries? The short answers to those two questions are 'yes' and 'no'. Our results below indeed show that, as we move to a more aggregate definition of industries, the degree of concentration changes. On average, we observe less geographical concentration at larger spatial scales. Even if there are also less dispersed industries (11 percent on average, compared to 27 percent at the 6-digit level), the geographic concentration at the 4-digit level is weaker than at the 6-digit level. This result is reminiscent of findings by Rosenthal and Strange (2003), who show that the average level of agglomeration increases as one moves from 4- to 6-digit industries. It also holds true as the geographical scale goes from economic regions to provinces, and from census divisions to economic regions. Table 6 summarizes our results for the 86 NAICS 4-digit industries. Table 7 reveals that there are systematic localization patterns by broad industry groups. Some 3-digit industries are made up of many concentrated 6-digit subindustries (e.g., 'Apparel manufacturing' or 'Chemical manufacturing'), whereas others are mostly dispersed (e.g., 'Beverage and Tobacco Product Manufacturing'). This suggests that localization trends extend across different 3-digit groupings.

Table 16 in the Appendix lists the ten most and the ten least localized industries at the 4-digit level for the years 2001, 2005, and 2009, respectively. The hierarchy of the different industries remains nearly unchanged when compared to the 6-digit level. Industries related to either textiles or to the extraction and processing of natural resources dominate again the group of the most localized industries, whereas industries related to dairy and tobacco remain among the least localized ones.

3.3 Location Patterns of Small Firms, Young Firms, and Exporters

We now stratify our sample into three sub-samples including small firms, young firms, and exporters, respectively. Since these firms are considered to be crucial for employment and economic development, understanding their localization patterns seems to be important from a policy perspective. Our aim is to analyze their locational trends and compare them to the overall trend in their sector. What patterns of industrial localization do we observe for these specific firms, and what are the changes in their geographical concentration over the last decade? Are small firms, young firms, or exporters more concentrated than firms in their industry in general, i.e., is there 'excess concentration'?

To answer these questions, we construct three industry subsamples. The first relates to small-scale firms. Instead of using Statistics Canada's definition of small-scale business – a firm

Table 5: Ten Most and Least Localized 6-Digit Industries, EG and EGspat indices.

NAICS 6	Most localized industries in 2001	EG	EGspat
315231	Women's and Girls' Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing	0.473	0.490
315233	Women's and Girls' Cut and Sew Dress Manufacturing	0.301	0.305
315221	Men's and Boys' Cut and Sew Underwear and Nightwear Manufacturing	0.294	0.321
315110	Hosiery and Sock Mills	0.282	0.292
321112	Shingle and Shake Mills	0.184	0.202
313240	Knit Fabric Mills	0.179	0.194
333130	Mining and Oil and Gas Field Machinery Manufacturing	0.173	0.181
339930	Doll, Toy and Game Manufacturing	0.141	0.144
325210	Resin and Synthetic Rubber Manufacturing	0.112	0.114
332113	Forging	0.110	0.120

NAICS 6	Least localized industries in 2001	EG	EGspat
327214	Glass Manufacturing	-0.039	-0.037
311823	Dry Pasta Manufacturing	-0.042	-0.039
315291	Infants' Cut and Sew Clothing Manufacturing	-0.047	-0.021
326121	Unlaminated Plastics Profile Shape Manufacturing	-0.051	-0.036
313320	Fabric Coating	-0.056	-0.047
311310	Sugar Manufacturing	-0.057	-0.045
326193	Motor Vehicle Plastic Parts Manufacturing	-0.079	-0.067
335229	Other Major Appliance Manufacturing	-0.122	-0.102
325110	Petrochemical Manufacturing	-0.124	-0.050
315227	Men's and Boys' Cut and Sew Trouser, Slack and Jean Manufacturing	-0.204	-0.176

NAICS 6	Most localized industries in 2005	EG	EGspat
315221	Men's and Boys' Cut and Sew Underwear and Nightwear Manufacturing	0.444	0.464
315231	Women's and Girls' Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing	0.431	0.446
325181	Alkali and Chlorine Manufacturing	0.314	0.368
315233	Women's and Girls' Cut and Sew Dress Manufacturing	0.292	0.298
315110	Hosiery and Sock Mills	0.235	0.249
321112	Shingle and Shake Mills	0.193	0.247
333130	Mining and Oil and Gas Field Machinery Manufacturing	0.189	0.200
311111	Dog and Cat Food Manufacturing	0.178	0.207
315234	Women's and Girls' Cut and Sew Suit, Coat, Tailored Jacket and Skirt Manufacturing	0.171	0.182
315190	Other Clothing Knitting Mills	0.167	0.184

NAICS 6	Least localized industries in 2005	EG	EGspat
324190	Other Petroleum and Coal Products Manufacturing	-0.061	-0.058
325120	Industrial Gas Manufacturing	-0.065	-0.053
315227	Men's and Boys' Cut and Sew Trouser, Slack and Jean Manufacturing	-0.065	-0.041
327310	Cement Manufacturing	-0.066	-0.060
336330	Motor Vehicle Steering and Suspension Components (except Spring) Manufacturing	-0.077	-0.049
335229	Other Major Appliance Manufacturing	-0.102	-0.081
311830	Tortilla Manufacturing	-0.112	-0.035
325110	Petrochemical Manufacturing	-0.129	0.007
335110	Electric Lamp Bulb and Parts Manufacturing	-0.150	-0.124
333611	Turbine and Turbine Generator Set Unit Manufacturing	-0.211	-0.179

Table 5 (continued): Ten Most and Least Localized 6-Digit Industries, EG and EGspat indices.

NAICS 6	Most localized industries in 2009	EG	EGspat
315231	Women's and Girls' Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing	0.524	0.543
315233	Women's and Girls' Cut and Sew Dress Manufacturing	0.437	0.446
315221	Men's and Boys' Cut and Sew Underwear and Nightwear Manufacturing	0.296	0.334
333130	Mining and Oil and Gas Field Machinery Manufacturing	0.269	0.286
313240	Knit Fabric Mills	0.196	0.216
315232	Women's and Girls' Cut and Sew Blouse and Shirt Manufacturing	0.195	0.204
315190	Other Clothing Knitting Mills	0.174	0.202
325181	Alkali and Chlorine Manufacturing	0.155	0.251
321112	Shingle and Shake Mills	0.154	0.214
311111	Dog and Cat Food Manufacturing	0.151	0.180

NAICS 6	Least localized industries in 2009	EG	EGspat
315227	Men's and Boys' Cut and Sew Trouser, Slack and Jean Manufacturing	-0.056	-0.034
339930	Doll, Toy and Game Manufacturing	-0.059	-0.056
336330	Motor Vehicle Steering and Suspension Components (except Spring) Manufacturing	-0.063	-0.033
335110	Electric Lamp Bulb and Parts Manufacturing	-0.072	-0.056
311830	Tortilla Manufacturing	-0.100	0.021
333611	Turbine and Turbine Generator Set Unit Manufacturing	-0.109	-0.099
327990	All Other Non-Metallic Mineral Product Manufacturing	-0.139	-0.137
312210	Tobacco Stemming and Redrying	-0.148	-0.072
325110	Petrochemical Manufacturing	-0.155	-0.012
321217	Waferboard Mills	-0.193	-0.129

Notes: EG and EGspat indices computed at the 6-digit NAICS level. The spatial scale used is census divisions (CD), and the weighting is inverse distance between CD centroids.

Table 6: Mean and Median EG and EGspat Indices at Different Spatial Scales, NAICS 4-Digit Industries.

	2001			2005			2009		
Geography	PROV	ER	CD	PROV	ER	CD	PROV	ER	CD
	Unweighted EG								
Mean	0.064	0.033	0.019	0.065	0.031	0.020	0.056	0.027	0.015
Median	0.018	0.023	0.012	0.017	0.016	0.011	0.018	0.010	0.007
Share < 0	16.279	6.977	6.977	22.093	11.628	11.628	23.256	16.279	15.116
Share $\in (0, 0.05]$	45.349	68.605	82.558	43.023	69.767	79.070	46.512	62.791	76.744
Share > 0.05	38.372	24.419	10.465	34.884	18.605	9.302	30.233	20.930	8.140
	EG weighted by the inverse distance matrix								
Mean	0.066	0.036	0.022	0.068	0.035	0.022	0.060	0.031	0.018
Median	0.019	0.027	0.014	0.017	0.021	0.013	0.020	0.013	0.011
Share < 0	18.605	5.814	6.977	23.256	8.140	5.814	22.093	12.791	11.628
Share $\in (0, 0.05]$	43.023	67.442	82.558	41.860	72.093	82.558	48.837	65.116	77.907
Share > 0.05	38.372	26.744	10.465	34.884	19.767	11.628	29.070	22.093	10.465
	EG weighted by the common border length								
Mean	0.064	0.040	–	0.071	0.039	–	0.064	0.033	–
Median	0.022	0.029	–	0.022	0.023	–	0.021	0.015	–
Share < 0	18.605	5.814	–	22.093	8.140	–	23.256	15.116	–
Share $\in (0, 0.05]$	44.186	63.953	–	43.023	65.116	–	38.372	61.628	–
Share > 0.05	37.209	30.233	–	34.884	26.744	–	38.372	23.256	–

Notes: Mean and median values for 86 NAICS 4-digit industries. Share < 0 means 'not clustered'. Share $\in (0, 0.05]$ means 'weakly clustered'. Share > 0.05 means 'strongly clustered'. See Ellison and Glaeser (1997) for details.

Table 7: Localization Patterns by Broad Industry Groups.

NAICS3	Industry name	Subsectors	# of localized subsectors			# of dispersed subsectors		
			2001	2005	2009	2001	2005	2009
	Unweighted EG index							
311	Food Manufacturing	33	17	19	18	16	14	15
312	Beverage and Tobacco Product Manufacturing	6	3	4	4	3	2	2
313	Textile Mills	7	5	6	5	2	1	2
314	Textile Product Mills	4	2	3	3	2	1	1
315	Apparel Manufacturing	17	15	15	14	2	2	3
316	Leather and Allied Product Manufacturing	3	3	2	3		1	
321	Wood Product Manufacturing	14	12	13	11	2	1	3
322	Paper Manufacturing	12	9	11	5	3	1	7
323	Printing and Related Support Activities	6	6	6	3			3
324	Petroleum and Coal Products Manufacturing	4	4	3	2		1	2
325	Chemical Manufacturing	20	15	16	16	5	4	4
326	Plastics and Rubber Products Manufacturing	14	8	11	12	6	3	3
327	Nonmetallic Mineral Product Manufacturing	12	7	9	7	5	3	5
331	Primary Metal Manufacturing	13	9	9	9	4	4	4
332	Fabricated Metal Product Manufacturing	21	15	13	16	6	8	5
333	Machinery Manufacturing	17	15	14	15	2	3	2
334	Computer and Electronic Product Manufacturing	9	8	7	6	1	2	3
335	Electrical Equipment, Appliances and Components	12	9	6	7	3	6	5
336	Transportation Equipment Manufacturing	18	14	12	14	4	6	4
337	Furniture and Related Product Manufacturing	10	9	8	9	1	2	1
339	Miscellaneous Manufacturing	7	6	6	5	1	1	2
	Total	259	191	193	184	68	66	76
	% of localized or dispersed		73.745	74.517	70.769	26.255	25.483	29.231
	Weighted by inverse distance EGspat index							
311	Food Manufacturing	33	23	22	26	10	11	7
312	Beverage and Tobacco Product Manufacturing	6	4	5	4	2	1	2
313	Textile Mills	7	5	7	6	2		1
314	Textile Product Mills	4	3	3	3	1	1	1
315	Apparel Manufacturing	17	15	16	14	2	1	3
316	Leather and Allied Product Manufacturing	3	3	3	3			
321	Wood Product Manufacturing	14	14	14	13			1
322	Paper Manufacturing	12	10	12	8	2		4
323	Printing and Related Support Activities	6	6	6	4			2
324	Petroleum and Coal Products Manufacturing	4	4	3	3		1	1
325	Chemical Manufacturing	20	19	19	17	1	1	3
326	Plastics and Rubber Products Manufacturing	14	10	13	12	4	1	2
327	Nonmetallic Mineral Product Manufacturing	12	8	9	8	4	3	4
331	Primary Metal Manufacturing	13	11	11	10	2	2	3
332	Fabricated Metal Product Manufacturing	21	17	15	17	4	6	4
333	Machinery Manufacturing	17	16	15	15	1	2	2
334	Computer and Electronic Product Manufacturing	9	8	8	6	1	1	3
335	Electrical Equipment, Appliances and Components	12	9	7	8	3	5	4
336	Transportation Equipment Manufacturing	18	15	14	15	3	4	3
337	Furniture and Related Product Manufacturing	10	10	8	9		2	1
339	Miscellaneous Manufacturing	7	6	6	7	1	1	1
	Total	259	216	216	208	43	43	52
	% of localized or dispersed		83.398	83.398	80.000	16.602	16.602	20.000

Notes: The measures are computed using the EG index at the CD level (NAICS 6-digit) unweighted (top panel) and weighted by inverse distance EGspat (bottom panel). Subsectors are identified at the 6-digit level. Blank cells indicate that there are no subsectors in the respective category (localized or dispersed or random).

with less than 50 full-time equivalent employees or having annual sales of less than $2 million – we consider a plant as being small if its size – as measured by the number of employees – is less than the industry median.[6] We repeat the same exercise to construct our young firms subsample. We consider a plant as being young if its age – measured since the year of its establishment – is less than the industry median. Our last subsample is for exporting plants. Here, we simply select all plants that report some exporting activity. Table 8 presents results for the spatially weighted and unweighted EG indices, respectively. A key difference with the EG indices we computed before is that we now take as benchmark the distribution of all firms in a particular industry. In other words, the question we are asking is the following: do specific subgroups of firms cluster more than the industry in general?

Not surprisingly, our results paint a very different picture from that of the baseline case. We find less concentration across both years and geographical scales.[7] At the census division level, less than 13 percent – around 34 industries – are found to be localized in terms of small firms, young firms, and exporters. Most of these industries are, however, strongly localized. We also find that more than 200 industries have dispersed patterns for these types of firms, i.e., more than 80 percent on average. These results are very different from those for industries as a whole. One needs, however, to keep in mind that the underlying benchmark is very different: conditional on the observed concentration of industries, small firms, young firms, and exporters do not generally concentrate more. Looking at the specific industries that underlie the foregoing figures, we find a very heterogeneous group of industries. The three industries with the most localized subgroups of firms in 2009, for example, are: for small firm: – Cold-Rolled Steel Shape Manufacturing (NAICS 331221), Railroad Rollling Stock Manufacturing (NAICS 336510), and Dog and Cat Food Manufacturing (NAICS 311111) –; for young firm: – Fur and Leather Clothing Manufacturing (NAICS 315292), Fibre, Yarn and Thread Mills (NAICS 313110), and Artificial and Synthetic Fibres and Filame (NAICS 325220) –; and for exporters: – Narrow Fabric Mills and Schiffli Machine (NAICS 313220), Other Snack Food Manufacturing (NAICS 311919), and Wet Corn Milling (NAICS 311221).

To summarize, only few sectors display patterns where small firms, young firms, or exporters are substantially more localized than the sector itself. In most cases, these subgroups of firms are more dispersed than the sector in general.

4 Continuous Measures: Methodology and Results

While discrete measures of localization, such as the EG index, are very popular and have been widely used – which is not the case of the EGspat index – they are known to be sensitive to

[6] Using a fixed employment threshold makes little sense, as the minimum operational scale varies widely across different industries.

[7] The only exception is the province level, but that level is rather coarse as spatial scale.

Table 8: Mean and Median EG Indices at Different Spatial Scales.

Geography	2001 PROV	2001 ER	2001 CD	2005 PROV	2005 ER	2005 CD	2009 PROV	2009 ER	2009 CD
Unweighted EG : NAICS 6-digit industries									
Results for small firms									
Mean	0.301	-0.332	-17.652	-0.518	-5.526	-0.314	0.286	-3.101	-0.727
Median	0.295	-1.010	-1.041	0.281	-1.012	-1.051	0.281	-1.007	-1.045
Share < 0	16.602	61.776	89.922	22.780	65.251	91.120	21.154	59.615	87.692
Share $\in (0, 0.05]$	0.772	2.703	0.388	0.772	1.931	0.772	0.769	3.462	1.923
Share > 0.05	82.625	35.521	9.690	76.448	32.819	8.108	78.077	36.923	10.385
Results for young firms									
Mean	0.061	-0.833	-1.984	-3.623	-0.626	-2.449	0.254	-0.695	-0.802
Median	0.364	-1.072	-1.122	0.328	-1.052	-1.158	0.336	-1.049	-1.140
Share < 0	20.463	67.829	88.372	23.166	62.162	88.417	23.462	62.308	86.538
Share $\in (0, 0.05]$	1.158	1.938	1.550	0.772	2.317	0.386	0.000	1.923	0.385
Share > 0.05	78.378	30.233	10.078	76.062	35.521	11.197	76.538	35.769	13.077
Results for exporters									
Mean	1.335	-1.060	-0.654	0.629	-1.210	-1.319	-0.452	-0.864	-1.112
Median	0.366	-1.061	-1.117	0.358	-1.078	-1.138	0.329	-1.066	-1.127
Share < 0	16.602	65.251	88.372	18.533	66.023	91.506	21.923	63.462	87.692
Share $\in (0, 0.05]$	0.772	2.703	0.775	1.158	3.089	1.544	1.154	1.923	0.385
Share > 0.05	82.625	32.046	10.853	80.309	30.888	6.950	76.923	34.615	11.923

Notes: Mean and median values for 259 (resp., 260 in 2009) NAICS 6-digit industries. Share < 0 means 'not clustered'. Share $\in (0, 0.05]$ means 'weakly clustered'. Share > 0.05 means 'strongly clustered'. See Ellison and Glaeser (1997) for details.

the choice of geographical units. To obtain sharper results on the geographical concentration of manufacturing industries in Canada, we therefore make now use of the micro-geographic nature of our data and compute continuous measures of localization, namely the Duranton-Overman index (Duranton and Overman, 2005, 2008).[8] This index is based on the kernel density of the distribution of bilateral distances across all plants in an industry – or, in its weighted version, of all employees in an industry – and compares that distribution to a counterfactual one that is obtained under the assumption of 'spatial randomness'. The key advantage of this index is that it retains the desirable properties of the EG index – namely to control for the size distribution of plants in an industry – while getting rid of the need to choose specific spatial units onto which to build the analysis.

To understand the logic underlying that index, we illustrate the possible patterns with the help of Figure 4. The observed distribution of distances in the industry is depicted by the red solid line, which we refer to as the K-density. The figure also depicts the 'local' (dashed black) and the 'global' (dashed blue) confidence bands (see the Appendix for details). These bands contain 90 percent of the counterfactual distributions, so that when the red line lies within them we cannot reject – at the 5 percent level – the null hypothesis that the location pattern

[8]The same methodology has been recently extended and can be applied to many economic problems where space matters and where micro-geographic data is available (see Murata, Nakajima, Okamoto, and Tamura, 2014, for an application to the localization of patents).

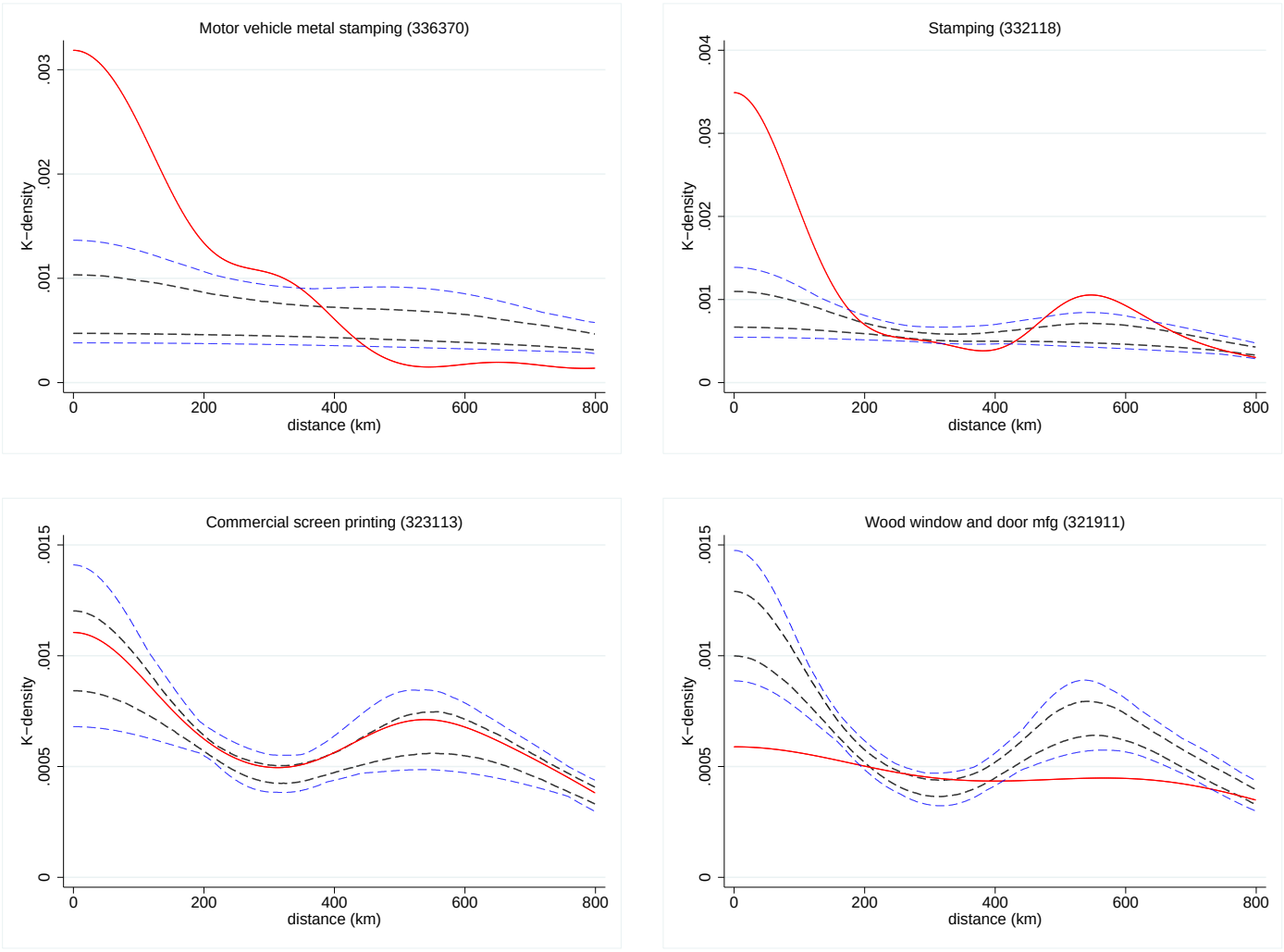

Figure 4: Selected Location Patterns of Industries in 2001 (Unweighted).

of the industry is one of 'spatial randomness'. If the red line lies above the upper bound of the confidence band, distances between firms are over-represented as compared to spatial randomness, which is interpreted as *localization*; whereas when the red line lies below the lower bound of the confidence band, distances between firms are under-represented as compared to spatial randomness, which is interpreted as *dispersion*.

The four industries depicted in Figure 4 display four different geographical patterns. The top-left panel depicts an industry that is localized at a regional scale (up to 200 kilometers), however dispersed at longer distances. This corresponds to the 'classical' location pattern where firms are disproportionately located at short distances, i.e., the industry is localized. The top-right panel depicts an industry that is both significantly concentrated at short distances, and also significantly agglomerated in major urban areas – 500 kilometers corresponds approximately to the distance between the metropolitan regions of Toronto and Montreal. The bottom-left panel depicts an industry that is neither significantly localized nor significantly dis-

persed. The location pattern of that industry is not significantly different from one that would be obtained by a purely random location process. Last, the bottom-right panel depicts an industry that is significantly dispersed, both at short distances and across major urban centers.

4.1 Baseline Results

We first examine the number of industries that are localized or dispersed according to the DO index. As can be seen from Table 9, using our strict definition of manufacturing firms (see the Appendix), we find that roughly 31 percent and 55 percent of industries were significantly localized in 2001 at the 6-digit and the 4-digit levels, respectively. These numbers were quite stable between 2001 and 2005, but they fall below 25 percent at the 6-digit level and below 49 percent at the 4-digit level in 2009. On average, the fraction of localized manufacturing industries in Canada is slightly smaller than the ones reported for the U.K. (52 percent), France (63 percent), Germany (71 percent), and Japan (50 percent) in earlier studies by Duranton and Overman (2005), Barlet, Briant, and Crusson (2013), Riedel and Hyun-Ju (2012), and Nakajima, Saito, and Uesugi (2012), respectively.

In line with our previous findings based on the EG indices, there is a tendency towards less localization between 2001 and 2009: the number of localized industries decreases, as well as the strength of localization as measured by the average Γ across all localized sectors. This trend affects both the 4- and the 6-digit industries, with and without employment weights. Although industries tend to display less localization when using the employment-weighted K-densities than in the unweighted case, the key results remain very similar. It is worth noting that the number of industries that do not significantly depart from randomness is quite large in our samples – around 59 percent in 2009 – which may be due to either the fine level of sectoral disaggregation, or to the presence of a large number of small firms in our samples, or to the specific structure of the Canadian economy.[9] Table 9 summarizes our results for the three years and for the different sample definitions – strict vs extended – and weighting schemes – unweighted vs weighted – and industry aggregation levels – 6-digit vs 4-digit.

Tables 17–19 in the Appendix list the ten most and the ten least localized industries – both at the 4- and at the 6-digit levels – for 2001, 2005, and 2009, respectively. Since the raw value of the DO index is hard to interpret, we report results using the cumulative distribution function (CDF) associated with the K-density, evaluated at a distance of 50 kilometers.[10] These results are summarize in Tables 10 and 11 below. Consider, e.g., 'Knit Fabric Mills' (NAICS 313240) in 2001. As can be seen from Table 10, the CDF at a distance of 50 kilometers is 0.4173. In words,

[9]Previous studies for the UK, France, Germany, or Japan, focus on 'compact countries', whereas Canada is geographically all but 'compact'.

[10]Note that the DO index and the CDF do not provide the same set of information. Yet, as can be seen from Tables 17–19, 10, and 11, industries with large values of the DO index also have a high CDF. This is because for most industries that are strongly localized, that localization occurs at quite short distances (below 50 kilometers).

Table 9: Summary Statistics for K-density Estimates.

4-digit industries

	2001, unweighted				2001, weighted			
	Strict		Extended		Strict		Extended	
	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
localized	47	54.651	52	60.465	47	54.651	49	56.977
random	26	30.233	20	23.256	29	33.721	29	33.721
dispersed	13	15.116	14	16.279	10	11.623	8	9.302
$\overline{\Gamma}\|_{\Gamma_i>0}$	0.051		0.050		0.057		0	0.059
$\overline{\Psi}\|_{\Psi_i>0}$	0.024		0.026		0.018		0.027	

	2005, unweighted				2005, weighted			
	Strict		Extended		Strict		Extended	
	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
localized	48	55.814	50	58.140	42	48.837	46	53.488
random	24	27.907	17	19.767	33	38.372	30	34.884
dispersed	14	16.279	19	22.093	11	12.791	10	11.628
$\overline{\Gamma}\|_{\Gamma_i>0}$	0.043		0.038		0.050		0.045	
$\overline{\Psi}\|_{\Psi_i>0}$	0.027		0.027		0.020		0.023	

	2009, unweighted				2009, weighted			
	Strict		Extended		Strict		Extended	
	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
localized	42	48.837	47	54.651	34	39.535	36	41.860
random	29	33.721	23	26.744	39	45.349	40	46.512
dispersed	15	17.442	16	18.605	13	15.116	10	11.628
$\overline{\Gamma}\|_{\Gamma_i>0}$	0.039		0.035		0.044		0.036	
$\overline{\Psi}\|_{\Psi_i>0}$	0.029		0.028		0.017		0.030	

6-digit industries

	2001, unweighted				2001, weighted			
	Strict		Extended		Strict		Extended	
	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
localized	79	30.620	100	38.610	88	34.109	105	40.541
random	153	59.302	120	46.332	157	60.853	132	50.965
dispersed	26	10.078	39	15.058	13	5.039	22	8.494
$\overline{\Gamma}\|_{\Gamma_i>0}$	0.082		0.062		0.072		0.059	
$\overline{\Psi}\|_{\Psi_i>0}$	0.018		0.018		0.008		0.016	

	2005, unweighted				2005, weighted			
	Strict		Extended		Strict		Extended	
	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
localized	78	30.116	105	40.541	69	26.641	96	37.066
random	150	57.915	108	41.699	170	65.637	139	53.668
dispersed	31	11.969	46	17.761	20	7.722	24	9.266
$\overline{\Gamma}\|_{\Gamma_i>0}$	0.069		0.044		0.085		0.047	
$\overline{\Psi}\|_{\Psi_i>0}$	0.016		0.019		0.012		0.014	

	2009, unweighted				2009, weighted			
	Strict		Extended		Strict		Extended	
	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
localized	64	24.710	94	36.293	62	23.938	80	30.888
random	163	62.934	120	46.332	180	69.498	148	57.143
dispersed	32	12.355	45	17.375	17	6.564	31	11.969
$\overline{\Gamma}\|_{\Gamma_i>0}$	0.071		0.044		0.077		0.047	
$\overline{\Psi}\|_{\Psi_i>0}$	0.016		0.018		0.012		0.012	

Notes: See Appendix C for details on how to compute Γ_i and Ψ_i. We denote their arithmetic average by $\overline{\Gamma}$ and $\overline{\Psi}$.

41.73 percent of firm pairs are located less than 50 kilometers apart in that sector. Alternatively, we can view this as the probability that two randomly drawn firms from that industry are less than 50 kilometers away from each other. Clearly, more than two chances in five is a very high value given the geographical extent of manufacturing in Canada. As can be seen from Tables 10 and 11 (and also from Tables 17–19 in the Appendix), various textile and metal-related sectors rank among the most strongly localized industries in the different years. These results largely concur with those obtained using the EG indices.

Table 10: Ten Most Localized Industries According to the DO CDF (Unweighted).

NAICS6	Industry name	CDF
	2001	
315231	Women's and Girls' Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing	0.471
313240	Knit Fabric Mills	0.417
315210	Cut and Sew Clothing Contracting	0.258
315292	Fur and Leather Clothing Manufacturing	0.234
333220	Rubber and Plastics Industry Machinery Manufacturing	0.206
333519	Other Metalworking Machinery Manufacturing	0.204
336110	Automobile and Light-Duty Motor Vehicle Manufacturing	0.178
325991	Custom Compounding of Purchased Resins	0.175
332118	Stamping	0.170
336370	Motor Vehicle Metal Stamping	0.159
NAICS6	Industry name	CDF
	2005	
315231	Women's and Girls' Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing	0.536
333220	Rubber and Plastics Industry Machinery Manufacturing	0.369
332118	Stamping	0.237
336110	Automobile and Light-Duty Motor Vehicle Manufacturing	0.230
312210	Tobacco Stemming and Redrying	0.200
315292	Fur and Leather Clothing Manufacturing	0.188
333519	Other Metalworking Machinery Manufacturing	0.188
336370	Motor Vehicle Metal Stamping	0.168
325991	Custom Compounding of Purchased Resins	0.166
315110	Hosiery and Sock Mills	0.158
NAICS6	Industry name	CDF
	2009	
315231	Women's and Girls' Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing	0.513
312210	Tobacco Stemming and Redrying	0.282
333220	Rubber and Plastics Industry Machinery Manufacturing	0.256
332991	Ball and Roller Bearing Manufacturing	0.252
336110	Automobile and Light-Duty Motor Vehicle Manufacturing	0.241
336370	Motor Vehicle Metal Stamping	0.228
315292	Fur and Leather Clothing Manufacturing	0.186
333519	Other Metalworking Machinery Manufacturing	0.180
332118	Stamping	0.180
332720	Turned Product and Screw, Nut and Bolt Manufacturing	0.151

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.

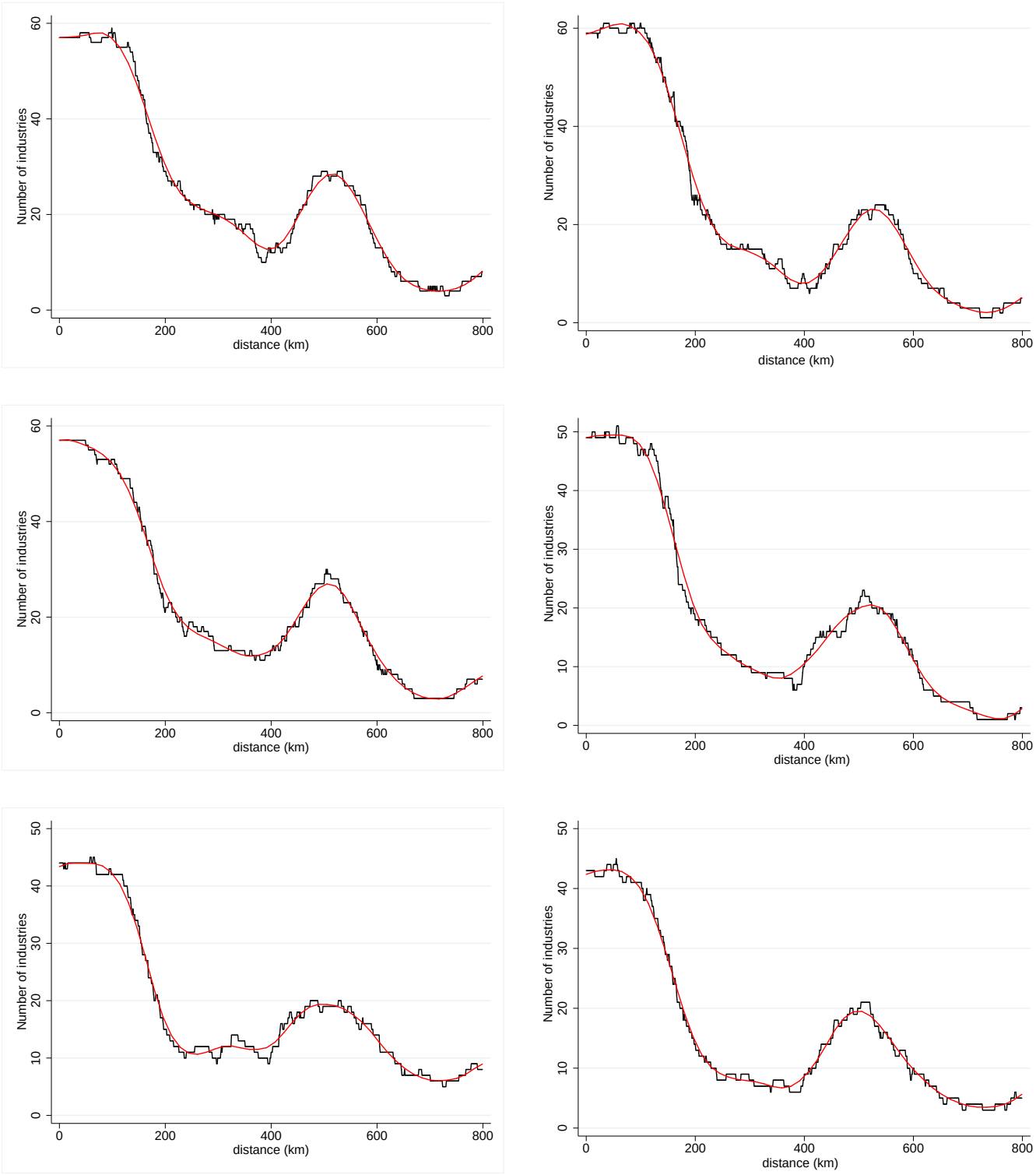

Figure 5: Global Localization, Unweighted (Left Panel) and Weighted (Right Panel) K-densities for 2001 (Top), 2005 (Middle), and 2009 (Bottom).

Table 11: Ten Most Localized Industries According to the DO CDF (Weighted).

NAICS 6	Industry name	CDF
	2001	
325110	Petrochemical Manufacturing	0.344
313240	Knit Fabric Mills	0.309
333220	Rubber and Plastics Industry Machinery Manufacturing	0.254
315231	Women's and Girls' Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing	0.247
336370	Motor Vehicle Metal Stamping	0.216
315110	Hosiery and Sock Mills	0.207
332118	Stamping	0.199
333519	Other Metalworking Machinery Manufacturing	0.169
336110	Automobile and Light-Duty Motor Vehicle Manufacturing	0.166
315233	Women's and Girls' Cut and Sew Dress Manufacturing	0.166
NAICS 6	Industry name	CDF
	2005	
333220	Rubber and Plastics Industry Machinery Manufacturing	0.277
312210	Tobacco Stemming and Redrying	0.241
336370	Motor Vehicle Metal Stamping	0.192
313240	Knit Fabric Mills	0.179
336110	Automobile and Light-Duty Motor Vehicle Manufacturing	0.169
332118	Stamping	0.162
315210	Cut and Sew Clothing Contracting	0.157
315231	Women's and Girls' Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing	0.157
333519	Other Metalworking Machinery Manufacturing	0.156
333511	Industrial Mould Manufacturing	0.155
NAICS 6	Industry name	CDF
	2009	
315231	Women's and Girls' Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing	0.459
312210	Tobacco Stemming and Redrying	0.249
336110	Automobile and Light-Duty Motor Vehicle Manufacturing	0.209
336370	Motor Vehicle Metal Stamping	0.207
333220	Rubber and Plastics Industry Machinery Manufacturing	0.188
332118	Stamping	0.158
333519	Other Metalworking Machinery Manufacturing	0.156
333511	Industrial Mould Manufacturing	0.142
332991	Ball and Roller Bearing Manufacturing	0.135
325520	Adhesive Manufacturing	0.132

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.

One advantage of the continuous measures is that they allow to finely assess at *what distances* localization or dispersion actually occur. The top panel of Figure 5 depicts the number of 6-digit industries that are globally localized at each distance between 1 and 800 kilometers, both in the unweighted (left panel) and the weighted (right panel) case in 2001. As one can see, most industries are localized at relatively short distances (up to 150–180 kilometers) or at intermediate distances (about 500 kilometers). The reason is that some industries cluster predominantly in an urban environment – short distances, or distances of about 500 kilometers

between major urban centers – whereas other industries cluster in more rural and semi-rural areas between major urban centers (about 200–400 kilometers). These industries are then naturally underrepresented at short distances, because dispersion at some distances is the flip-side of agglomeration at other distances. Observe also that: (i) less industries are localized in 2009 than in 2001, especially at short distances and at intermediate inter-city distances; and (ii) this trend is stronger in the unweighted case, thereby suggesting that the change in the pattern is driven by smaller firms that either disappear or change location.

Last, Figure 6 plots the rank-ordered distribution of the Γ_i (blue line) and the Ψ_i (red line) measures of localization and dispersion. As one can see, there is a limited number of highly localized or dispersed industries. Furthermore, most of the industries do not have extreme values, which is similar to results for the UK and Japan. One can also see that the number of localized industries decreases over time, both in the unweighted and in the weighted case, whereas there is not much change in the degree and strength of dispersion, as well as in the number of dispersed industries. Last, some of the most strongly localized industries tend to get even more strongly localized. These findings suggest an interesting insight: over the 2001–2009 period, manufacturing industries got generally less localized in Canada, but localization increased at the very top of the distribution. The general trend of spatial deconcentration thus does not affect all industries in the same way.

4.2 Sectoral Scope of Localization

We next look at the differences in the degree of localization across broad industry categories. As can be seen from Table 9, there is generally more localization at the 4-digit level than at the 6-digit level. Another way to see that there are broad industry location patterns at a more aggregate level is to look at the 21 3-digit industries and to compute the ratio of localized 6-digit industries in the total number of 6-digit industries that make up a particular 3-digit industry. The results of this exercise are summarized in Table 12. As can be seen from that table, 6-digit industries belonging to the 3-digit industries 313 ('Textile Mills'), 315 ('Clothing Manufacturing'), 323 ('Printing and Related Support Activities'), 333 ('Machinery Manufacturing'), and 334 ('Computer and Electronic Product Manufacturing') are made up of subindustries that display strong localization patterns. On the contrary, 6-digit sectors belonging to industries 324 ('Petroleum and Coal Products'), 312 ('Beverage and Tobacco'), and 321 ('Wood products') display only very weak patterns of localization. These findings are similar to those for the UK, where textile (SIC 17-19) and publishing (SIC 22) industries are among the most localized industries, while food and drink (SIC 15), wood (SIC 20), and petroleum (SIC 23) industries are among the least localized industries (see Duranton and Overman, 2005). The pattern is also similar to that observed in Japan by Nakajima, Saito, and Uesugi (2012), where the most localized industries are related to 'Textile Mill Products' (JSIC 11), 'Electrical Machinery' (JSIC 27),

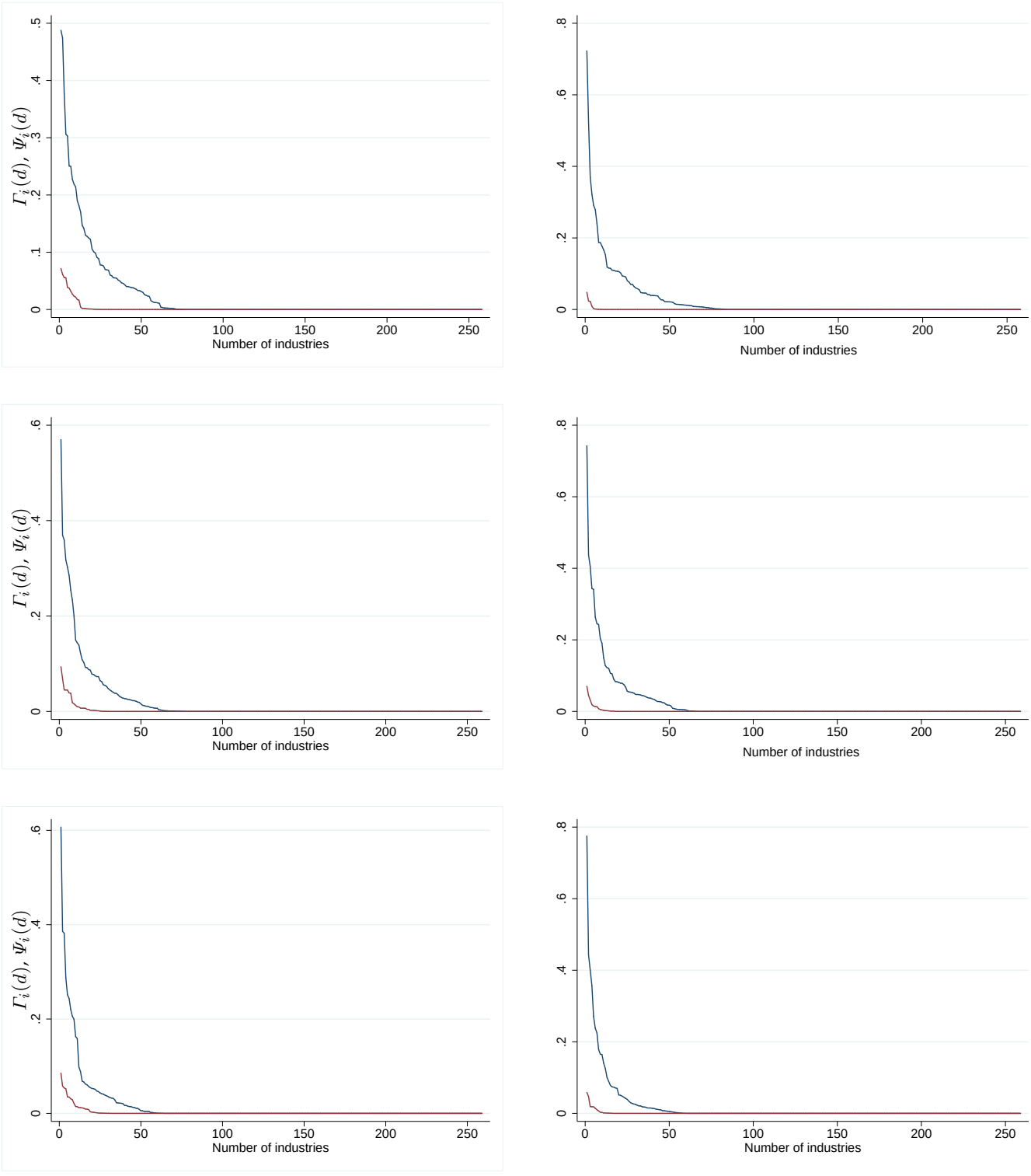

Figure 6: Rank-order Distribution of Γ_i and Ψ_i for each Industry, Unweighted (Left Panel) and Weighted (Right Panel) K-densities for 2001 (Top), 2005 (Middle), and 2009 (Bottom).

whereas the least localized are related to 'Petroleum and Coal Products' (JSIC 18), and 'Lumber and Wood Products' (JSIC 13).

4.3 Location Patterns of Small Firms, Young Firms, and Exporters

We now use the continuous approach to analyze the location patterns of small firms, young firms, and exporters (see also Duranton and Overman, 2008). Do small firms, young firms, or exporters locate closer to each other than establishments in the industry in general? Table 13 summarizes our results. Across years, we find that only 7 to 11 industries (3 to 4 percent) exhibit localization of small firms, whereas 13 to 19 industries (5 to 7 percent) exhibit dispersion of small firms. This leaves more than 90 percent of industries with location patterns of small firms that do not differ significantly from randomness. These findings suggest that small establishments in an industry do not locate differently than its establishments in general. This weak tendency for clustering of small establishments is consistent with Duranton and Overman's (2008) findings for the UK. We obtain very similar results for young firms, as can be seen from Table 13. Turning to exporters, these firms exhibit somewhat more localization. There are indeed 36 to 41 industries (14 to 16 percent) that exhibit localization of exporters, whereas only 11 to 28 (4 to 11 percent) exhibit dispersion of exporters. Even though these figures are larger than for small firms and young firms, three-quarter of industries display no clear pattern with respect to the geographical distribution of their exporters. In a nutshell, there is only little evidence that small firms, young firms, or exporters display 'excess concentration' with respect to their industries.

One may worry that our finding that many industries seem to display random patterns is driven by small sample sizes. To check the robustness of our results, we thus restrict our industries conservatively to subsamples with at least 25 firms and run our estimations again. Doing so leaves us with 170 to 190 industries – depending on the year and the subsample. As one can see from the right part of Table 13, the results are similar, thus suggesting that they are not biased because of sectors with small sample sizes.

When looking at the specific industries that underlie the foregoing figures, we find again a very heterogeneous group of industries. The three industries with the most localized subgroups of firms in 2009, for example, are: for small firm: – All Other Plastic Product Manufacturing (NAICS 326198), Other Motor Vehicle Parts Manufacturing (NAICS 336390), and Coating, Engraving, Heat Treating and Allied Activities (NAICS 332810) –; for young firm: – Pottery, Ceramics and Plumbing Fixture Manufacturing (NAICS 327110), All Other Industrial Machinery Manufacturing (NAICS 333299), and All Other Plastic Product Manufacturing (NAICS 326198) –; and for exporters: – Sawmills (except Shingle and shake Mills) (NAICS 321111), Prefabricated Wood Building Manufacturing (NAICS 321992), and Other Animal Food Manufacturing (NAICS 311119).

Table 12: Localization Patterns by Broad Industry Groups.

NAICS3	Industry name	#subsectors	#localized	#random	#dispersed	% localized
	Unweighted K-density estimates					
311	Food Manufacturing	32	3	26	3	9.375
312	Beverage and Tobacco Product Manufacturing	6	0	5	1	0.000
313	Textile Mills	7	3	4	0	42.857
314	Textile Product Mills	4	0	3	1	0.000
315	Clothing Manufacturing	17	13	4	0	76.471
316	Leather and Allied Product Manufacturing	3	1	2	0	33.333
321	Wood Product Manufacturing	14	4	6	4	28.571
322	Paper Manufacturing	12	3	8	1	25.000
323	Printing and Related Support Activities	6	3	3	0	50.000
324	Petroleum and Coal Products Manufacturing	4	0	4	0	0.000
325	Chemical Manufacturing	20	7	12	1	35.000
326	Plastics and Rubber Products Manufacturing	14	4	10	0	28.571
327	Non-Metallic Mineral Product Manufacturing	12	1	9	2	8.333
331	Primary Metal Manufacturing	13	3	10	0	23.077
332	Fabricated Metal Product Manufacturing	21	8	12	1	38.095
333	Machinery Manufacturing	17	9	5	3	52.941
334	Computer and Electronic Product Manufacturing	9	5	2	2	55.556
335	Electrical Equipment, Appliance and Component Manufacturing	12	2	10	0	16.667
336	Transportation Equipment Manufacturing	18	3	9	6	16.667
337	Furniture and Related Product Manufacturing	10	4	6	0	40.000
339	Miscellaneous Manufacturing	7	3	3	1	42.857
	Weighted K-density estimates					
311	Food Manufacturing	32	3	28	1	9.375
312	Beverage and Tobacco Product Manufacturing	6	0	6	0	0.000
313	Textile Mills	7	6	1	0	85.714
314	Textile Product Mills	4	1	3	0	75.000
315	Clothing Manufacturing	17	16	0	1	94.118
316	Leather and Allied Product Manufacturing	3	1	2	0	33.333
321	Wood Product Manufacturing	14	2	5	7	14.286
322	Paper Manufacturing	12	7	4	1	58.333
323	Printing and Related Support Activities	6	0	3	1	0.000
324	Petroleum and Coal Products Manufacturing	4	0	4	0	0.000
325	Chemical Manufacturing	20	10	10	0	50.000
326	Plastics and Rubber Products Manufacturing	14	7	7	0	50.000
327	Non-Metallic Mineral Product Manufacturing	12	1	9	2	8.333
331	Primary Metal Manufacturing	13	3	10	0	23.077
332	Fabricated Metal Product Manufacturing	21	9	12	9	42.857
333	Machinery Manufacturing	17	9	5	3	52.941
334	Computer and Electronic Product Manufacturing	9	8	1	0	88.889
335	Electrical Equipment, Appliance and Component Manufacturing	12	6	6	0	50.000
336	Transportation Equipment Manufacturing	18	8	9	1	44.444
337	Furniture and Related Product Manufacturing	10	2	7	1	20.000
339	Miscellaneous Manufacturing	7	3	2	2	42.857

Notes: The year is 2001 and the measures are computed using the unweighted K-densities (top panel) and the employment-weighted K-densities (bottom panel). Subsectors are identified at the 6-digit level.

Table 13: Summary Statistics for Small, Young, and Exporter Subsamples.

	Small firms		Young firms		Exporters		Small firms		Young firms		Exporters	
	2001, all 6-digit industries						2001, restricted 6-digit industries					
Status	Number	%	Number	%	Number	%	Number	%	Number	%	Number	%
localized	10	3.891	16	6.226	41	15.953	9	4.945	11	6.077	37	19.271
random	228	88.716	239	92.996	205	79.767	153	84.066	168	92.818	146	76.042
dispersed	19	7.393	2	0.778	11	4.280	20	10.989	2	1.105	9	4.688
$\overline{\Gamma}\|_{\Gamma_i>0}$	0.021		0.003		0.023		0.021		0.004		0.025	
$\overline{\Psi}\|_{\Psi_i>0}$	0.008		0.006		0.006		0.007		0.006		0.007	
	2005, all 6-digit industries						2005, restricted 6-digit industries					
Status	Number	%	Number	%	Number	%	Number	%	Number	%	Number	%
localized	11	4.264	8	3.113	36	13.900	10	5.464	4	2.210	30	15.306
random	232	89.922	242	94.163	195	75.290	158	86.339	171	94.475	141	71.939
dispersed	15	5.814	7	2.724	28	10.811	15	8.197	6	3.315	25	12.755
$\overline{\Gamma}\|_{\Gamma_i>0}$	0.006		0.062		0.013		0.007		0.123		0.015	
$\overline{\Psi}\|_{\Psi_i>0}$	0.006		0.003		0.002		0.006		0.003		0.002	
	2009, all 6-digit industries						2009, restricted 6-digit industries					
Status	Number	%	Number	%	Number	%	Number	%	Number	%	Number	%
localized	7	2.713	11	4.280	37	14.341	6	3.550	3	1.775	29	15.847
random	238	92.248	238	92.607	198	76.744	152	89.941	159	94.083	133	72.678
dispersed	13	5.039	8	3.113	23	8.915	11	6.509	7	4.142	21	11.475
$\overline{\Gamma}\|_{\Gamma_i>0}$	0.002		0.008		0.020		0.002		0.011		0.026	
$\overline{\Psi}\|_{\Psi_i>0}$	0.005		0.002		0.002		0.006		0.002		0.003	

Notes: See the Appendix for details on how to compute Γ_i and Ψ_i. We denote their arithmetic average by $\overline{\Gamma}$ and $\overline{\Psi}$. The restricted industries case includes only industries with samples of more than 25 firms of that specific type.

We next depict in Figures 7 and 8 the global localisation (left panel) and the rank-order distribution of localized and dispersed industries (right panel) for small firms, young firms, and exporters in 2001 and 2009, respectively. Despite some differences – especially for small and young firms – the general shape of these graphs is similar to the baseline case: most industries are localized at relatively short or at intermediate distances. The number of industries in these ranges is far smaller than the number in the baseline results. The number of dispersed industries (not shown here) is increasing over the entire range of distances between 0 and 800 kilometers. It is also increasing across years. This mirrors our general finding that industries – and specific subgroups of firms – have a tendency to geographically disperse in Canada over the first decade of 2000.

Last, as one can see from Figure 7, the rank-order distributions of localized and dispersed industries are quite similar to those in the baseline case. The number of localized industries decreases betwen 2001 and 2009, and there is not much change in the degree of dispersion. Not surprinsingly, we have more localized industries in the exporters subsample. Clearly, a few industries only are highly localized, which mirrors our findings using the EG index.

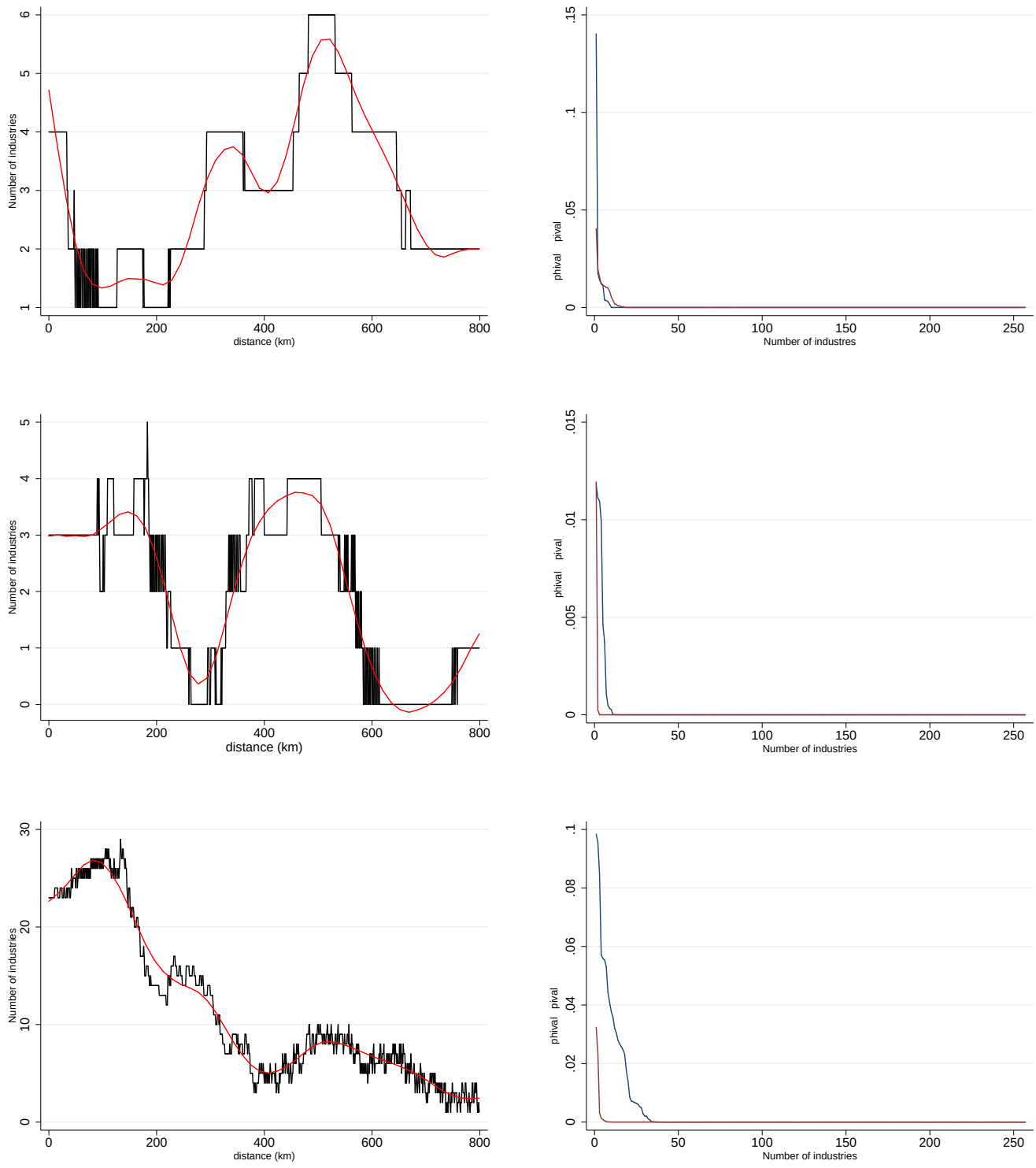

Figure 7: Unweighted Global Localization (Left Panel) and Rank-order Distribution of Γ_i and Ψ_i (Right Panel) K-densities for Small Firms (Top), Young Firms (Middle), and Exporters (Bottom) in 2001.

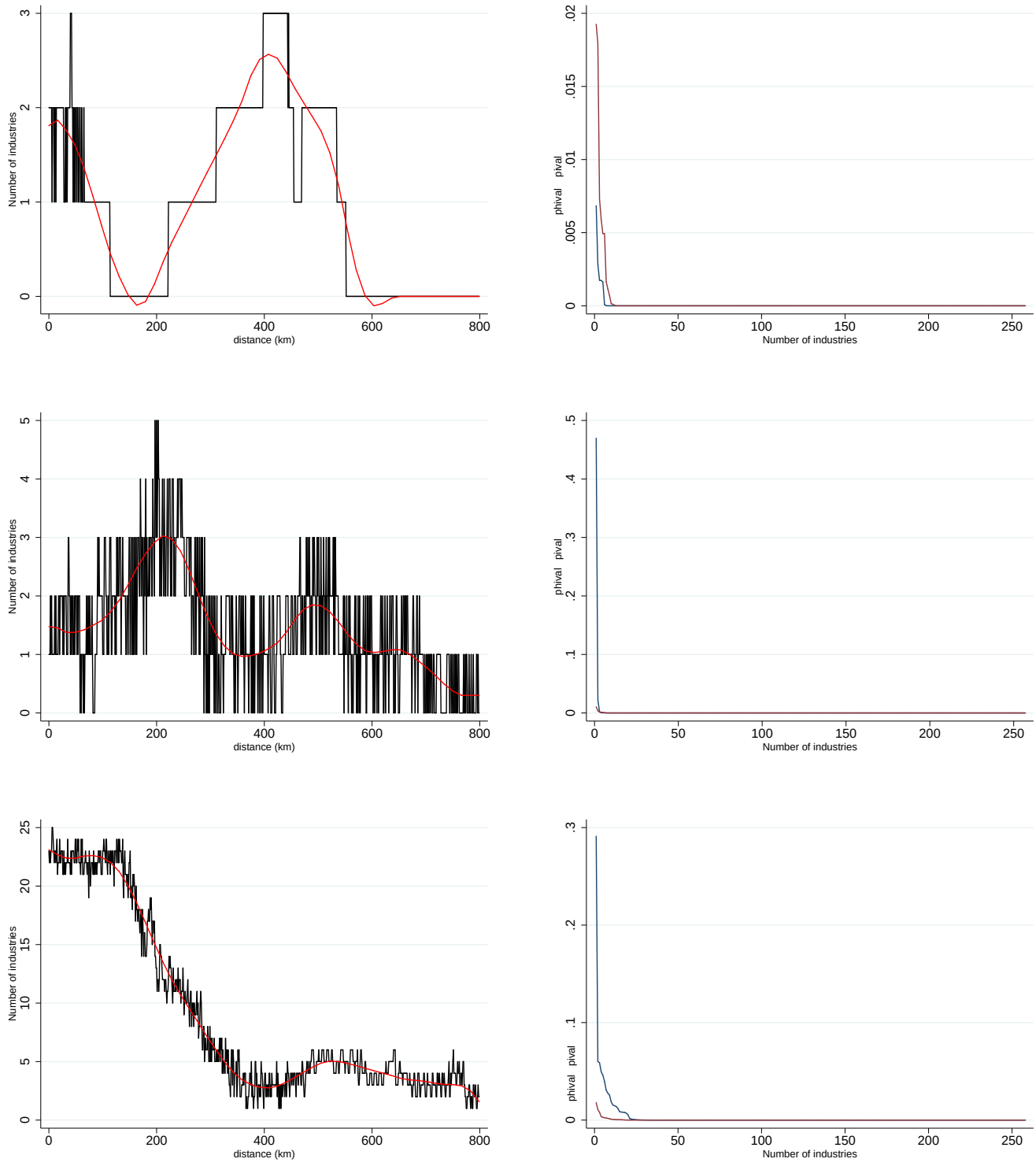

Figure 8: Unweighted Global Localization (Left Panel) and Rank-order Distribution of Γ_i and Ψ_i (Right Panel) K-densities for Small Firms (Top), Young Firms (Middle), and Exporters (Bottom) in 2009.

5 Concluding Remarks

We have used extensive micro-geographic data to provide what we believe is to date the most comprehensive anatomy of the geographical concentration of manufacturing industries in Canada. Looking at the changes between 2001 and 2009 allowed us also to examine the 'dynamics' of localization in a detailed way. The following key results stand out. First, depending on industry definitions and years, 40 to 60 percent of manufacturing industries are clustered, mainly at short distances and at distances of about 500 kilometers. This finding suggests that there is less industrial localization in Canada than in other developed countries. Second, according to all measures we computed – discrete, spatially weighted discrete, and continuous – localization has been decreasing from 2001 to 2009. Third, industries related to textiles and the extraction and processing of natural resources dominate the group of the most localized industries. This finding is in accord with previous results for other countries in the literature. Fourth, while there has been a general trend towards less concentration, some of the most strongly localized industries tend to get even more localized. Last, small firms and young firms are, in general, not more strongly concentrated than all firms in their respective industries – there is little evidence that these firms obey a location logic that is different from that of their industry in general. There is some evidence for 'excess concentration' of exporters, but that effect tends to weaken over the first decade of 2000.

Despite the depth of our analysis, two issues remain unresolved. First, our paper remains silent on the causes for localization. Yet, we need to better understand what agglomeration forces contribute to the clustering of Canadian manufacturing industries. Previous studies – such as Rosenthal and Strange (2001), and Ellison, Glaeser, and Kerr (2010) – have addressed that question for the US. Disentangling the relative contribution of the different sources of agglomeration in Canada – labor market pooling, input-output linkages, transportation costs, and knowledge spillovers – is logically the next item on our agenda. More research is called for here. Second, although continuous measures of localization obviate the need for using rather arbitrary spatial subdivisions, they still do rely on equally arbitrary subdivisions of industries. As shown by our analysis, the results do somewhat depend on industrial classifications. Hence, extending our measures to analyze location patterns in terms of 'firm similarity', like similarity in terms of labor requirements or in terms of input-output structures, seems a necessary step for deriving more robust results on agglomeration patterns and may provide valuable insights into what is driving agglomeration more generally. We leave this very important question again open for future work.

Finally, our findings also have a number of implications for 'cluster policy' and 'regional development'. As countries and regions strive to remain competitive in the face of globalization, governments – both local and national – seek increasingly to support competitive regional clusters – see, e.g., Canada's 'National Research Council Cluster Initiatives', the French 'Pôles

de compétitivité' Program, and the German 'BioRegio' Program. The 2007 OECD report on 'National Policy Approaches to Cluster Strategies' highlights the increasing focus on building strategic research capacity in selected regions as the basis for promoting clusters. Recent economic studies, however, increasingly question the use of cluster policies (e.g., Duranton, 2011; Duranton, Martin, Mayer, and Mayneris, 2012; Behrens, 2013). There is indeed little evidence that more clustering will have significant effects on average productivity or wages in manufacturing industries. Given the weak impacts compared to the costs of these initiatives, building clusters might not provide strong and cost-efficient solutions for regional development. This seems all the more true for Canada, where the trend over the first decade of 2000 is towards less industrial localization, and where there is only little evidence that targeting small firms, young firms, or exporters may give rise to clusters that could help to foster regional development and growth.

Acknowledgements. We thank Mark Brown for very helpful comments and suggestions. Katherine Pineault provided excellent research assistance. Financial support 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. We also gratefully acknowledge financial support from FQRSC Québec (Grant NP-127178), and from CIRPÉE. The views expressed in this paper, as well as all remaining errors, are ours.

References

[1] Axtell, Robert L. (2001) Zipf Distribution of US Firm Sizes, *Science* 293, 1818–1820.

[2] Barlet, Muriel, Briant, Anthony, and Crusson, Laure (2013) Location patterns of service industries in France: A distance-based approach, *Regional Science and Urban Economics* 43**(2)**, 338–351.

[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] Bernard, A.B., and J.B. Jensen (1995) Exporters, Jobs, and Wages in U.S. Manufacturing: 1976-1987, *Brookings Papers on Economic Activity, Microeconomics* 1995, 67–119.

[6] Cabral, Luís M.B., and José Mata (2003) On the Evolution of the Firm Size Distribution: Facts and Theory, *American Economic Review* 93**(4)**, 1075–1090.

[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 (2011) California Dreamin': The Feeble Case for Cluster Policies, *Review of Economic Analysis* 3, 3–45.

[10] Duranton, Gilles, Philippe Martin, Thierry Mayer and Florian Mayneris (2011) *The Economics of Clusters: Lessons from The French Experience*. Oxford Univ. Press and CEPREMAP.

[11] Duranton, Gilles, and Henry G. Overman (2005) Testing for Localisation Using Micro-Geographic Data, *Review of Economic Studies* 72**(4)**, 1077–1106.

[12] Duranton, Gilles, and Henry G. Overman (2008) Exploring the Detailed Location Patterns of UK Manufacturing Industries Using Microgeographic Data, *Journal of Regional Science* 48**(1)**, 213–243.

[13] 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.

[14] 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.

[15] Guimaraes, Paulo, Octavio Figueiredo, and Douglas Woodward (2011) Accounting for Neighboring Effects in Measures of Spatial Concentration, *Journal of Regional Science* 51**(4)**, 753–774.

[16] 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.

[17] Krugman, Paul R. (1991) Increasing Returns and Economic Geography, *Journal of Political Economy* 99**(3)**, 483–499.

[18] Marcon, Eric, and Florence Puech (2003) Evaluating the Geographic Concentration of Industries using Distance-based Methods, *Journal of Economic Geography* 3**(4)**, 678–693.

[19] Maurel, Françoise, and Beatrice Sedillot (1999) A Measure of the Geographic Concentration in French Manufacturing Industries *Regional Science and Urban Economics* 29**(5)**, 575–604.

[20] Murata, Y., R. Nakajima, R. Okamoto, and R. Tamura (2014) Localized knowledge spillovers and patent citations: A distance-based approach, *Review of Economics and Statistics*, forthcoming.

[21] Nakajima, Kentaro, Yukiko Umeno Saito, and Iichiro Uesugi (2012) Measuring economic localization: Evidence from Japanese firm-level data, *Journal of the Japanese and International Economies* 26**(2)**, 201–220.

[22] Openshaw, Stan (1983) *The Modifiable Areal Unit Problem*. Geo Books: Norwick, Norfolk, UK.

[23] Openshaw, Stan, and Peter Taylor (1979) A Million or so Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem. In: Neil Wrigley (ed.), *Statistical Applications in the Spatial Sciences*. Pion, London: pp. 127–144.

[24] Riedel, Nadine, and Hyun-Ju, Koh (2012) Assessing the Localization Pattern of German Manufacturing and Service Industries. Forthcoming in *Regional Studies*, DOI: 10.1080/00343404.2012.677024.

[25] Ripley, Brian David (1976) The Second-Order Analysis of Stationary Point Processes, *Journal of Applied Probability* 13**(2)**, 255–266.

[26] Ripley, Brian David (1977) Modelling Spatial Patterns, *Journal of the Royal Statistical Society B* 39**(2)**, 172–212.

[27] 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.

[28] Rosenthal, Stuart S., and William C. Strange (2003) Geography, industrial organization, and agglomeration, *Review of Economics and Statistics* 85**(2)**, 377–393.

[29] 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 briefly presents the Ellison-Glaeser and the spatially weighted Ellison-Glaeser indices. Appendix C provides details on the Duranton-Overman K-density approach. 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 *Scott's National All Business Directories Database*. This establishment-level database contains information on firms operating in Canada, with an extensive coverage of the manufacturing sector. It comprises 54,379 plants in 2001, 50,404 in 2005, and 46,391 in 2009 (see Table 14 for a breakdown by province). Our data cover the years 2001, 2005, and 2009, and our focus is on manufacturing firms only. For every etablishment, 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.

The Scott's database constitutes probably the best alternative to Statistics Canada's proprietary *Annual Survey of Manufacturers Longitudinal Microdata File*. As can be seen from Table 14, which provides a comparison of the Scott's National All 2001, 2005, and 2009 databases with Statistics Canada's province-level data from the 2003 and 2005 Annual Survey of Manufacturers (ASM; CANSIM Tables 301–0003 and 301–0006), it has a wide and similar coverage. Observe that the 2003 ASM covers only incorporated businesses with at least one employee and sales of manufactures in excess of C$30,000, whereas the 2005 data no longer has a minimum sales criterion and includes forestry. As can be seen from Table 14 by comparing either our 2001 or 2005 data to the ASM, the coverage of manufacturing firms in the Scott's data is very good, though large firms in the economic core provinces (Ontario, Québec, British Columbia, and Alberta) seem slightly under-represented, whereas small firms in the other provinces seem slightly over-represented. Even when compared to the 2005 ASM – though not directly comparable – the coverage of our database remains good (about 60 percent of the firms), though firms in western Canada seem slightly under-represented. To summarize, our data is very similar to that of the ASM in terms of coverage and province-level breakdown of firms and should, therefore, provide a fairly accurate picture of the overall manufacturing structure in Canada.

We consider that a firm is a manufacturer in the strict sense if it reports a manufacturing sector (NAICS 31–33) as its primary sector of activity. Since firms in our dataset also report up to four secondary NAICS codes, we can construct two different industry-level samples for the analysis: (i) a *strict sample*, restricted to firms that report a manufacturing sector as their primary sector of activity; and (ii) an *extended sample* that includes all firms that report a manufacturing sector as one of their sectors of activity, either primary or secondary. We thus can associate firms with industries at different levels of detail.

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

Table 14: Comparing Scott's National All to the Annual Survey of Manufacturers.

Province	Statcan ASM 2003	Statcan ASM 2005	Scott's 2001	Scott's 2005	Scott's 2009
Alberta	4,882	7,750	3,933	3,455	3,581
British Columbia	6,933	11,942	6,219	5,371	4,991
Manitoba	1,481	2,307	1,654	1,481	1,263
New Brunswick	963	1,533	1,395	1,258	1,175
Newfoundland and Labrador	522	765	576	540	472
Nova Scotia	1,106	1,944	1,676	1,495	1,296
Ontario	21,470	34,184	21,306	20,966	19,637
Prince Edward Island	211	351	328	327	280
Québec	15,251	23,042	15,939	14,166	12,560
Saskatchewan	1,008	1,804	1,353	1,305	1,091
Territories			0	40	45
Total	53,827	85,622	54,379	50,404	46,391

Notes: Province-level breakdown of manufacturing firms (NAICS 31–33) in the 2001, 2005, and 2009 Scott's National All databases versus Statistics Canada's 2003 Annual Survey of Manufacturers (ASM; CANSIM Table 301–0003) and 2005 ASM (CANSIM Table 301–0006). The 2003 ASM reports only employer firms with sales exceeding C$30,000 whereas the 2005 ASM reports information for manufacturing firms (including forestry, which is absent in the 2003 ASM) without a sales threshold. The Canadian Business Patterns 2008 of Industry Canada report about 56,000 employer establishments in manufacturing.

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. The census geography of 1996 and the postal codes as of May 2002 (818,907 unique postal codes) were associated with our 2001 sample. We also matched our 2005 sample with the 2001 Census geography and the postal codes as of January 2007 (861,765 unique postal codes). Finally, our 2009 sample was matched with the census geography of 2006, and the postal codes as of October 2010 (890,317 unique postal codes).

Table 15: Geographical Structure of the Census and PCCF Data.

	Census 1996 in the PCCF	Census 2001 in the PCCF	Census 2006 in the PCCF
Provinces and territories	13	13	13
Economic regions	74	76	76
Census divisions	285	288	288
Census subdivisions	4,410	4,088	3,692
Dissemination areas	34,940	42,297	45,904
Geographical concordance:			
Scott's All year	2001	2005	2009
PCCF version	May 2002	Jan 2007	Oct 2010
Census geography	1996	2001	2006
#unique postal codes	818,907	861,765	890,317

Notes: Geography of the 1996, 2001, and 2006 Censuses and concordances between *Scott's National All* databases and Statistic Canada's PCCFs.

Table 15 summarizes the geographical structure for the three years and provides details on

postal codes and census geographies. The highest level of geographical aggregation is that of the 10 provinces and 3 territories (PR); the second-highest level is that of economic regions (ER); the third-highest level is census divisions (CD); the fourth-highest level is census subdivisions (CS); and finally, the finest level is dissemination areas (DA). Census subdivisions, census divisions, and economic regions are useful spatial scales for computing discrete measures of localization like the Ellison and Glaeser (1997) index. Provinces are too coarse a spatial scale, whereas dissemination areas are too fine – most of the time, they contain no firms for any 4- or 6-digit NAICS industries. Note also that each postal code can be associated with multiple DAs. In that case, only one DA figures in the PCCF, so that the total number of DAs in the PCCF is smaller than that in the Census. This problem does not arise for larger geographical scales (provinces, regions, census divisions, and census subdivisions). To alleviate presentation, we only report results using census divisions. Results for other spatial scales are roughly similar and available upon request.

B. The Ellison-Glaeser and Spatially Weighted Ellison-Glaeser Indices

In this appendix, we briefly recall the logic underlying our discrete measures of localization. The Ellison-Glaeser index (henceforth, EG; Ellison and Glaeser, 1997), computed using employment data, is given by the following formula:[11]

$$\gamma_i \equiv \frac{G_i - (1 - \sum_r x_r^2) H_i}{(1 - \sum_r x_r^2)(1 - H_i)}, \tag{B.1}$$

where:

- $G_i \equiv \sum_r (s_{ri} - x_r)^2$ is the spatial Gini coefficient of industry i;
- x_r is the share of total employment in each region r;
- s_{ri} is the share of employment of region *r* in industry i;
- $H_i \equiv \sum_j z_{ji}^2$ is the Herfindahl index of the plant size distribution of industry i;
- z_{ji} represent the employment share of a particular firm j in industry i.

Given one well-known limit of the EG index – namely that it ignores the geographical positions of regions in space, the so-called 'checkerboard problem' – Guimaraes, Figueiredo, and Woodward (2011) propose an ad hoc solution. They introduce 'neighborhood effects' via a correction that leads to a *spatially weighted* version of the EG index (which we henceforth refer

[11] See Maurel and Sédillot (1999) for the definition of a very similar measure.

to as EGspat). Formally, they use the Ellison and Glaeser (1997) procedure and modify it to derive its spatially weighted version, which can be expressed as follows in matrix notation:

$$\widehat{\gamma}_S \equiv \frac{G_S - H_i(1 - X'\Psi X)}{(1 - H_i)(1 - X'\Psi X)}, \tag{B.2}$$

where:

- H_i is a Herfindahl index measuring the industry concentration in terms of plant-level employment;
- $G_S = (S - X)'\Psi(S - X)$ is the spatially weighted version of the raw concentration index;
- $X' = [x_1\ x_2\ \ldots\ x_J]$ is a vector containing the elements of the reference distribution;
- Ψ is a spatial weight matrix with generic element Ψ_{ij} and non-zero elements on the main diagonal. It is designed to account for spillovers that extend outside of the areal boundaries for which the EG index is computed. In general, $\Psi = \mathbf{I} + \mathbf{W}$, where $\mathbf{I}$ is the the identity matrix, and where $\mathbf{W}$ is a weight matrix for adjacent units. Adjacent units – also called contiguous units – are usually considered neighbors. In this study, we use two different matrices for Ψ, where the coefficients are either the inverse distance or the length of the common border between adjacent areal units. The latter measure has been computed using Canadian GIS data. A larger coefficient means that two adjacent units share a larger common border, so that there is greater potential that economic activity in one sector straddles the border. The latter effect increases the EGspat coefficient, which takes into account the spatial concentration across geographical units.

C. The Distance-Based Duranton-Overman Approach

In this appendix, we briefly recall the logic underlying our continuous measure of localization. Duranton and Overman (2005) propose a methodology that uses bilateral distances across pairs of establishments to identify localized industries. The idea is to apply sampling and bootstrapping techniques to determine the distribution of bilateral distances between the firms in an industry, and to compare it to a set of bilateral distances obtained from samples of randomly drawn firms. There are four steps.

First, we compute the pairwise distances between all firms in an industry and estimate a kernel density function of the distance distribution. Second, we construct a distribution of counterfactuals to assess whether the location pattern of a given industry departs statistically significantly from randomness. The counterfactuals are constructed on the basis that the firms in a given industry are located randomly among all possible locations where we do observe manufacturing activity. Third, we construct confidence intervals using our counterfactual random location distributions. Last, we test whether an industry is localized or dispersed,

by comparing the actual distribution of bilateral distances with the confidence bands derived from the sampling. We provide more information on the four steps in what follows.

First Step (Kernel Densities). Consider industry A with n firms. We compute the great circle distance, using postal code centroids, between each pair of firms in that industry. This yields $n(n-1)/2$ bilateral distances for industry A. Let us denote the distance between firms i and j by d_{ij}. Given n establishments, the kernel-smoothed estimator of the density of these pairwise distances, which we henceforth call K-density as in Duranton and Overman (2005), at any distance d is:

$$\widehat{K}(d) = \frac{1}{n(n-1)h} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} f\left(\frac{d - d_{ij}}{h}\right), \tag{C.1}$$

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:

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

We also compute the employment-weighted version of the K-density, which is given by

$$\widehat{K}_W(d) = \frac{1}{h\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(e_i + e_j)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (e_i + e_j) f\left(\frac{d - d_{ij}}{h}\right), \tag{C.2}$$

where e_i and e_j are the employment levels of establishment i and j, respectively.[12] The weighted K-density thus describes the distribution of bilateral distances between employees in a given industry, whereas the unweighted K-density describes the distribution of bilateral distances between plants in that industry.

Since the K-density is a distribution function, 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. Alternatively, we can view this as the probability that two randomly drawn establishments in an industry will be at most d kilometers away.

Second Step (Counterfactual Samples). 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. 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

[12] Contrary 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. Since our sample features many small firms, this appears preferrable to us. Using a multiplicative scheme would imply that our results may be too strongly driven by a few very large firms in a given industry.

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. This yields a set of 1,000 estimated values of the K-density at each distance d.

Third Step (Confidence Bands). 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.[13] 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.

Fourth Step (Identification of Location Patterns). 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. We can also define an index of global localization, $\gamma_i(d) \equiv \max\{\widehat{K}(d) - \overline{K}(d), 0\}$, as well as an index of global dispersion

$$\psi_i(d) \equiv \begin{cases} \max\{\underline{K}(d) - \widehat{K}(d)\} & \text{if} \quad \sum_{d=0}^{800} \gamma_i(d) = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{C.3}$$

Intuitively, if we observe a higher K-density than that of randomly drawn distributions, we consider the industry as localized. Similarly, if we observe a lower K-density than that of randomly drawn distributions, we consider the industry as dispersed. Last, the strength of localization and dispersion can be measured by $\Gamma_i \equiv \sum_d \gamma_i(d)$ and $\Psi_i \equiv \sum_d \Psi_i(d)$, which corresponds roughly to a measure of the 'surface' between the observed distribution and the upper- and lower-bounds of the confidence band.

[13] The 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.

D. Additional Tables and Results

Table 16: Ten Most and Least Localized 4-Digit Industries, EG and EGspat indices.

NAICS4	Most localized industries in 2001	EG	EGspat
3151	Clothing Knitting Mills	0.191	0.194
3152	Cut and Sew Clothing Manufacturing	0.068	0.068
3252	Resin, Synthetic Rubber, and Artificial and Synthetic Fibres and Filaments Manufacturing	0.064	0.067
3335	Metalworking Machinery Manufacturing	0.057	0.060
3313	Alumina and Aluminum Production and Processing	0.057	0.060
3369	Other Transportation Equipment Manufacturing	0.055	0.063
3117	Seafood Product Preparation and Packaging	0.053	0.064
3341	Computer and Peripheral Equipment Manufacturing	0.052	0.053
3159	Clothing Accessories and Other Clothing Manufacturing	0.050	0.051
3256	Soap, Cleaning Compound and Toilet Preparation Manufacturing	0.049	0.050

NAICS4	Least localized industries in 2001	EG	EGspat
3339	Other General-Purpose Machinery Manufacturing	0.002	0.002
3312	Steel Product Manufacturing from Purchased Steel	0.002	0.003
3149	Other Textile Product Mills	0.001	0.001
3324	Boiler, Tank and Shipping Container Manufacturing	0.001	0.001
3353	Electrical Equipment Manufacturing	-0.002	-0.001
3322	Cutlery and Hand Tool Manufacturing	-0.002	-0.001
3325	Hardware Manufacturing	-0.003	-0.001
3115	Dairy Product Manufacturing	-0.004	-0.003
3272	Glass and Glass Product Manufacturing	-0.009	-0.008
3122	Tobacco Manufacturing	-0.026	-0.015

NAICS4	Most localized industries in 2005	EG	EGspat
3151	Clothing Knitting Mills	0.267	0.274
3241	Petroleum and Coal Products Manufacturing	0.083	0.086
3152	Cut and Sew Clothing Manufacturing	0.080	0.080
3335	Metalworking Machinery Manufacturing	0.063	0.065
3162	Footwear Manufacturing	0.058	0.067
3159	Clothing Accessories and Other Clothing Manufacturing	0.058	0.060
3311	Iron and Steel Mills and Ferro-Alloy Manufacturing	0.057	0.058
3369	Other Transportation Equipment Manufacturing	0.053	0.062
3117	Seafood Product Preparation and Packaging	0.049	0.064
3313	Alumina and Aluminum Production and Processing	0.047	0.050

NAICS4	Least localized industries in 2005	EG	EGspat
3333	Commercial and Service Industry Machinery Manufacturing	-0.001	0.001
3272	Glass and Glass Product Manufacturing	-0.001	0.000
3321	Forging and Stamping	-0.002	0.002
3115	Dairy Product Manufacturing	-0.003	-0.002
3274	Lime and Gypsum Product Manufacturing	-0.004	0.002
3326	Spring and Wire Product Manufacturing	-0.005	-0.003
3365	Railroad Rolling Stock Manufacturing	-0.009	0.013
3161	Leather and Hide Tanning and Finishing	-0.013	0.004
3122	Tobacco Manufacturing	-0.022	-0.010
3336	Engine, Turbine and Power Transmission Equipment Manufacturing	-0.028	-0.027

Table 16 (continued): Ten Most and Least Localized 4-Digit Industries, EG and EGspat indices.

NAICS4	Most localized industries in 2009	EG	EGspat
3151	Clothing Knitting Mills	0.193	0.203
3311	Iron and Steel Mills and Ferro-Alloy Manufacturing	0.113	0.114
3152	Cut and Sew Clothing Manufacturing	0.079	0.079
3241	Petroleum and Coal Products Manufacturing	0.077	0.080
3331	Agricultural, Construction and Mining Machinery Manufacturing	0.065	0.068
3159	Clothing Accessories and Other Clothing Manufacturing	0.060	0.063
3117	Seafood Product Preparation and Packaging	0.053	0.067
3366	Ship and Boat Building	0.049	0.057
3313	Alumina and Aluminum Production and Processing	0.049	0.056
3254	Pharmaceutical and Medicine Manufacturing	0.048	0.049

NAICS4	Least localized industries in 2009	EG	EGspat
3121	Beverage Manufacturing	-0.002	-0.002
3133	Textile and Fabric Finishing and Fabric Coating	-0.003	-0.001
3326	Spring and Wire Product Manufacturing	-0.008	-0.006
3115	Dairy Product Manufacturing	-0.016	-0.014
3365	Railroad Rolling Stock Manufacturing	-0.017	0.002
3345	Navigational, Measuring, Medical and Control Instruments Manufacturing	-0.018	-0.017
3122	Tobacco Manufacturing	-0.023	-0.009
3336	Engine, Turbine and Power Transmission Equipment Manufacturing	-0.028	-0.028
3346	Manufacturing and Reproducing Magnetic and Optical Media	-0.028	-0.018
3279	Other Non-Metallic Mineral Product Manufacturing	-0.125	-0.123

Notes: EG and EGspat indices computed at the 4-digit NAICS level. The spatial scale used is census divisions (CD), and the weighting is inverse distance between CD centroids.

Table 17: Ten Most and Least Localized Industries According to the DO Index in 2001.

NAICS4	Industry name		DO
	Most localized industries	# of firms	Γ_A
3335	Metalworking Machinery Manufacturing	1011	0.297
3152	Cut and Sew Clothing Manufacturing	1803	0.197
3132	Fabric Mills	183	0.179
3131	Fibre, Yarn and Thread Mills	46	0.147
3321	Forging and Stamping	205	0.140
3361	Motor Vehicle Manufacturing	273	0.131
3372	Office Furniture (including Fixtures) Manufacturing	702	0.123
3344	Semiconductor and Other Electronic Component Manufacturing	391	0.121
3151	Clothing Knitting Mills	172	0.113
3351	Electric Lighting Equipment Manufacturing	207	0.095
	Least localized industries	# of firms	Ψ_A
3331	Agricultural, Construction and Mining Machinery Manufacturing	747	0.070
3366	Ship and Boat Building	290	0.047
3116	Meat Product Manufacturing	784	0.044
3212	Veneer, Plywood and Engineered Wood Product Manufacturing	326	0.042
3121	Beverage Manufacturing	400	0.035
3149	Other Textile Product Mills	877	0.034
3362	Motor Vehicle Body and Trailer Manufacturing	382	0.028
3346	Manufacturing and Reproducing Magnetic and Optical Media	350	0.003
3241	Petroleum and Coal Products Manufacturing	309	0.002
3363	Motor Vehicle Parts Manufacturing	685	0.0009

NAICS6	Industry name	# of firms	DO
	Most localized industries		Γ_A
313240	Knit Fabric Mills	31	0.488
315231	Women's and Girls' Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing	69	0.474
315210	Cut and Sew Clothing Contracting	212	0.377
315292	Fur and Leather Clothing Manufacturing	233	0.306
333519	Other Metalworking Machinery Manufacturing	739	0.303
336370	Motor Vehicle Metal Stamping	41	0.251
336110	Automobile and Light-Duty Motor Vehicle Manufacturing	211	0.250
333220	Rubber and Plastics Industry Machinery Manufacturing	43	0.227
332118	Stamping	172	0.219
333511	Industrial Mould Manufacturing	272	0.214
	Least localized industries	# of firms	Ψ_A
321911	Wood Window and Door Manufacturing	676	0.061
311614	Rendering and Meat Processing from Carcasses	363	0.056
312110	Soft Drink and Ice Manufacturing	220	0.056
336612	Boat Building	253	0.039
321215	Structural Wood Product Manufacturing	200	0.037
333110	Agricultural Implement Manufacturing	323	0.032
336215	Motor Home, Travel Trailer and Camper Manufacturing	104	0.027
336212	Truck Trailer Manufacturing	105	0.023
314990	All Other Textile Product Mills	583	0.022

Notes: The measures of localization and dispersion are defined as in Duranton and Overman (2005): $\Gamma = \sum_d \Gamma(d)$, where $\Gamma(d)$ is the maximum between zero and the difference between the empirical K-density and the upper bound of the global confidence band at distance d. Analogously, $\Psi = \sum_d \Psi(d)$, where $\Psi(d)$ is the maximum between zero and the difference between the lower bound of the global confidence band and the empirical K-density at distance d, provided that the empirical K-density does not exceed the upper bound over the whole distance range. See Appendix C.

Table 18: Ten Most and Least Localized Industries According to the DO Index in 2005.

NAICS4	Industry name		DO
	Most localized industries	# of firms	Γ_A
3335	Metalworking Machinery Manufacturing	957	0.283
3321	Forging and Stamping	185	0.224
3152	Cut and Sew Clothing Manufacturing	1548	0.178
3361	Motor Vehicle Manufacturing	207	0.156
3372	Office Furniture (including Fixtures) Manufacturing	680	0.116
3151	Clothing Knitting Mills	113	0.106
3344	Semiconductor and Other Electronic Component Manufacturing	368	0.097
3222	Converted Paper Product Manufacturing	657	0.071
3132	Fabric Mills	156	0.066
3328	Coating, Engraving, Heat Treating and Allied Activities	632	0.060
	Least localized industries	# of firms	Ψ_A
3331	Agricultural, Construction and Mining Machinery Manufacturing	703	0.082
3121	Beverage Manufacturing	384	0.063
3366	Ship and Boat Building	293	0.051
3116	Meat Product Manufacturing	737	0.046
3212	Veneer, Plywood and Engineered Wood Product Manufacturing	335	0.042
3279	Other Non-Metallic Mineral Product Manufacturing	714	0.021
3149	Other Textile Product Mills	869	0.020
3362	Motor Vehicle Body and Trailer Manufacturing	354	0.019
3326	Spring and Wire Product Manufacturing	120	0.013
3241	Petroleum and Coal Products Manufacturing	291	0.009
NAICS6	Industry name		DO
	Most localized industries	# of firms	Γ_A
315231	Women's and Girls' Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing	53	0.570
333220	Rubber and Plastics Industry Machinery Manufacturing	38	0.369
336110	Automobile and Light-Duty Motor Vehicle Manufacturing	146	0.359
332118	Stamping	157	0.317
315110	Hosiery and Sock Mills	34	0.302
333519	Other Metalworking Machinery Manufacturing	694	0.285
315292	Fur and Leather Clothing Manufacturing	190	0.254
336370	Motor Vehicle Metal Stamping	43	0.234
333511	Industrial Mould Manufacturing	263	0.201
315210	Cut and Sew Clothing Contracting	281	0.149
	Least localized industries	# of firms	Ψ_A
327320	Ready-Mix Concrete Manufacturing	600	0.094
312110	Soft Drink and Ice Manufacturing	187	0.069
321911	Wood Window and Door Manufacturing	538	0.045
336612	Boat Building	255	0.045
321215	Structural Wood Product Manufacturing	208	0.045
311614	Rendering and Meat Processing from Carcasses	360	0.039
333110	Agricultural Implement Manufacturing	310	0.038
327990	All Other Non-Metallic Mineral Product Manufacturing	667	0.019
336310	Motor Vehicle Gasoline Engine and Engine Parts Manufacturing	158	0.016
321114	Wood Preservation	48	0.013

Notes: The measures of localization and dispersion are defined as in Duranton and Overman (2005): $\Gamma = \sum_d \Gamma(d)$, where $\Gamma(d)$ is the maximum between zero and the difference between the empirical K-density and the upper bound of the global confidence band at distance d. Analogously, $\Psi = \sum_d \Psi(d)$, where $\Psi(d)$ is the maximum between zero and the difference between the lower bound of the global confidence band and the empirical K-density at distance d, provided that the empirical K-density does not exceed the upper bound over the whole distance range. See Appendix C.

Table 19: Ten Most and Least Localized Industries According to the DO Index in 2009.

NAICS4	Industry name		DO
	Most localized industries	# of firms	Γ_A
3335	Metalworking Machinery Manufacturing	829	0.268
3321	Forging and Stamping	161	0.171
3152	Cut and Sew Clothing Manufacturing	1096	0.153
3361	Motor Vehicle Manufacturing	169	0.145
3372	Office Furniture (including Fixtures) Manufacturing	588	0.092
3359	Other Electrical Equipment and Component Manufacturing	260	0.076
3222	Converted Paper Product Manufacturing	579	0.074
3341	Computer and Peripheral Equipment Manufacturing	154	0.059
3344	Semiconductor and Other Electronic Component Manufacturing	358	0.059
3328	Coating, Engraving, Heat Treating and Allied Activities	645	0.043
	Least localized industries	# of firms	Ψ_A
3273	Cement and Concrete Product Manufacturing	980	0.078
3331	Agricultural, Construction and Mining Machinery Manufacturing	661	0.074
3366	Ship and Boat Building	276	0.062
3219	Other Wood Product Manufacturing	1953	0.044
3121	Beverage Manufacturing	409	0.041
3212	Veneer, Plywood and Engineered Wood Product Manufacturing	284	0.037
3116	Meat Product Manufacturing	682	0.029
3362	Motor Vehicle Body and Trailer Manufacturing	324	0.022
3149	Other Textile Product Mills	764	0.020
3241	Petroleum and Coal Products Manufacturing	276	0.006

NAICS6	Industry name		DO
	Most localized industries	# of firms	Γ_A
315231	Women's and Girls' Cut and Sew Lingerie, Loungewear and Nightwear Manufacturing	37	0.607
336370	Motor Vehicle Metal Stamping	39	0.386
336110	Automobile and Light-Duty Motor Vehicle Manufacturing	115	0.382
333519	Other Metalworking Machinery Manufacturing	619	0.287
333220	Rubber and Plastics Industry Machinery Manufacturing	33	0.252
315292	Fur and Leather Clothing Manufacturing	147	0.244
332118	Stamping	140	0.220
315110	Hosiery and Sock Mills	21	0.206
335920	Communication and Energy Wire and Cable Manufacturing	39	0.199
333511	Industrial Mould Manufacturing	210	0.163
	Least localized industries	# of firms	Ψ_A
327320	Ready-Mix Concrete Manufacturing	559	0.085
321215	Structural Wood Product Manufacturing	183	0.058
336612	Boat Building	235	0.054
312110	Soft Drink and Ice Manufacturing	162	0.052
321911	Wood Window and Door Manufacturing	489	0.035
333110	Agricultural Implement Manufacturing	275	0.034
311614	Rendering and Meat Processing from Carcasses	326	0.031
314990	All Other Textile Product Mills	514	0.029
336310	Motor Vehicle Gasoline Engine and Engine Parts Manufacturing	170	0.021
312130	Wineries	140	0.015

Notes: The measures of localization and dispersion are defined as in Duranton and Overman (2005): $\Gamma = \sum_d \Gamma(d)$, where $\Gamma(d)$ is the maximum between zero and the difference between the empirical K-density and the upper bound of the global confidence band at distance d. Analogously, $\Psi = \sum_d \Psi(d)$, where $\Psi(d)$ is the maximum between zero and the difference between the lower bound of the global confidence band and the empirical K-density at distance d, provided that the empirical K-density does not exceed the upper bound over the whole distance range. See Appendix C.