How Densely Do Manufacturing Establishments Occupy Land?*

Kristian Behrens† Florian Mayneris‡ Théophile Ndjanmou Biéda§

February 27, 2023

Abstract

We construct a dataset on the size of parcels and on the footprint of buildings occupied by manufacturing establishments in Canada. We find that establishments in big cities and close to city centers occupy their parcels more densely, both in terms of employment and building footprint. Moreover, establishments with more employees use less land per worker. These facts cannot be rationalized by a Cobb-Douglas production function featuring land and labor as production factors. They are, however, consistent with a modified-CES production function in which land has a fixed cost component. The calibrated elasticity of substitution between land—as measured by parcel size—and labor is low, between 0.2 and 0.4.

Keywords: Land use; Land and production; Production function; Georeferenced data; Building and parcel polygons; Manufacturing.
JEL Classification: R32; R14; L60.

*We thank our discussant Matt Freedman, as well as Théophile Bougna, Julien Martin, Marlon Seror, Nate Baum-Snow, and participants at various seminars and conferences for helpful comments and suggestions. The authors acknowledge financial support from SSHRC (Insight Development Grant #430-2018-00959). Behrens gratefully acknowledges 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*. Any remaining errors are ours.

†Université du Québec à Montréal, Canada; and CEPR. Email: behrens.kristian@uqam.ca

‡Université du Québec à Montréal, Canada; and CEPR. Email: mayneris.florian@uqam.ca

§Université du Québec à Montréal, Canada. Email: ndjanmou_bieda.theophile@courrier.uqam.ca.

1 Introduction

The monocentric city model and its many refinements have given rise to a rich empirical literature on issues related to residential land, such as the extent and causes of urban sprawl or the effects of land use regulations on housing supply and prices. However, little is known—both theoretically and empirically—on the role of land in firms' production processes (Duranton and Puga, 2015). We make progress in that direction by analyzing the determinants of the employment density and structural density of the parcels occupied by manufacturing establishments in Canada.

We define *employment density* as the number of workers of an establishment divided by the surface of the parcel on which it is located. Similarly, we define *structural density* as the building-to-parcel ratio, i.e., the footprint of the building(s) divided by the surface of the parcel on which it is located.[1] We uncover three new facts on the employment and structural density of parcels occupied by manufacturing establishments in urban areas: (i) establishments occupy their parcels more densely both in terms of employment and structure in larger urban areas; (ii) both types of density decrease with distance to city centers; and (iii) large establishments occupy their parcels more densely in terms of employment, but not in terms of structure. These empirical regularities provide guidance on how to model land in the production function of manufacturing establishments. We can rationalize them through the lens of a conceptual framework—featuring a modified-CES production function—in which land: (i) has a fixed cost component; and (ii) has a low elasticity of substitution with labor. When taken together, our results show that big manufacturing firms located in large urban areas and close to city centers allow for more compact land use than small manufacturing firms located in smaller cities or farther away from city centers. These are interesting findings at a time when the preservation of open space and city compactness become more important to fight climate change.

Our analysis proceeds in four steps. First, we combine proprietary data on georeferenced manufacturing establishments in Canada with open-source data for parcel- and building polygons to construct a detailed dataset on the amount of land occupied by manufacturing establishments. Doing so, we fill an important gap since in firm-level balance-sheet data—when data on land is available at all—it is almost always conflated with data on non-land capital inputs, and quantity measures are usually not reported.[2]

In a second step, we regress our measures of employment density and of structural density on several characteristics of the establishments and their geographic en-

[1] Our definition of structural density is related to, yet slightly different from, the one generally used in the literature. For example, in Brueckner (1987), structural density is the capital-to-land ratio and is a measure of building height.

[2] Even when separate information on land is available, usually only its value is reported.

vironment, controlling for industry and geographic fixed effects. City size, distance to city centers, and/or establishment size in terms of employees are among the determinants with the highest explanatory power of the two variables we use to measure how densely manufacturing establishments occupy land. The results prove resistant to a wide range of robustness checks. They are also very similar when restricting the analysis to establishments in Montréal only, for which we can measure parcel size more precisely based on administrative data from the property assessment roll.

In a third step, we interpret our empirical findings through the lens of a conceptual framework where firms combine land and labor using a modified-CES production function with a minimum land requirement, similar to the Stone-Geary utility function used to model consumer preferences. The minimum land requirement—which may be viewed as a fixed cost—allows to rationalize the increasing relationship between employment density and establishment size. Our conceptual framework shows how the employment density of parcels depends on relative factor prices and their elasticity of substitution. More precisely, the CES structure of the production function implies that the elasticity of parcel size per worker to city size is the product of the elasticity of substitution between land and labor and the elasticity of (relative) land prices to city size. In the same vein, the semi-elasticity of parcel size per worker to the distance to city centers is the product of the elasticity of substitution between land and labor and the land price gradient.

In the last step, we use estimates of the density-elasticity of land prices from Combes et al. (2019) and of land price gradients from Albouy et al. (2018) to back out the elasticity of substitution between land and labor for manufacturing firms. We find positive values between 0.2 and 0.4, much smaller than unity as would be implied by the Cobb-Douglas production function. This finding suggests that land and labor are fairly bad substitutes in the production function of manufacturing firms. Note that the data we have and our focus in this paper is on the amount of land occupied by manufacturing establishments, not on the floor space they consume. For the subset of firms located in Montréal, we have information on floor space and our results show that floor space per worker is insensitive to the distance to city centers. This suggests that floor space and labor are even worse substitutes than land and labor.[3]

Our findings shed light on a topic we know little about empirically and contribute to the literature in a number of ways. First, while density, housing prices, and land price gradients for residential use have been extensively studied—both theoretically and empirically—"*little to no work has been devoted to the predicted gradients of the capital intensity of housing development and of housing consumption per household*" (Duranton and Puga, 2015, p. 523). Even less is thus known regarding the capital intensity of

[3]Put differently, firms can partly save on outdoor space or compensate for a smaller parcel size by occupying taller buildings, but they can hardly change the amount of floor space per worker they use.

production sites and of land consumption for production establishments. Our results reveal a decreasing gradient for structural density in line with estimates for residential housing (e.g., McMillen, 2006), although our gradients are weaker, and thus imply a low elasticity of substitution between land and labor.

Second, the few empirical studies on land used for production mostly relate to the effects of land-use regulations on productivity in the retail sector (e.g., Haskel and Sadun, 2012; Cheshire et al., 2014) or to the determinants of commercial real estate prices (e.g., Drennan and Kelly, 2010; Ahlfeldt and McMillen, 2018; Liu et al., 2018). Except for a couple of works on the patterns and determinants of the floor-to-area ratio (FAR; see, e.g., Barr and Cohen, 2014; Brueckner et al., 2017), we are not aware of studies on the quantity of land used by firms. This is likely due to a lack of data, and our work partially fills this gap by generating a new dataset.

Third, our results contribute to the literature on the costs and benefits of urban density (see Ahlfeldt and Pietrostefani, 2019, for a recent review) by emphasizing that the concentration of production in big firms and the concentration of big firms in big cities allows to save space compared to patterns that would be more dispersed across both smaller firms and smaller cities.

Last, our results have also theoretical ramifications. Canonical urban models generally assume that production is concentrated in dimensionless 'business districts', i.e., production requires no land. Notable exceptions—where firms and residents compete for land—include Fujita and Ogawa (1982), Lucas and Rossi Hansberg (2002), Pflüger and Tabuchi (2010), and Wrede (2013), but the way land enters the production function varies greatly across existing models—sometimes it is a pure fixed cost or Leontief, but more often it is via a Cobb-Douglas production function. To the best of our knowledge, no empirical studies corroborate to date a particular modelling strategy. Our empirical results reject the Cobb-Douglas specification for manufacturing firms and show that land should enter the production function with both a fixed component and a component where land and labor are (fairly) imperfect substitutes.[4] There are potentially several important implications arising from a low elasticity of substitution between labor and land for firms in an urban setting. For example, the concentration of high value-added services in urban cores may be due more to land prices pushing manufacturing out because of the low elasticity of substitution between labor and land rather

[4]Several recent studies show that the production function of many manufacturing sectors is not Cobb-Douglas (see, e.g., Oberfield and Raval 2021 for the US; Imbert et al. 2022 for China; and Mayneris 2022 for France). However, these studies analyze the substitutability between labor and capital without accounting explicitly for land. The construction sector is the only one for which we are aware of studies analyzing the role of land in the production process. Epple et al. (2010) and Combes et al. (2019), for example, estimate production functions for housing where land and non-land inputs (called "capital" are the two production factors). The latter show that for newly-built single-family homes, the production function for housing is well, though not perfectly, approximated by a Cobb-Douglas function.

than strong agglomeration economies pulling the service firms into the core. More work is called for here, but we conjecture that sorting patterns very much depend on how much of a firm's land is a fixed- compared to a variable-cost component.

The remainder of the paper is organized as follows. Section 2 explains the construction of our dataset and shows its representativeness. Section 3 provides sectoral descriptive statistics on the amount of land occupied by Canadian manufacturing establishments. Section 4 estimates the elasticity of plant-level land consumption to city- and plant-level characteristics. We develop a conceptual framework in Section 5 that puts some structure on our regression results and guides us in the quantification of the elasticity of substitution between land and labor. Section 6 concludes. Details on the construction of our database and robustness checks are relegated to the Appendix.

2 Data construction

We collect information on the amount of land occupied by manufacturing establishments to construct the two main measures we use throughout the paper.[5] The first measure is *parcel size per worker* and is an inverse measure of employment density. To compute it, we need the surface area of the parcels where the plants are located, which is derived from the polygons of the parcels. This measure captures both the building footprint and the outdoor space used by the plant for storage, parking, or green space. The second measure is the *building-to-parcel ratio*, i.e., the share of the parcel covered by the footprint of the buildings where the plants are located. To compute it, we need the surface area of the building footprint, which we derive from the polygons of the buildings where the plants are grounded.[6] Note that we do not observe the floor space used by manufacturing establishments and that we cannot infer it from our data. This is why our analysis is not about the amount of floorspace used by establishments, but about the amount of land they occupy.[7]

2.1 Methodology

We first briefly present the methodology used to construct the dataset. Details for each step and an extensive discussion of the quality of the final dataset are relegated to Appendices A and B. The dataset we build combines proprietary data on geo-referenced Canadian manufacturing establishments with open-source data for parcel- and building polygons. We use GIS tools to associate each establishment with specific building

[5] In what follows, we interchangeably use the terms 'establishment', 'plant', and 'firm'.

[6] The building footprint is also known as the 'gross area floor' in the assessment roll terminology.

[7] We can provide an analysis of the determinants of floorspace per worker using an auxiliary dataset covering the establishments located in Montréal (see Section 4.4).

and parcel polygons. We then compute the parcel size and the building footprint for each plant using the surface of its associated polygons.

2.1.1 Data collection and processing

Establishment-level data. We use the proprietary Scotts National All Business Directories, a dataset that draws information on plants operating in Canada from Business Register records and telephone surveys. It provides a fairly exhaustive coverage of the manufacturing sector, less so for services, which is one of the reasons why we focus on manufacturing in this paper. We use cross-sectional data for 2017, the closest year to the reference year for the polygon datasets that we use. This choice reduces potential measurement error due to changes in the delineation of buildings and parcels. It also allows for more precise geocoding as street names and configurations may change over time. The variables of interest for our analysis include the address of each plant, its industrial classifications (North American Industry Classification System, NAICS 6-digit level), an estimate of the number of workers at the site, and dummy variables for whether the plant reports an export activity and whether or not it is recorded as a headquarter. The dataset also contains information on the products manufactured by the plants and the broad sectors it is active in (manufacturing, wholesale, professional, scientific and technical services etc.). We geocode the plants using the procedure explained in Appendix A.[8]

Polygon datasets. We collect parcel and building polygons in Canada from numerous provincial and metropolitan sources. The full list of sources from which we collect these datasets, as well as a discussion of their quality, are relegated to Appendix A (see Table A1). Concerning parcels, we collect more than 4.5 million polygons covering the entire provinces of British Columbia (BC), Quebec (QC), and New Brunswick (NB) as well as the cities of Toronto, Oshawa, Windsor, and York in Ontario (ON). For the other provinces, we obtain data for Banff in Alberta (AB), Winnipeg in Manitoba (MB) and Regina and Saskatoon in Saskatchewan (SK). We did not obtain data for Nova Scotia (NS), Newfoundland and Labrador (NL), Prince Edward Island (PE), and the three Territories. Concerning buildings, we collect information from the Open Source data on building footprints in Canada released by Microsoft. These datasets contain 12,663,475 building footprints covering all provinces and territories.[9]

[8] The dataset already reports geographic coordinates for each plant but some of these coordinates are based on postal code centroids obtained from Post Canada's Postal Code Conversion files. These are necessarily less accurate than coordinates obtained from rooftop geocoding and do not permit to precisely associate plants with building- or parcel polygons.

[9] For additional information, see `https://blogs.bing.com/maps/2019-03/microsoft-releases-12-million-canadian-building-footprints-as-open-data`.

Other datasets. To make use of spatially fine-grained population census data, we collect the shapefiles of the boundaries of all dissemination areas (DA; the smallest geographic unit at which census data are publicly released), census metropolitan areas (CMA), census agglomerations (CA), economic regions (ER), and provinces and territories in Canada.[10] Combined with data from the population census of 2016, we obtain files that contain information on the population, the surface area, and the relations between the different levels of the census geography. We also obtain polygon files released by DMTI which record basic zoning restrictions in Canada, namely the main type of activity allowed in each area by local zoning policies (commercial/industrial, residential, recreational). Last, we collect information on some major infrastructures such as highway junctions (from the Canadian road network files), as well as the location of airports, seaports, and train freight stations (from the Open Government geographic data portal).

2.1.2 Construction of the surface measures

We use GIS tools to relate each geocoded plant in the establishment dataset to parcel and building polygons (see Appendix A for technical details). The mapping between plants and polygons allows us to construct the measures of land occupied by manufacturing establishments. The parcel size is the surface of the parcel polygon that contains the establishment, while the building footprint is the ground floor area of the building polygon that contains the establishment.

There is no one-to-one mapping between establishments, on the one hand, and parcel and building polygons, on the other hand. Sometimes, several establishments are on the same parcel. Put differently, there is some sharing of parcels and buildings. This should, however, not be a major problem for our analysis: compared to service establishments, manufacturing plants are less likely to have many neighbors. In the sample used for the analysis of parcel size, the average number of neighbors identified for each establishment based on the Scotts data is 1.3 and the median is 0. Since there are still some establishments that share parcels or buildings, we control in the

[10] A census metropolitan area (CMA) or a census agglomeration (CA) is formed by one or more adjacent municipalities centered on a population center (known as the core). A CMA must have a total population of at least 100,000 of which 50,000 or more must live in the core based on adjusted data from the previous Census of Population Program. A CA must have a core population of at least 10,000 also based on data from the previous Census of Population Program. To be included in the CMA or CA, other adjacent municipalities must have a high degree of integration with the core, as measured by commuting flows derived from data on place of work from the previous Census Program. An economic region (ER) is a grouping of complete census divisions (CDs), created as a standard geographic unit for analysis of regional economic activity. There are 76 economic regions in Canada that constitute a partition of the country. They are much smaller than provinces but, except for the very largest metropolitan areas, much bigger than cities. Finally, the 10 provinces and 3 territories are the federated political units in Canada.

regressions for the number of neighbors using a flexible polynomial function. We also verify that our main results are unchanged when focusing on establishments with no identified neighbors.[11]

Furthermore, the locations occupied by manufacturing establishments may be composed of several contiguous parcels and not just the ones on which the establishments fall during the geocoding process. Following discussions we had with employees from the Land Register of Québec, we think that this situation occurs rarely. This is coherent with the fact that, as shown by Brooks and Lutz (2016), assembled parcels have a much higher value than the sum of the values of the individual parcels, so that owners of contiguous parcels have an incentive to assemble them. Moreover, for the city of Montréal, we have exhaustive information from the property assessment roll that allows us to compute the number of parcels associated with the establishments' tax lots. We find that 85% of the manufacturing establishments in Montréal are on tax lots composed of a single parcel, confirming our discussion with the Land Register of Québec. Hence—provided that Montréal is representative of the rest of the country—measurement error remains very limited here.

Finally, for some plants the surface area of the parcel is smaller than the building footprint. This is because the assignment of establishments to parcels, on the one hand, and to buildings, on the other hand, is done independently. Hence, an establishment can be assigned to a building and to a parcel that do not correspond to the same lot. Furthermore, polygons may be mis-identified by the automatic recognition procedure. In particular, adjacent buildings may get amalgamated (or are just visible as a single polygon in the data), thus yielding larger polygons that straddle several parcels. In a robustness check, we reproduce our main results using only the sample of establishments for which we have both the parcel size and the building footprint, and for which the parcel size exceeds the building footprint.

2.2 Quality assessment

Assigning geocoded data to polygons delineated from satellite imagery inherently raises issues regarding data quality and accuracy. We relegate the detailed discussion of these issues to Appendix B. We simply want to mention here that to gauge the quality of the data obtained after the geocoding and assignment processes, we make use of the subset of data for the province of Québec (QC). The reason is that the polygon identifiers in the QC dataset are the same as the official identifiers of the polygons

[11] Although the Scotts dataset provides an almost exhaustive coverage of the manufacturing sector, it does not constitute the universe of plants in Canada, especially for services which are more sparsely present in that database. Thus, there is potential measurement error in the count of neighbors. We discuss the implications of this for the estimation results in Section 4.

as recorded on the government website of the Land Register "Infolot". We can thus randomly draw from our dataset a subset of plants in QC and compare their parcel identifier from "Infolot" (obtained from the address of the establishment) to the one obtained with our assignment procedure. Based on this comparison we can build a variable that measures 'data quality'—which we construct for the whole dataset, not just Québec—with three categories: excellent, good, and poor (see Appendix B for additional details). In the remainder of the paper, we only keep observations of 'excellent' quality (77.1% of the observations for which we have a measure of parcel size). We verify later that our results hold when: (i) including observations of lower quality; and (ii) when restricting our analysis to the Montréal dataset for which we have high-quality administrative data from the property assessment roll.

2.3 The final dataset

Our final dataset contains the plants from the Scotts database that: (i) are precisely geocoded; and (ii) have an excellent quality in terms of assignment to parcel- or building polygons.

We now discuss the representativeness of our data. Out of the $32,417$ manufacturing plants recorded in the Scotts database for 2017, we can assign parcel size of excellent quality to $8,708$ (26.86%) of them and building footprint size of excellent quality to $20,443$ (73%) of them. We are able to compute a building footprint-to-parcel ratio for 8,514 plants. The loss of data is mainly due to the absence of polygons for some provinces and cities and, to a lesser extent, the accuracy of the geocoding and polygon assignments (see Table A2 for more details). Concerning the sectoral representativeness of our data, Table A3 in Appendix D shows that the distribution of the 3-digit industries is broadly similar to that in the raw Scotts database. The correlation between these distributions exceeds 0.98.

From a geographic perspective, Table A4 in Appendix D shows the distribution of plants across provinces. As explained before, we lack parcel polygons for entire provinces which are thus missing from the estimation sample. Yet, the correlation of the geographic distribution with the raw Scotts data remains reasonably high at 0.77. Moreover, the provinces that host the majority of manufacturing in Canada are very well represented: Ontario, Quebec, and British Columbia account for 80.1% of the number of manufacturing plants in Canada and represent 87.3% of the plants in our parcel sample.

To conclude, our final dataset has good sectoral and geographic coverage and appears generally representative of manufacturing in Canada. There could still be some selection based on establishment characteristics. We hence use a probit model to assess the extent to which the establishments in the final sample exhibit specific observable characteristics compared to those for which we do not have reliable information on parcel size. Table A5 in Appendix D shows that, beyond the geographic fixed effects, very few establishment characteristics are related to the probability to be included in the regression sample. Moreover, the pseudo R-square of the regression is quite low at 0.32, of which only 5 percentage points are related to establishment characteristics. Put differently, there is little selection in the sample used for the analysis, and most of it is related to selection across provinces due to the fact that we do not have parcel data for entire provinces in Canada.

3 Land occupied by manufacturing establishments: Some sectoral statistics

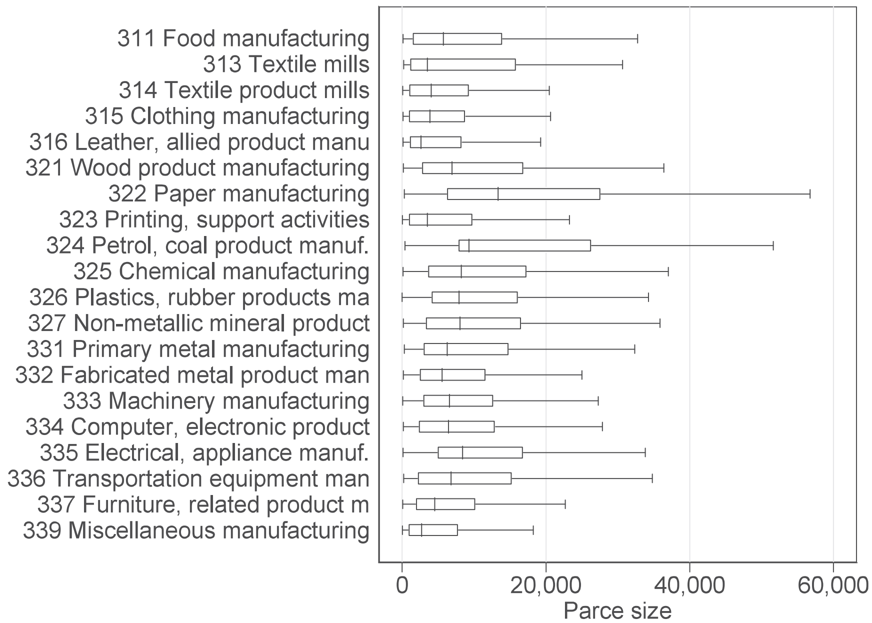
We now present sectoral statistics on the size of the parcels occupied by manufacturing establishments in Canada, on parcel size per worker, and on the share of their surface area covered by the footprint of their buildings (building-to-parcel ratio). These statistics are based on our final dataset.

3.1 Size of parcels by sectors

Figure 1 reveals substantial heterogeneity in the amount of land occupied by manufacturing establishments, both between and within sectors (see Table A6 in Appendix D for the associated figures). The average parcel size of a plant is around 13,354m^2, but the median is more than twice smaller, thus suggesting a very right-skewed distribution of parcel sizes. Moreover, the coefficient of variation equals 270%, revealing substantial heterogeneity in our sample. Part of that heterogeneity reflects between-sector differences, with some sectors having on average large parcels (e.g., "312 Beverages and Tobacco" (not on the graph); "331 Primary Metal Products"; "324 Petroleum and Coal Products"; and "322 Paper Manufacturing") whereas others have much smaller parcels (e.g., "316 Leather and Allied Products"; "315 Clothing and Textile"; and "323 Printing and Support Activities"). The coefficient of variation is large in all sectors, ranging from 130% to 530%, thus showing that the sizes of parcels are not only heterogeneous between sectors but also within sectors.

Figure 2 focuses on parcel size per worker instead of parcel size (see Table A7 in Appendix D for the exact figures). Parcel size per worker is an inverse measure of how densely land is occupied by manufacturing establishments in terms of employees. The patterns in Figure 2 reveal even more heterogeneity between and within sectors than for parcel size: the coefficient of variation is 430% for the whole sample, ranging from 130% to 880% across industries. In addition, the industry-level rank correlation

Figure 1: Parcel size by industry



Notes: This graph shows the distribution of parcel sizes across industries. The industry "312 Beverages and Tobacco" has been removed to keep the graph readable.

between parcel size and parcel size per worker is not statistically significant: sectors with the largest parcel sizes are not necessarily those with the most or the least densely occupied parcels in terms of workers.

3.2 Building-to-parcel ratio

We can compute the building-to-parcel ratio as another way to measure how densely a parcel is occupied: the higher this ratio, the more densely built the parcel is. Figure 3 shows a fair amount of heterogeneity in terms of the building-to-parcel ratio, both between and within sectors, even though this heterogeneity is less pronounced than the one observed for the parcel measure alone (the coefficient of variation equals 60% for the whole sample, and ranges from 37% to 79%). Some sectors, like "324 Petroleum and Coal Products", "331 Primary Metal Products", or "321 Wood Products" have small building-to-parcel ratios. On the contrary, "315 Clothing and Textile" and "323 Printing and Support Activities" exhibit high ratios, thus showing that they use relatively less outdoor space.

The heterogeneity we observe across sectors certainly reflects the different needs in terms of land across industries. The sectors with the lowest ratios are sectors that rely either on outdoor resources (wood, coal, non-metallic mineral products) or for which space for storage tanks (petrol, beverages) is important. By contrast, the sectors with

Figure 2: Parcel size per worker by industry

Notes: This graph shows the distribution of parcel sizes per worker across industries. The industry "312 Beverages and Tobacco" has been removed to keep the graph readable.

Figure 3: Building-to-parcel ratio by industry

Notes: This graph shows the distribution of the building-to-parcel ratios across industries. The industry "312 Beverages and Tobacco" has been removed to keep the graph readable. We constrain the sample to observations where the parcel size exceeds the building footprint.

the highest ratios are historically located in denser areas and belong to what could be called 'light manufacturing' (clothing, printing, textiles). Note that plants may react to higher land prices by reducing either their building footprint, their parcel footprint, or both; and that the ease with which either type of land ('indoors' or 'outdoors') can be adjusted may depend on the use (e.g., parking vs storage) and the industry. This might have important implications for the spatial sorting of sectors and their propensity to agglomerate. We return to several of these points in the following.

4 Land occupied by manufacturing establishments: An econometric analysis

We now analyze in detail the characteristics of the manufacturing establishments and their environment that determine how densely they occupy land. We focus on two quantities: the surface area of the parcel occupied by an establishment, divided by the number of workers; and the building-to-parcel ratio. As will become clear in Section 5, the first variable lends itself well to a structural interpretation of the regression coefficients. We first explain the equation we estimate, then present the benchmark results, and finally propose a variety of robustness checks.

4.1 Estimated equation

The equation we bring to the data is the following:

$$y_{i(s,z)} = \alpha \mathrm{Env}_i + \beta \mathrm{Estab}_i + \gamma \mathrm{Infra}_i + \theta_s + \eta_z + \epsilon_{i(s,z)}, \tag{1}$$

where i indexes an establishment, s is the 4-digit industry it belongs to, and z the economic region it is located in. The dependent variable $y_{i(s,z)}$ is either parcel size per worker—measuring the inverse employment density—or the building-to-parcel ratio—measuring structural density.

In equation (1), Env_i is a vector of characteristics related to the environment of establishment i. It includes the (log) size of the urban area it is located in, both in terms of population and of surface area (as per the 2016 Census); the weighted distance of the establishment to the city centers of the urban area; fixed effects identifying the type of zoning (commercial/industrial, residential, or recreational) in use at its location; and a fourth-degree polynomial in the number of neighbors on the same parcel.[12] In some specifications, we also control for the local density around the plant, measured by the (log) population density in the dissemination areas within a 500 meters radius.

[12]The urban area corresponds either to the Census Metropolitan Area or the Census Agglomeration. To identify city centers, we use a routine that identifies clusters of densely populated dissemination areas. The details of the procedure, as well as the weighting scheme, are presented in Appendix C.

Estab_i is a vector of characteristics related to the size and type of the activities carried out by establishment i. We control, in particular, for the (log) number of employees of the plant, dummies identifying headquarters and exporting plants, as well as counts of the number of 4-digit industries, of products, and of broad types of activity the plant is involved in.

Proximity to specific infrastructures might influence the amount of land per worker occupied by manufacturing establishments and their building-to-parcel ratio, either because of the size of the parcels available close to these infrastructures or because of how 'packed' establishments accept to be in order to enjoy proximity to these infrastructures. Infra_i is thus a vector containing the (log) distance from establishment i to the closest major airport, major seaport, train freight station, and highway junction.

Finally, θ_s and η_z stand for sector and economic regions fixed effects. They capture technological parameters and regional determinants that may drive how densely manufacturing establishments occupy land.

To account for auto-correlation between observations within urban areas, we cluster all standard errors at the CMA/CA level (Moulton, 1990). As mentioned in Section 2.2, we restrict the sample to observations for which our data on parcel size are of the highest quality.

4.2 Benchmark results

Employment density. Table 1 shows results with the parcel size per worker as the dependent variable. All regressions include industry- and economic region fixed effects, as well as a fourth-degree polynomial in the number of neighbors of the establishment on its parcel. In column (1), the only other covariates are the characteristics of the geographic environment of the establishment. In column (2), the log employment of the establishment is added to the set of regressors. Column (3) considers other individual characteristics of the establishment, while we control in column (4) for the log distance to various transport infrastructures. We retain this regression as the benchmark specification and use its estimates—together with others available in the literature—in Section 5 to infer the elasticity of substitution between land and labor in the production function through the lens of a theoretical framework. In column (5), we add to the set of environment characteristics the log population density in a 500 meters radius around the establishment. Finally, we present in column (6) the standardized coefficients of the benchmark specification from column (4).

Our regression results exhibit several robust patterns. First, with regard to the local environment, plants in populated urban areas, as well as plants that are located closer to city centers within urban areas, use less land per worker (the coefficient on the weighted distance to city centers being positive). In our benchmark specification, the elasticity of parcel size per worker to city population equals -0.227 and the semi-elasticity of parcel size to the weighted distance to city centers equals 0.033. Note that both coefficients decrease in absolute value when we control for the population density in the immediate surroundings of the establishment in column (5) (this latter variable has a negative and highly significant coefficient). This reflects the fact that population density is not homogeneous within cities yet decreases, on average, with distance to the center. Hence, when local population density is accounted for, the effect of the other two variables in weaker. Land prices are higher in big cities and lower at higher distances from city centers. We will show in Section 5 how we can use these elasticities to recover values for the elasticity of substitution between land and labor in the production function. Regarding zoning, not surprisingly, compared to establishments in commercial/industrial areas (the reference category), manufacturing plants in residential areas occupy less land per worker. This may be because land use is restricted to certain industries or parcels are smaller in the residential parts of cities so that they attract establishments with lower land requirements.

Turning to establishment characteristics, as expected, headquarters occupy less land per worker, while the opposite is true for exporters: 'office' functions require less space than functions related to production and exports for which factory space and warehousing are needed. The results also show that plants with a broader range of activity—measured in terms of the number of NAICS and product codes—tend to occupy less land per worker, even though the relationship is statistically weaker. Finally, the highest correlation is found for establishment size in terms of employees, with an elasticity between -0.65 and -0.60 depending on the specification: bigger establishments use less land per worker.

We see several explanations to this negative correlation. First, moving, opening, or closing a facility is costly so that firms adjust the size of the parcel they use less easily than their workforce; only when shocks are large and permanent enough do firms adjust their land consumption, most likely by moving or by opening and closing establishments (Bergeaud and Ray, 2021). This means that when firms grow or shrink, they first do so by adjusting their number of employees only, especially if they face transitory shocks. Then, if big firms are those that have grown more compared to their initial size, the negative correlation between parcel size per worker and establishment size could be related to the existence of adjustment cost. However, we ran regressions controlling for plant-level employment growth between 2013 and 2017, and the coefficient on establishment size is unaffected.[13] Second, as already mentioned before, the Scotts data are exhaustive for manufacturing but not for services, so that we possibly mis-measure the number of neighbors on the parcel. If there is measurement error, it is arguably more severe for small firms: they are more likely to share their location

[13] These results are available upon request.

Table 1: Determinants of parcel size per worker (inverse of employment density)

	ln Parcel size per worker					
	(1)	(2)	(3)	(4)	(5)	(6)
Characteristics of the local environment						
Ln Population CMA	-0.264^a	-0.242^a	-0.241^a	-0.227^a	-0.125^b	-0.237^a
	(0.070)	(0.056)	(0.055)	(0.065)	(0.051)	
Ln CMA surface area	0.081	0.064	0.068	0.088	0.025	0.047
	(0.091)	(0.067)	(0.067)	(0.083)	(0.069)	
Weighted Distance to city centers	0.040^a	0.038^a	0.038^a	0.033^a	0.025^a	0.220^a
	(0.003)	(0.004)	(0.004)	(0.006)	(0.004)	
$\mathbb{1}$ Residential	-0.504^a	-0.924^a	-0.918^a	-0.889^a	-0.707^a	-0.289^a
	(0.080)	(0.105)	(0.103)	(0.104)	(0.097)	
$\mathbb{1}$ Recreational	0.114	0.084	0.083	0.089	-0.000	0.022
	(0.094)	(0.107)	(0.107)	(0.105)	(0.091)	
Ln Population density 500m					-0.148^a	
					(0.012)	
Characteristics of the establishment						
Ln Employment		-0.619^a	-0.624^a	-0.626^a	-0.645^a	-0.590^a
		(0.016)	(0.016)	(0.016)	(0.017)	
$\mathbb{1}$ Headquarter			-0.070^b	-0.077^a	-0.058^b	-0.016^a
			(0.026)	(0.025)	(0.023)	
$\mathbb{1}$ Exporter			0.114^a	0.107^a	0.086^a	0.036 ^{a}
			(0.021)	(0.021)	(0.020)	
# functions in the estab.			-0.019	-0.019	-0.007	-0.007
			(0.039)	(0.037)	(0.035)	
# 4-digit NAICS in the estab.			-0.021^c	-0.020^c	-0.020^c	-0.018^c
			(0.012)	(0.012)	(0.012)	
# products produced in the estab.			-0.010^b	-0.010^b	-0.009^b	-0.017^b
			(0.004)	(0.004)	(0.004)	
Distance to transport infrastructure						
Ln Distance to major airport				-0.134^c	-0.009	-0.122^c
				(0.068)	(0.076)	
Ln Distance to major seaport				0.129^b	0.085^b	0.135^b
				(0.054)	(0.041)	
Ln Distance to freight station				0.020	0.022	0.016
				(0.052)	(0.026)	
Ln Distance to junction				-0.033	-0.007	-0.027
				(0.041)	(0.041)	
Observations	8,707	8,707	8,707	8,707	8,707	8,708
R-squared	0.287	0.561	0.564	0.568	0.588	0.568
Industry (4-digit) fixed effects	✓	✓	✓	✓	✓	✓
Economic region fixed effects	✓	✓	✓	✓	✓	✓
Controls for # neighbors	✓	✓	✓	✓	✓	✓

Notes: All regressions include a polynomial function of degree 4 in the number of neighbors of the establishment on its parcel. Only observations with the highest reliable information on parcel size are included. $\mathbb{1}$ denotes $\{0,1\}$ dummy variables. Standard errors clustered at the CMA/CA level in parentheses. ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.

with other businesses, and we may thus overestimate their parcel size per worker. This could also explain the negative correlation between parcel size per worker and establishment size we find. However, as shown in Figure 4, when we run our benchmark regression separately by establishment-size bins, the correlation between parcel size per worker and establishment size is close to -1 for establishments with 1–5 employees, and close to -0.6, the coefficient found for the whole sample, for establishments with 5–15, 15–50 and 50+ employees. This pattern of heterogeneity is inconsistent with the idea that the negative correlation estimated on the whole sample mostly reflects an overestimation of parcel size per worker for smaller plants. Thus, two explanations are more likely to explain why parcel size per worker decreases with establishment size.

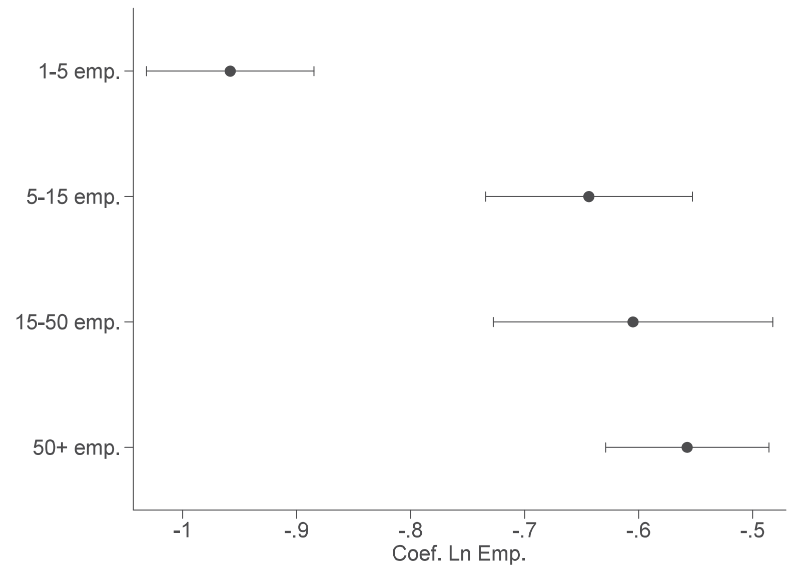
First, land likely has a fixed cost component. Indeed, if land were a variable cost only, under usual functional forms of the production function (e.g., Cobb-Douglas or CES, see Section 5 below), land per worker would be independent of the firm size in terms of employees. However, part of the land used by firms has the nature of a fixed cost: corridors, bathrooms, office spaces, or production spaces have a size that is partly independent of the number of workers using them. A second possible explanation is that—even though it is not the most frequent situation—some manufacturing firms occupy buildings with several floors. In Montréal, for example, we know the number of floors of the buildings occupied by establishments: 64% of the manufacturing establishments occupy a one-floor building, 16% a two-floor building, 5% a 3-floor building, and about 15% a 4-plus floor building.[14] It is likely that larger establishments occupy taller buildings but not necessarily much bigger parcels, which would show up in a lower parcel size per worker.

Finally, among the various types of transport infrastructure we consider, distance to major seaports is the only one robustly correlated with the amount of land per worker used by manufacturing establishments: plants located close to major seaports use less land per worker.

From a quantitative perspective, it is worth noting that the R^2 in our model is fairly large in all specifications, between 0.5 and 0.6, and that it is not soley driven by the industry- and economic region fixed effects. Thus, although we work with micro data at the establishment level, the model explains a substantial part of the variation in land per worker for the establishments. Among the regressors we consider, the standardized coefficients in column (6) show that four characteristics particularly stand out: the establishment size in terms of employees, being located in a residential zone, the population of the CMA and the weighted distance to city centers. Thus, city size and distance to the center, land use regulations, and establishment size have first-order effects on plants' land use per worker.

[14] This information is derived from the property assessment roll we have for Montréal. Unfortunately, we do not have it for the other cities in our sample.

Figure 4: Parcel size per worker and establishment size across size bins



Notes: This figure shows the coefficient and the 95% confidence interval on establishment size for the benchmark regression in column (4) of Table 1 run separately for each employment-size bin.

Structural density. That parcels in larger cities, in more central locations, and in residential zones within cities are more densely occupied in terms of employees does not mean that the establishments occupying these parcels use less floor space per worker: buildings on these parcels could cover a greater share of the parcel or could have multiple floors. The data we have do not allow us to recover floor space information, but we have the footprint of the buildings so that we can compute the building-to-parcel ratio, our measure of structural density.

Table 2 shows our results when the building-to-parcel ratio is used as the dependent variable (the regression sample is thus composed of establishments for which we have information on both building footprint and parcel size). Two of the four main determinants of the parcel size per worker appear as important determinants of the building-to-parcel ratio too: the city population and the weighted distance to city centers, with standardized coefficients well above 0.2 in absolute value. The building-to-parcel ratio increases with city size and decreases with the weighted distance to city centers. This latter effect likely reflects the fact that outdoor space is partly used for parking lots or green space, two dimensions on which firms can accept restrictions when land prices are high. We also know that the cost of surface parking increases with the value of land, which implies that firms and households save on land by investing in underground or structural parking when being closer to the city center (Brueckner and Franco, 2017). Indoor space probably exhibits, on the contrary, a stronger complementarity with the other production factors and can less easily be compressed. Interestingly, the coefficient on establishment size, by far the main determinant of parcel size per worker,

Table 2: Determinants of building-to-parcel ratio (structural density)

	ln Building-to-parcel ratio					
	(1)	(2)	(3)	(4)	(5)	(6)
Characteristics of the local environment						
Ln Population CMA	0.366^{a}	0.367^{a}	0.370^{a}	0.391^{a}	0.321^{a}	0.457^{a}
	(0.074)	(0.075)	(0.074)	(0.086)	(0.079)	
Ln CMA surface area	-0.147	-0.147	-0.150	-0.209^{c}	-0.164	-0.126^{c}
	(0.099)	(0.100)	(0.099)	(0.123)	(0.116)	
Weighted Distance to city centers	-0.045^{a}	-0.045^{a}	-0.045^{a}	-0.040^{a}	-0.034^{a}	-0.298^{a}
	(0.005)	(0.005)	(0.005)	(0.007)	(0.006)	
$\mathbb{1}$ Residential	0.432^{a}	0.411^{a}	0.413^{a}	0.386^{a}	0.261^{a}	0.142^{a}
	(0.073)	(0.072)	(0.072)	(0.085)	(0.074)	
$\mathbb{1}$ Recreational	-0.228^{b}	-0.230^{b}	-0.226^{b}	-0.232^{b}	-0.172^{c}	-0.062^{b}
	(0.090)	(0.090)	(0.089)	(0.095)	(0.094)	
Ln Population density 500m					0.101^{a}	
					(0.017)	
Characteristics of the establishment						
Ln Employment		-0.031^{a}	-0.028^{a}	-0.029^{a}	-0.016	-0.031^{a}
		(0.008)	(0.009)	(0.009)	(0.010)	
$\mathbb{1}$ Headquarter			-0.051	-0.041	-0.055	-0.010
			(0.038)	(0.034)	(0.034)	
$\mathbb{1}$ Exporter			-0.024	-0.019	-0.006	-0.007
			(0.020)	(0.020)	(0.020)	
# functions in the estab.			-0.035	-0.032	-0.040	-0.014
			(0.041)	(0.038)	(0.037)	
# 4-digit NAICS			0.004	0.001	0.000	0.001
			(0.016)	(0.015)	(0.016)	
# products			0.009^{c}	0.009^{b}	0.009^{b}	0.019 ^{b}
			(0.004)	(0.004)	(0.004)	
Distance to transport infrastructure						
Ln Distance to major airport				0.211^{a}	0.125	0.216^{a}
				(0.074)	(0.079)	
Ln Distance to major seaport				-0.120^{b}	-0.090^{c}	-0.143^{b}
				(0.054)	(0.053)	
Ln Distance to freight station				-0.077	-0.078^{c}	-0.068
				(0.057)	(0.040)	
Ln Distance to junction				-0.030	-0.048	-0.027
				(0.051)	(0.049)	
Observations	8,513	8,513	8,513	8,513	8,513	8,514
R-squared	0.290	0.291	0.292	0.299	0.311	0.299
Industry (4-digit) fixed effects	✓	✓	✓	✓	✓	✓
Economic region fixed effects	✓	✓	✓	✓	✓	✓

Notes: All regressions include a polynomial function of degree 4 in the number of neighbors of the establishment on its parcel. Only observations with the highest reliable information on parcel size are included. $\mathbb{1}$ denotes $\{0,1\}$ dummy variables. Standard errors clustered at the CMA/CA level in parentheses. ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1.

is now close to 0. All else equal, structural density is very modestly correlated with establishment size.

4.3 Robustness checks

Table 3 summarizes robustness checks on the determinants of the parcel size per worker and the building-to-parcel ratio. We present the coefficients for three variables of interest only, namely city population, weighted distance to city centers, and establishment size, but all covariates of the benchmark specification are included in the different regressions.

Table 3: Robustness checks

(a) Employment density:	ln parcel size per worker								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Ln Population CMA	-0.227^a	-0.191^a	-0.224^a	-0.221^a	-0.142^a	-0.213^a	-0.222^a	-0.245^a	-0.165^c
	(0.065)	(0.055)	(0.059)	(0.066)	(0.049)	(0.066)	(0.066)	(0.067)	(0.085)
Weighted Distance to city centers	0.033^a	0.027^a	0.030^a	0.032^a	0.021^a	0.034^a	0.033^a	0.036^a	0.028^a
	(0.006)	(0.004)	(0.005)	(0.005)	(0.003)	(0.007)	(0.005)	(0.006)	(0.007)
Ln Employment	-0.626^a	-0.655^a	-0.568^a	-0.625^a	-0.630^a	-0.709^a	-0.542^a	-0.633^a	-0.620^a
	(0.016)	(0.017)	(0.014)	(0.016)	(0.019)	(0.018)	(0.015)	(0.019)	(0.020)
Observations	8,707	12,138	8,531	8,048	7,254	6,947	5,333	8,707	6,793
R-squared	0.568	0.488	0.549	0.570	0.596	0.547	0.458	0.538	0.579
(b) Structural density:	ln building-to-parcel ratio								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Ln Population CMA	0.391^a	0.358^a	0.285^a	0.368^a	0.252^a	0.386^a	0.362^a	0.379^a	0.289^a
	(0.086)	(0.059)	(0.072)	(0.082)	(0.049)	(0.095)	(0.096)	(0.080)	(0.097)
Weighted Distance to city centers	-0.040^a	-0.036^a	-0.032^a	-0.038^a	-0.022^a	-0.043^a	-0.041^a	-0.040^a	-0.031^a
	(0.007)	(0.005)	(0.005)	(0.006)	(0.004)	(0.008)	(0.006)	(0.006)	(0.007)
Ln Employment	-0.029^a	-0.012	-0.010	-0.034^a	0.018^c	0.018	-0.032^b	-0.016	-0.013
	(0.009)	(0.012)	(0.010)	(0.008)	(0.010)	(0.018)	(0.014)	(0.010)	(0.010)
Observations	8,513	11,793	8,341	8,048	7,254	6,830	5,206	8,513	6,629
R-squared	0.299	0.251	0.285	0.292	0.220	0.312	0.314	0.300	0.283

Notes: All regressions include industry (4-digit) fixed effects, economic region fixed effects, and a polynomial function of degree 4 in the number of neighbors of the establishment on its parcel. Only observations with the highest reliable information on parcel size are included. $\mathbb{1}$ denotes $\{0,1\}$ dummy variables. Standard errors clustered at the CMA/CA level in parentheses. ^{a} p<0.01, ^{b} p<0.05, ^{c} p<0.1. See main text for a description of the sample used in each regression.

We show eight different robustness checks. Column (1) shows the benchmark results as a point of comparison. In column (2), we expand the sample to include all establishments for which we have information on the dependent variable, irrespective of the quality of the geocoding and polygon assignment procedures. In column (3), we eliminate the 1% tails of the distribution in terms of the dependent variable. In column (4), the sample only contains observations for which the information on both the parcel size and the building footprint is of excellent quality. In column (5), we eliminate observations for which parcel size is smaller than building footprint (which reflects misidentified polygons, mis-assignment to parcel and/or building polygons, or problems in the raw Microsoft building data as explained before). In column (6), we restrict the sample to manufacturing establishments with less than 50 employees to address the fact that large establishments may occupy several adjacent parcels, in which case we underestimate the amount of space they use (even though, as mentioned in Section 2.1.2, this case is certainly rare). In column (7), we restrict the sample to those establishments that have no identified neighbors on the same parcel or in the same building. Indeed, despite the fact that we control for the number of neighbors in the benchmark regressions using a fourth-degree polynomial, it is still possible that we mis-measure the actual amount of land occupied by establishments when several manufacturing firms occupy the same parcel. In the same vein, in column (8), we replace the number of neighbors by the mean of the number of neighbors on the parcel and in the building when both values are available. Indeed, the spatial mappings between establishments and parcels, on the one hand, and establishments and buildings, on the other hand, being done independently, these two figures may not be exactly the same. Finally, column (9) restricts the sample to establishments that are located farther than 5 kilometres from a city center, to ensure the patterns we uncover are not driven by what happens in very central locations.

Irrespective of whether we consider parcel size per worker or the building-to-parcel ratio, our results for city population and weighted distance to city centers are remarkably stable, both qualitatively and quantitatively (most of the time). The same applies to the relationship between parcel size per worker and establishment size, while the sign and significance of the correlation between building-to-parcel ratio and establishment size is less stable, but always close to zero and almost always negative. We are thus confident in our key results from the benchmark econometric analysis: controlling for establishment size, manufacturing establishments occupy parcels more densely in big cities and in central locations within cities, both in terms of employment- and structural density. Moreover, controlling for location, larger establishments use less land per worker—as measured by parcel size per worker.

4.4 Zooming in on Montréal

We have access to the property assessment roll for the city of Montréal.[15] These data contain information on the floor space of the properties and on the surface area of the parcels on which they are built (surface of the parcels being reported more often

[15]Property assessment rolls are decentralized in Canada and unfortunately, we could not get similar data for other cities or provinces.

than floor space). Using the geographic coordinates of the properties in the assessment roll, we can merge this information with the data used for the core analysis for 1,115 establishments (out of which 937 have information on floor space). Reassuringly, when both variables are available, the correlation between the surface area of the parcels we have used so far and the ones filled in the property assessment roll equals 0.883. This confirms that the spatial join procedure we have implemented generally allows us to recover reliable information on parcel sizes.

Table 4 reproduces the benchmark analyses of Tables 1 and 2 using the sample of Montréal establishments for which we have information from the assessment roll. We run regressions using alternatively the original and the roll data. We complement the results on parcel size per worker and building-to-parcel ratios with a regression where (the log of) floor space per worker is used as the dependent variable.[16] The results convey three main messages. First, when comparing columns (1) and (3) of Table 4 to the last column of Tables 1 and 2 respectively, it appears that Montréal is very representative of the whole Canadian sample: for the two variables of interest (average distance to city centers and establishment size) and for most of the covariates, the results are both qualitatively and quantitatively similar when using the whole sample and the sample restricted to Montréal. Second, the comparisons of columns (1) and (2) and of columns (3) and (4) of Table 4 reveal that using the original or the roll data does not affect the results, be it in terms of sign, size, or significance of the coefficients. Manufacturing establishments occupy their parcels more densely both in terms of employment and in terms of building footprint when they are located close to city centers. And bigger establishments occupy their parcels more densely too, especially in terms of parcel size per worker. Finally, the results in column (5) of Table 4 show that bigger establishments occupy less floor space per worker, but floor space per worker is not significantly related to the average distance to city centers. If—as for parcel size per worker and building-to-parcel ratio—the results on the Montréal sample are representative of the whole sample, this means that establishment do not reduce the amount of floor space they use when locating closer to city centers; they simply occupy the parcels more densely.

Overall, we believe the results on the Montréal sample strongly confirm the reliability of the benchmark results obtained with the data we assembled for a broader set of Canadian cities and provinces.

[16]Note that the distance to major seaports is removed since its correlation with the weighted distance to city centers equals 0.84 in Montréal. Population and surface area of the CMA are also excluded since there is no variation for these variables when focusing the analysis on a single CMA.

Table 4: Focus on Montréal

	ln Parcel size per worker		ln Building-to-parcel ratio		ln Floor space per worker
	Original data	Role data	Original data	Role data	Role data
	(1)	(2)	(3)	(4)	(5)
Characteristics of the local environment					
Weighted Distance to city centers	0.162^{a}	0.221^{a}	-0.255^{a}	-0.286^{a}	-0.001
$\mathbb{1}$ Residential	-0.340^{a}	-0.307^{a}	0.176^{a}	0.131^{a}	-0.159^{a}
$\mathbb{1}$ Recreational	0.054^{b}	0.035	-0.008	0.012	0.010
Characteristics of the establishment					
Ln Employment	-0.590^{a}	-0.549^{a}	-0.120^{a}	-0.113^{a}	-0.524^{a}
$\mathbb{1}$ Headquarter	-0.066^{a}	-0.074^{a}	0.038	0.047	-0.009
$\mathbb{1}$ Exporter	0.026	0.044^{c}	-0.019	-0.042	0.041^{c}
# functions in the estab.	-0.087^{a}	-0.056^{b}	0.074^{c}	0.032	-0.045^{c}
# 4-digit NAICS in the estab.	0.114^{a}	0.062	-0.058	0.007	0.044
# products produced in the estab.	-0.033	-0.010	0.029	-0.003	-0.020
Distance to transport infrastructure					
Ln Distance to major airport	-0.127^{a}	-0.130^{a}	0.112^{a}	0.109^{a}	-0.063^{a}
Ln Distance to freight station	0.014	-0.002	-0.126^{a}	-0.091^{b}	-0.000
Ln Distance to junction	-0.066^{a}	-0.033	0.082^{b}	0.035	-0.009
Observations	1,053	1,053	1,039	1,039	882
R-squared	0.641	0.562	0.380	0.309	0.721
Industry (4-digit) fixed effects	✓	✓	✓	✓	✓
Economic region fixed effects	✓	✓	✓	✓	✓
Controls for # neighbors	✓	✓	✓	✓	✓

Notes: The regression sample only includes establishments in Montreal. All regressions include a polynomial function of degree 4 in the number of neighbors of the establishment on its parcel. Only observations with the highest reliable information on parcel size are included. $\mathbb{1}$ denotes $\{0,1\}$ dummy variables. The table reports standardized coefficients. ^{a} $p<0.01$, ^{b} $p<0.05$, ^{c} $p<0.1$.

5 Conceptual framework and structural interpretation

We now propose a simple conceptual framework to interpret the empirical regularities described in the previous section and use it to give a structural interpretation for some of the elasticities we estimate.

5.1 Setup

We assume that input markets are competitive and that firms are price takers in factor markets. Let i denote firms, s sectors, and z zones. We index firms by $i(s,z)$. When there is no confusion, we use i for short. Firm i has the following production function:[17]

$$Y_{i(s,z)} = A_i \left\{ \alpha_{i(s,z)} \left[\kappa_{i(s,z)}(P_i - \overline{P}_s) \right]^{\frac{\sigma_s - 1}{\sigma_s}} + (1-\alpha_{i(s,z)}) L_i^{\frac{\sigma_s-1}{\sigma_s}} \right\}^{\frac{\sigma_s}{\sigma_s - 1}} \tag{2}$$

where Y_i, P_i, and L_i stand for firm-level output, parcel (land) inputs, and the number of workers, respectively; A_i is a Hicks-neutral productivity shifter; $\alpha_{i(s,z)}$ is a technological parameter—with both a sectoral and a firm component—that influences the intensity of the production function in land and labor; and $\kappa_{i(s,z)}$ is a land-augmenting productivity parameter specific to firm i (and/or its industry and zone). Observe that there is a minimum land requirement $\overline{P}_s$ in sector s, which captures the presence of some fixed costs or indivisibilities in land consumption. Furthermore, σ_s is the elasticity of substitution between land and labor. We assume this parameter is industry specific and $\sigma_s > 0$, i.e., land and labor inputs are imperfect substitutes in production. Note that (2) exhibits constant returns to scale at the firm level in $P_i - \overline{P}_s$ and L_i.[18]

Letting p_z and w_z denote the unit price for parcels and labor in zone z, respectively, standard unit cost minimization yields:

$$p_z = \frac{\alpha_{i(s,z)}\kappa_{i(s,z)}[\kappa_{i(s,z)}(P_i - \overline{P}_s)]^{-\frac{1}{\sigma}} Y_{i(s,z)}}{\alpha_{i(s,z)}[\kappa_{i(s,z)}(P_i - \overline{P}_s)]^{\frac{\sigma_s-1}{\sigma_s}} + (1-\alpha_{i(s,z)})L_i^{\frac{\sigma_s-1}{\sigma_s}}} \tag{3}$$

$$w_z = \frac{(1-\alpha_{i(s,z)})L_i^{-\frac{1}{\sigma_s}} Y_{i(s,z)}}{\alpha_{i(s,z)}[\kappa_{i(s,z)}(P_i - \overline{P}_s)]^{\frac{\sigma_s-1}{\sigma_s}} + (1-\alpha_{i(s,z)})L_i^{\frac{\sigma_s-1}{\sigma_s}}} \tag{4}$$

[17] We can easily add capital as a third production factor. If we consider that its price is constant across space—as usual in the literature—this does not change the analysis. With different prices across locations, the analysis becomes more involved. Since we are mostly interested in parcel size per worker, we do not develop the case with capital in more detail in this paper.

[18] We do not rule out the presence of increasing returns to scale external to the firm. This would, e.g., be the case when $A_i = A_i(\mathbf{L}_{s,z})$ depends on aggregate employment $\mathbf{L}_{s,z}$ in sector s and zone z.

We focus on the ratio P_i/L_i as this is the theoretical equivalent of the parcel size per worker used in our empirical analysis. Since

$$\frac{p_z}{w_z} = \frac{\alpha_{i(s,z)}}{1-\alpha_{i(s,z)}} \left(\kappa_{i(s,z)}\right)^{\frac{\sigma_s - 1}{\sigma_s}} \left(\frac{L_i}{P_i - \overline{P}_s}\right)^{\frac{1}{\sigma_s}},$$

we can express parcel size per unit of labor as follows:

$$\frac{P_i}{L_i} = \left(\frac{\alpha_{i(s,z)}}{1-\alpha_{i(s,z)}}\right)^{\sigma_s} \left(\kappa_{i(s,z)}\right)^{\sigma_s - 1} \left(\frac{p_z}{w_z}\right)^{-\sigma_s} + \frac{\overline{P}_s}{L_i}. \tag{5}$$

Observe that P_i/L_i is independent of L_i if land has no fixed-cost component (i.e., if $\overline{P}_s = 0$). We have seen in our empirical analysis that this is clearly not the case: P_i/L_i decreases with L_i, which suggests that $\overline{P}_s > 0$.

5.2 Determinants of P_i/L_i

Our framework highlights three main types of determinants of firm-level parcel size per worker: (i) the relative price of land; (ii) technological/productivity parameters; and (iii) an additional term that depends on firm size and the importance of land as a fixed production factor.

Relative price of land. Conditional on technological parameters, productivity parameters, and firm size, the firm-level parcel size per worker is a decreasing function of its relative price, its sensitivity being determined by the elasticity of substitution σ_s between production factors.

Technological/productivity parameters. Parcel size per worker also depends on technological parameters and, due to the spatial sorting of plants, these are probably correlated in the data with the local relative price of production factors. For example, the land-intensity of the firm-level production function, as determined by the technological parameter $\alpha_{i(s,z)}$, matters for the relative quantity of land used by firms. For a given relative price of land, firms with low $\alpha_{i(s,z)}$ use relatively less land. We should then observe that firms with low $\alpha_{i(s,z)}$ sort into places where land is relatively expensive. The discussion is more involved for the land-augmenting productivity parameter $\kappa_{i(s,z)}$. Whether high or low $\kappa_{i(s,z)}$ firms sort into zones where land is expensive depends on the value of σ_s. Indeed, equation (5) above requires us to distinguish three cases:

(i) When $\sigma_s = 1$, any variation in factor-augmenting productivity $\kappa_{i(s,z)}$ leaves the relative demand P_i/L_i unaffected. There is no obvious spatial sorting of firms based on land-augmenting productivity in this case.

(ii) When $\sigma_s < 1$, firms cannot easily substitute labor for land. For a given relative price of land, firms with a high $\kappa_{i(s,z)}$ will then use less land per worker. Put differently, it is optimal for firms with a high land-augmenting productivity to tilt their demand towards non-land inputs when production factors are not easily substitutable. In this case, we should observe that firms with a high land-augmenting productivity sort into places where land is relatively expensive.

(iii) When $\sigma_s > 1$, firms can easily substitute labor for land. For a given relative price of land, firms with a high $\kappa_{i(s,z)}$ will then use more land per worker. Put differently, it is optimal for firms with a high land-augmenting productivity to tilt their demand towards the land input when production factors are easily substitutable. In this case, we should see firms with a relatively low land-augmenting productivity sort into places where land is relatively expensive.

Before proceeding, it is worth noting that high-A_i establishments are more likely to locate in zones with high factor costs (parcel prices p_z and wages w_z) since only they can afford the high production costs there. However, as equation (5) reveals, the parcel size per worker P_i/L_i used by establishments does not depend on their total factor productivity A_i, since the latter is a Hicks-neutral productivity shifter.[19]

Fixed land requirements. Last, for a given relative price of land and technological parameters, firms in sectors with larger fixed requirements mechanically use more land per worker.

5.3 Implications for empirical estimation

Because of the presence of the fixed land requirements $\overline{P}_s$, we cannot readily log-linearize equation (5). However, we can proceed as follows. We first rewrite equation (5) as:

$$\begin{aligned}\frac{P_i}{L_i} &= \underbrace{\left(\frac{\alpha_{i(s,z)}}{1-\alpha_{i(s,z)}}\right)^{\sigma_s} \left(\kappa_{i(s,z)}\right)^{\sigma_s - 1} \left(\frac{p_z}{w_z}\right)^{-\sigma_s}}_{\equiv \xi_{i(s,z)}} + \frac{\overline{P}_s}{L_i} \\ &= \left(\frac{\alpha_{i(s,z)}}{1-\alpha_{i(s,z)}}\right)^{\sigma_s} \left(\kappa_{i(s,z)}\right)^{\sigma_s - 1} \left(\frac{p_z}{w_z}\right)^{-\sigma_s} \left(1 + \frac{\overline{P}_s}{\xi_{i(s,z)} L_i}\right). \qquad (6)\end{aligned}$$

We then log-linearize (6), introducing a constant term β_0 and a reduced-form error term $\widetilde{\epsilon}_i$, and obtain

$$\ln\left(\frac{P_i}{L_i}\right) = \beta_0 + \beta_1 \ln\left(\frac{p_z}{w_z}\right) + \epsilon_i \qquad (7)$$

[19] Hence, external returns to scale that may be subsumed in the Hicks-neutral productivity shifter also do not affect firms' parcel size per worker. This vindicates the fact that we do not attempt to control for agglomeration effects in our empirical analysis.

where

$$\epsilon_i = \sigma_s \ln\left(\frac{\alpha_{i(s,z)}}{1-\alpha_{i(s,z)}}\right) + (\sigma_s - 1)\ln \kappa_{i(s,z)} + \ln\left(1 + \frac{\overline{P}_s}{\xi_{i(s,z)} L_i}\right) + \widetilde{\epsilon}_i \tag{8}$$

is a structural error term. In the previous section, we estimated the elasticity of parcel size per worker to city population and the semi-elasticity of parcel size per worker to the weighted distance to city centers. Through the lens of the model, $\beta_1 = -\sigma_s$ so that:

$$\frac{\partial \ln\left(\frac{P_i}{L_i}\right)}{\partial \ln \mathrm{Pop}_z} = -\sigma_s \frac{\partial \ln\left(\frac{p_z}{w_z}\right)}{\partial \ln \mathrm{Pop}_z} \quad \text{and} \quad \frac{\partial \ln\left(\frac{P_i}{L_i}\right)}{\partial \mathrm{Dist}_i} = -\sigma_s \frac{\partial \ln\left(\frac{p_z}{w_z}\right)}{\partial \mathrm{Dist}_i}.$$

From the literature, $\partial \ln (p_z/w_z)/\partial \ln \mathrm{Pop}_z > 0$ and $\partial \ln (p_z/w_z)/\partial \mathrm{Dist}_i < 0$, and our regressions show $\partial \ln (P_i/L_i)/\partial \ln \mathrm{Pop}_z < 0$ and $\partial \ln (P_i/L_i)/\partial \mathrm{Dist}_i > 0$. Hence, our estimates imply a positive value for σ_s, which is reassuring. Quantitatively, considering for now that the elasticities $\partial \ln (p_z/w_z)/\partial \ln \mathrm{Pop}_z$ and $\partial \ln (p_z/w_z)/\partial \mathrm{Dist}_i$ are given, we can infer the value of σ_s as:

$$\sigma_s = -\frac{\partial \ln\left(\frac{P_i}{L_i}\right)}{\partial \ln \mathrm{Pop}_z} \Bigg/ \frac{\partial \ln\left(\frac{p_z}{w_z}\right)}{\partial \ln \mathrm{Pop}_z} = -\frac{\partial \ln\left(\frac{P_i}{L_i}\right)}{\partial \mathrm{Dist}_i} \Bigg/ \frac{\partial \ln\left(\frac{p_z}{w_z}\right)}{\partial \mathrm{Dist}_i} \tag{9}$$

However, the theoretical discussion in Section 5.2 shows that, before we can structurally interpret our estimated elasticities, we need to discuss the endogeneity issues arising from the presence of $\alpha_{i(s,z)}$, $\kappa_{i(s,z)}$, and $\ln\left(1 + \frac{\overline{P}_s}{\xi_{i(s,z)} L_i}\right)$ in the structural error term (8).

Spatial sorting of firms and endogeneity. The type of bias arising from the firm-specific requirements in terms of land and labor is straightforward: low $\alpha_{i(s,z)}$ firms sort into high p_z/w_z zones and—in the absence of controls or valid instruments—the OLS estimate of $\partial \ln (P_i/L_i)/\partial \ln \mathrm{Pop}_z$ is likely to be biased downward and the one of $\partial \ln (P_i/L_i)/\partial \mathrm{Dist}_i$ is likely to be biased upward. The spatial sorting of firms based on $\kappa_{i(s,z)}$ induces the same type of biases, but this is less straightforward to establish. Indeed, as discussed before, three cases need to be distinguished:

(i) When $\sigma_s = 1$, spatial sorting of firms based on land-augmenting productivity is not an issue and there is no endogeneity bias.

(ii) When $\sigma_s < 1$, firms with high $\kappa_{i(s,z)}$ sort into places where the relative price of land is high. In this case, the naive estimates of $\partial \ln\left(\frac{P_i}{L_i}\right)/\partial \ln \mathrm{Pop}_z$ and $\partial \ln\left(\frac{P_i}{L_i}\right)/\partial \mathrm{Dist}_i$ suffer from a downward and an upward bias repsectively.

(iii) When $\sigma_s > 1$, firms with a low land-augmenting productivity sort into places where land is relatively expensive, meaning that again the naive estimates of the coefficients suffer from the same biases.

To summarize, the direction of the bias related to the spatial sorting of firms based on their technological and productivity parameters is always the same: in the absence of adequate controls or valid instruments, $\partial \ln (P_i/L_i)/\partial \ln \text{Pop}_z$ is likely to be under-estimated and $\partial \ln (P_i/L_i)/\partial \text{Dist}_i$ over-estimated. In both cases, we thus obtain an *upper bound* for σ_s.

Fixed costs and firm size. The second type of bias arises if land has a fixed-cost component. More productive and thus larger firms are more likely to be found in high p_z/w_z zones. Since they also have a smaller parcel size per worker—because the fixed requirements are distributed over a larger workforce—not controlling for this leads to a downward biased estimate of $\partial \ln (P_i/L_i)/\partial \ln \text{Pop}_z$ and an upward biased estimate of $\partial \ln (P_i/L_i)/\partial \text{Dist}_i$.

5.4 Inferring σ_s

Following the above discussion, the possible fixed-cost components of land as well as the technological and productivity parameters need to be controlled for to infer meaningful values of σ_s from the estimation of $\partial \ln (P_i/L_i)/\partial \ln \text{Pop}_z$ and $\partial \ln (P_i/L_i)/\partial \text{Dist}_i$. The former is relatively straightforward to do: this is the reason why we included firm size as an additional explanatory variable into our econometric analysis. As we have shown, the elasticity of P_i/L_i with respect to L_i is negative, highly significant, and very stable across specifications. Thus, land has a fixed component and we control for it.

Let us now discuss the technological and productivity parameters, which are less straightforward to control for. We neither directly observe firms' relative land requirements $\alpha_{i(s,z)}$ nor their land-augmenting productivity parameters $\kappa_{i(s,z)}$. However, the various controls we use in our regressions in Section 4 are likely correlated with these parameters and thus should be reasonable proxies for them.

Firm-level relative land requirements are certainly partly determined by some sectoral parameters of the production function. These are controlled for by the NAICS 4-digit industry fixed effects in our regressions. Still, the part of $\alpha_{i(s,z)}$ that is specific to establishments and $\kappa_{i(s,z)}$ are not controlled for by such fixed effects. However, we believe that four sets of our controls partly deal with this. First, land requirements vary with the functions that the establishment carries out and with its international exposure. We actually showed that headquarters occupy less land per worker while the opposite is true for exporters. Second, local governments have various zoning policies that affect the quantity of land available for production. In particular, zones that are specifically dedicated to a commercial or industrial use are probably more attractive to firms that require a lot of land per worker, while the opposite should be true for residential zones where land suitable for production is likely to be scarce. This is why we also included zoning type fixed effects. Third, the proximity of transport infrastructure could affect the quantity of land available for production or attract firms that are specific in terms of their needs for land (e.g., the exporters we mentioned above). This is why we control for the distance to the closest major airport, major seaport, train freight station, and highway junction, even though these variables do not appear to be major determinants of parcel size per worker in the end. Finally, establishments involved in a large number of activities in terms of products or sectors may require different types of facilities (multi-floor buildings for example to separate the various product lines), which could imply specific technological and productivity parameters regarding their land inputs. This is why we added controls for the number of products, 4-digit NAICS industries, and broad sectors covered by the plants' operations to our regressions.

Using the relationships in equation (9) and the results in column (4) of Table 1, we now back out the value of σ_s. We start with $\partial \ln (P_i/L_i)/\partial \ln \text{Pop}_z$, which we estimate to equal -0.227. To back out σ_s, we need a value for $\partial \ln (p_z/w_z)/\partial \ln \text{Pop}_z$. Data on commerical land prices are notoriously difficult to find, and we do not have them for Canada. In other words, we do not observe the relative price p_z/w_z. We thus rely on estimates available in the existing literature. Based on French data, Combes et al. (2019) find that the elasticity of the price of parcels (per square metre) to city population is roughly equal to 0.6, while using French data too, Combes et al. (2008) find an elasticity of individual wages to population density of 0.03.[20] In France, the elasticity of relative land prices to city size/city population density is thus equal to 0.57. Taking this as a reference value for Canada, equation (9) implies a value of $\sigma_s = 0.227/0.57 = 0.4$.

We repeat the same exercise for the estimate of $\partial \ln (P_i/L_i)/\partial \text{Dist}_i$, equal to 0.033. Again, we are not aware of clean estimates of land price gradients for Canadian cities. However, Albouy et al. (2018) provide estimates of the ratio of land values per acre in the city center (0.5 miles from downtown) and 10 miles away from it for more than 300 urban areas in the US. The weighted average ratio equals 6.5 (using urban area population as weights). This corresponds to a semi-log gradient of 0.197.[21] Since urban areas are delineated following commuting patterns, the average wage should not vary too much within them. Then, taking -0.197 as the reference value for $\partial \ln (p_z/w_z)/\partial \text{Dist}_i$ for Canada and using equation (9), it follows that $\sigma_s = 0.033/0.197 = 0.17$.

[20] These two elasticities are not estimated for the same period and at the exact same spatial scale, but they are cleanly estimated with very detailed data. We are not aware of better estimates in the literature to obtain a measure of $\partial \ln \left(\frac{p_z}{w_z}\right)/\partial \ln \text{Pop}_z$. Moreover, the two regressions from which these estimates derive both contain the surface area of the unit over which population and population density are computed. In such an empirical framework, the elasticity to population and to population density are equivalent.

[21] Assuming that the log of land price linearly depends on the distance to the city center, and since Albouy et al. (2018) estimate the ratio of land values at 0.5 and 10 mile from downtown to equal 6.5 on average, the gradient is given by $-\ln(6.5)/9.5 = -0.197$.

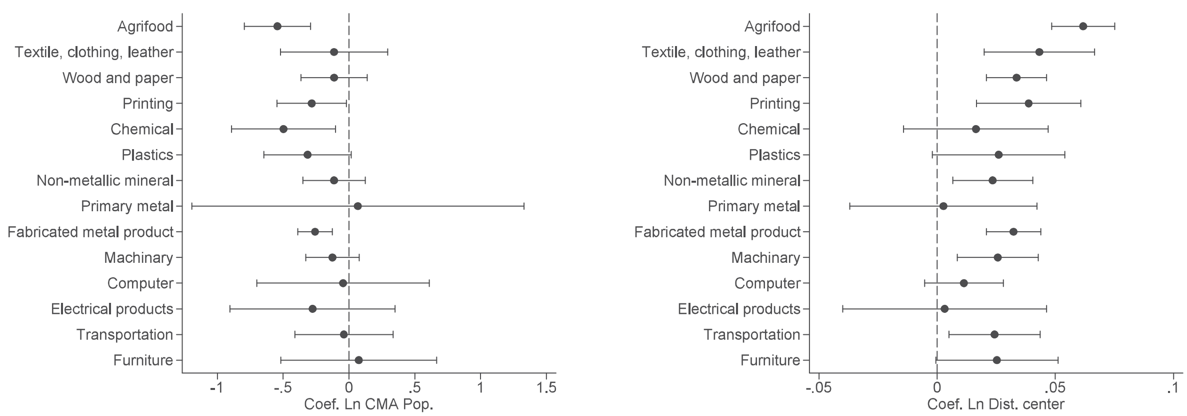
The two values of σ_s implied by the quantification exercises we propose, 0.17 and 0.4, are quite far from the ubiquitous Cobb-Douglas specification that has been used in the existing literature (and which implies that $\sigma_s = 1$). They suggest that labor and land, as measured by parcel size, are complements rather than substitutes in the production function of manufacturing establishments. As already mentioned throughout the paper, some substitutability exists because conditional no their number of employees, establishments can partly give up on outdoor space or occupy taller buildings. However, the results based on the Montréal sample in section 4 show that floor space per worker is not significantly related to the average distance to city centers. This suggests that the relationship between floor space (instead of parcel size) and labor is Leontieff, i.e. they are perfect complements (once the fixed cost dimension is taken into account).

In our conceptual framework, we assumed σ_s is sector specific since there is no reason a priori to believe that land and labor are equally substitutable in all sectors. To see whether the average σ_s masks heterogeneity, we investigate the cross-sectoral heterogeneity in the two elasticities we can estimate from our data.

Figure 5: Heterogeneity of coefficients by sector and by parcel size per worker

(a) ln CMA population

(b) Weighted distance to city centers



Notes: The graphs shows the point estimate and the 90% confidence intervals of the associated coefficients with the explanatory variable indicated in the heading. Regressions are run separately for the different sectors using the benchmark specification in column (4) of Table 1.

Figure 5 reports the sectoral estimates for the two covariates of interest, city population and weighted distance to city centers. In line with the pooled results, the coefficients we obtain are most of the time negative (but not always significant due to noisy estimates sometimes) for city population, and they are always positive (and most of the time significant) for weighted distance to city centers. However, once standard errors are accounted for, it is hard to see any significant heterogeneity across sectors. This suggests that σ_s is low in every sector, so that most of the heterogeneity across sectors highlighted in Section 3 comes from productivity or sectoral parameters.

6 Conclusion

Using a uniquely detailed dataset on the quantity of land used by manufacturing establishments in Canada, we provide what we believe is the first evidence on the individual and local determinants of the amount of land used for manufacturing production. Big establishments occupy less land per worker. Establishments located in big cities and close to city centers occupy their parcels more densely too, both in terms of employment and building footprint. Land thus has a strong fixed cost component, and a quantification of our regression results suggests that for the variable part of land consumption, land and labor are not easily substitutable. Since Canadian manufacturing is likely representative of manufacturing in other developed countries, our data and results are potentially interesting for other researchers, e.g., to calibrate models or for structural estimation exercises. Furthermore, the increasing availability of big open-source data on parcel and building polygons provides ample opportunities to replicate and extend our analysis to other countries in the future. In a rapidly urbanizing world where the structure of cities and the density of economic activity become increasingly important—e.g., to mitigate climate change—this seems like a worthwhile endeavour.

References

Ahlfeldt, Gabriel M. and Daniel P. McMillen, "Tall Buildings and Land Values: Height and Construction Cost Elasticities in Chicago, 1870-2010," *The Review of Economics and Statistics*, 2018, *100* (5), 861–875.

— **and Elisabetta Pietrostefani**, "The economic effects of density: A synthesis," *Journal of Urban Economics*, 2019, *111* (C), 93–107.

Albouy, David, Gabriel Ehrlich, and Minchul Shin, "Metropolitan Land Values," *The Review of Economics and Statistics*, July 2018, *100* (3), 454–466.

Barr, Jason and Jeffrey P. Cohen, "The floor area ratio gradient: New York City, 1890–2009," *Regional Science and Urban Economics*, 2014, *48*, 110–119.

Behrens, Kristian, Brahim Boualam, and Julien Martin, "Are clusters resilient? Evidence from Canadian textile industries," *Journal of Economic Geography*, 2020, *20* (1), 1–36.

Bergeaud, A. and S. Ray, "Adjustment Costs and Factor Demand: New Evidence From Firms' Real Estate," *Economic Journal*, 2021, *131* (633), 70–100.

Brooks, Leah and Byron Lutz, "From Today's City to Tomorrow's City: An Empirical Investigation of Urban Land Assembly," *American Economic Journal: Economic Policy*, 2016, *8* (3), 69–105.

Brueckner, Jan K., "The structure of urban equilibria: A unified treatment of the muth-mills model," in E. S. Mills, ed., *Handbook of Regional and Urban Economics*, Vol. 2 of *Handbook of Regional and Urban Economics*, Elsevier, 1987, chapter 20, pp. 821–845.

_ and Sofia F. Franco, "Parking and Urban Form," *Journal of Economic Geography*, 2017, *17* (1), 95–127.

_ , Shihe Fu, Yizhen Gu, and Junfu Zhang, "Measuring the Stringency of Land Use Regulation: The Case of China's Building Height Limits," *The Review of Economics and Statistics*, July 2017, *99* (4), 663–677.

Cheshire, Paul C., Christian AL Hilber, and Ioannis Kaplanis, "Land use regulation and productivity–land matters: evidence from a UK supermarket chain," *Journal of Economic Geography*, 2014, *15* (1), 43–73.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon, "Spatial wage disparities: Sorting matters!," *Journal of urban economics*, 2008, *63* (2), 723–742. Publisher: Elsevier.

_ , _ , and _ , "The Costs of Agglomeration: House and Land Prices in French Cities," *The Review of Economic Studies*, July 2019, *86* (4), 1556–1589.

Drennan, Matthew P. and Hugh F. Kelly, "Measuring urban agglomeration economies with office rents," *Journal of Economic Geography*, 2010, *11* (3), 481–507.

Duranton, Gilles and Diego Puga, "Urban land use," in "Handbook of regional and urban economics," Vol. 5, Elsevier, 2015, pp. 467–560.

Epple, Dennis, Brett Gordon, and Holger Sieg, "A new approach to estimating the production function for housing," *American Economic Review*, 2010, *100* (3), 905–24.

Fujita, Masahisa and Hideaki Ogawa, "Multiple equilibria and structural transition of non-monocentric urban configurations," *Regional science and urban economics*, 1982, *12* (2), 161–196.

Haskel, Jonathan and Raffaella Sadun, "Regulation and UK retailing productivity: evidence from microdata," *Economica*, 2012, *79* (315), 425–448.

Imbert, Clément, Marlon Seror, Yifan Zhang, and Yanos Zylberberg, "Migrants and Firms: Evidence from China," *American Economic Review*, 2022, *112* (6), 1885–1914.

Liu, Crocker H., Stuart Rosenthal, and William Strange, "The vertical city: Rent gradients, spatial structure, and agglomeration economies," *Journal of Urban Economics*, 2018, *106* (C), 101–122.

Lucas, Robert E. and Esteban Rossi Hansberg, "On the internal structure of cities," *Econometrica*, 2002, *70* (4), 1445–1476.

Mayneris, Florian, "Does the urban wage premium imply higher firm-level labor shares in cities?," Technical Report 2022.

McMillen, Daniel P., "Testing for monocentricity," *A Companion to Urban Economics. Oxford: Blackwell*, 2006, pp. 128–140.

Moulton, Brent R, "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Unit," *The Review of Economics and Statistics*, May 1990, *72* (2), 334–338.

Oberfield, Ezra and Devesh Raval, "Micro Data and Macro Technology," *Econometrica*, 2021, *89* (2), 703–732.

Pflüger, Michael and Takatoshi Tabuchi, "The size of regions with land use for production," *Regional Science and Urban Economics*, 2010, *40* (6), 481–489.

Wrede, Matthias, "Heterogeneous skills and homogeneous land: segmentation and agglomeration," *Journal of Economic Geography*, 2013, *13* (5), 767–798.

Appendix material

A. Data Appendix

Geocoding. Geocoding consists in providing an address to a geocoder—a particular Application Programming Interface (API) used to recover geographic coordinates of addresses–which returns the latitude and longitude of the corresponding address. The geocoder also provides the address related to the coordinates of the points it returns so that we can verify if the input address and the return address match.

For the sake of precision, we use three different options to perform the geocoding. The first option uses the commercial API of the Google Map server to geocode each plant based on the address recorded in the Scotts database. The second option uses the same API but combines the company's name with the address as the input for the geocoder. In doing so, small errors in the address reported in the Scotts data can be corrected and the accuracy of the geocoding improved. The third option uses the point coordinates provided in the DMTI database, which is an extensive database containing more than 15 million feature points representing Canadian addresses and their related geographic coordinates with 'rooftop' precision. We merge the Scotts addresses with the DMTI address using the API of ArcGIS, a commercial Geographic Information Systems (GIS) software.

Once we have geocoded the addresses, we compare the coordinates (latitude, longitude) returned by the three options and assign to each plant the coordinates that are most likely the accurate ones. Accuracy is based on two criteria: (i) the distances between the point coordinates yielded by the three options (so as to identify probable errors, i.e., points that are very far away from the other return values); and (ii) the match between the postal codes recorded in the Scotts database and the postal codes returned by the geocoder for each option (so as to keep only the points for which the postal code corresponds to the one recorded in the Scotts database). If several different points are returned for the same establishment, the coordinates retrieved from Google Maps based on the company name and the address are preferred to the coordinates obtained via Google Map using the address only, which are themselves preferred to the DMTI coordinates.

Finally, we construct a variable with three categories to grade the accuracy of the geocoding process for each plant based on how convergent the three options are in terms of establishment location. We retain only observations that are either 'rooftop' (i.e., exactly coded) or 'range interpolated' (i.e., interpolated based on a range of address numbers); we do not consider the rest (e.g., postal-code level) as being accurate enough to assign plants to polygons.

Data sources. We extensively explored existing open-access data sources on various websites and got in touch with several institutions to obtain information on parcel- and buildings polygons and footprints in Canada. The main relevant data sources for our work are the following:

- Statistics Canada, via the official website of the Canadian Government, provides several datasets including data on buildings that are open for public use.
- Some Assessment Rolls of different municipalities—which are in charge of computing the value of the tenure taxes based on the nature, the location, and the scope of the properties—provide open-access data.
- Cadastral information: Some provinces and cities in Canada do have information on the parcels where buildings are located.
- GIS databases of cities: The websites of some cities provide GIS data which record parcels polygons and/or footprints of buildings of their localities.
- Open source data on building footprints in Canada released by Microsoft: These datasets contain 12,663,475 building footprints covering all provinces and territories.[22]

The table below provides the complete list of polygon datasets that we collected along with the links where they can be accessed.

Polygon dataset quality. We collected polygon datasets from the above sources. These datasets come in different data formats (KML, shapefile, geodataset, etc) and are for different reference years. During their processing, we identified and solved the following challenges linked to the quality of the data:

- Quality of the collected files: The polygon datasets we collected are not homogeneous. The formats of the files are not always the same and the reference units of the polygon datasets are different in some cases (feet, meters, etc.) and sometimes not indicated at all in the files. To solve this problem we converted all the files into shapefile format (.shp), harmonized the units to meters, and projected each dataset into a suitable coordinate system according to the position of the locality it refers to. We consider as a suitable coordinate system one which does not alter distances. In most cases, the 'Albers conic conformal system' is used, as generally recommended for Canada. We also construct for each polygon dataset the following key variables: a unique identifier, the surface area, and the number

[22]For additional information, see `https://blogs.bing.com/maps/2019-03/microsoft-releases-12-million-canadian-building-footprints-as-open-data`.

Table A1: Overview of datasources

Locality	Coverage	Last update		Polygon type	Licence	Links
Locality	Coverage	Last update		Polygon type	Licence	Links
Alberta	AB province	2019	province	Building footprints	OSM/Statcan	https://github.com/Microsoft/CanadianBuildingFootprints
Alberta	Banff	2017	Banff	Parcels	open data	http://banffmaps.ca/opendata/
Alberta	Winnipeg	2017	Winnipeg	Parcels	open data	https://data.winnipeg.ca/Assessment-Taxation-Corporate/Map-of-Assessment-Parcels/rt7t-3m4m
British Columbia	CB province	2019	province	Building footprints	OSM/Statcan	https://github.com/Microsoft/CanadianBuildingFootprints
British Columbia	CB province	2016	province	Parcels	Open data	https://catalogue.data.gov.bc.ca/dataset/parcelmap-bc-parcel-fabric
Manitoba	MB province	2019	province	Building footprints	OSM/Statcan	https://github.com/Microsoft/CanadianBuildingFootprints
Manitoba	Brandon	2017	Brandon	Parcels	open data	http://opengov.brandon.ca/OpenDataService/opendata.html
New Brunswick	province	2019	province	Building footprints	OSM/Statcan	https://github.com/Microsoft/CanadianBuildingFootprints
New Brunswick	province	2019	province	Parcels	open data	https://gnb.socrata.com/api/geospatial/rzzg-85tb?method=export\
Newfoundland and Labrador	NL province	2019	province	Building footprints	OSM/Statcan	https://github.com/Microsoft/CanadianBuildingFootprints
Newfoundland and Labrador	St john	2019	St John	Parcels	open data	http://catalogue-saintjohn.opendata.arcgis.com/
North-West Territories	NT territories	2019	province	Building footprints	opendata	http://opendata.yellowknife.ca
Nova Scotia	NS province	2019	province	Building footprints	open data	https://github.com/Microsoft/CanadianBuildingFootprints
Nunavut	NU territories	2019	province	Building footprints	OSM/Statcan	https://github.com/Microsoft/CanadianBuildingFootprints
Ontario	ON province	2019	province	Building footprints	OSM/Statcan	https://github.com/Microsoft/CanadianBuildingFootprints
Ontario	Oshawa	2017	Oshawa	Parcels	open data	https://city-oshawa.opendata.arcgis.com/datasets?t=Durham\%20Housing
Ontario	York	2019	York	Parcels	open data	https://insights-york.opendata.arcgis.com/datasets/parcel
Ontario	Toronto	2017	Toronto	Parcels	open data	https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/
Ontario	Windsor	2017	Windsor	Parcels	open data	www.citywindsor.ca/opendata/Lists/OpenData/Attachments/20/Land\%20Parcels.kmz
Prince-Edward Island	PE province	2019	province	Building footprints	OSM/Statcan	https://github.com/Microsoft/CanadianBuildingFootprints
Quebec	QC province	2019	province	Building footprints	OSM/Statcan	https://github.com/Microsoft/CanadianBuildingFootprints
Quebec	QC province	2018	province	Parcels	InfoLot	UQAM data warehouse (https://appli.mern.gouv.qc.ca/infolot
Saskatchewan	Regina	2017	Regina	Parcels	open data	http://open.regina.ca/
Saskatchewan	SK province	2019	province	Building footprints	OSM/Statcan	https://github.com/Microsoft/CanadianBuildingFootprints
Yukon Territories	YT territories	2019	province	Building footprints	OSM/Statcan	https://github.com/Microsoft/CanadianBuildingFootprints

Notes: Listing of all the datasources that we use in the paper.

of neighbors of each polygon recorded in the dataset. The latter variable is useful to check for the quality of the area assignation process for each plant. The LIDAR dataset source gives the building footprints along with an estimation of the height of the building. These files report the minimum and the maximum height detected by the signal used to scan the space.

- Matching buildings to parcels: The polygon datasets we collected have two different features. The first one is the parcel-polygon that represents the amount of land used by a plant to host its main building and possibly some other spaces (auxiliary buildings, parking, storage, etc.). The second polygon type is the building-polygon that represents only the building of the plant. Theoretically, the building footprint should be included in the parcel outline. Yet, in some cases the building overlaps with more than one parcel. As a result, the surface of the building footprint is greater than the surface of the parcel to which its is related. We solve this issue by aggregating up all the parcels that overlap with the building.

Assignment to polygons. We have, on the one hand, a geocoded establishment-level dataset and, on the other hand, different polygon datasets. To merge them, we use the spatial join tools available in the open-source software Quantum GIS (QGIS) to map each plant to a polygon. More precisely, we overlay the polygon datasets (parcels and buildings) with the coordinate point layers representing the geocoded establishments. Figure A1 shows an example of how the geocoded Scotts plants are overlaid on the building polygon layer for the spatial join process.

As is well known, spatial join can be a somewhat noisy process. Hence, not all plants fall exactly onto a polygon (neither parcels nor buildings). For each plant, we thus perform three assignment options. The first option relates each plant to the polygon onto which it falls; in that case, the distance between the plant and the polygon is assumed to be 0. If the plant does not fall exactly onto a polygon, it has no associated polygon. The second option then relates each plant to the polygon whose centroid is the closest, and we compute the distance between the plant and that centroid. Finally, the third option relates each plant to the polygon whose border is the closest; we again compute the distance between the plant and that border. We then compare the three (or two) distances obtained in the three option and we take as the final assignment the polygon corresponding to the shortest distance. Obviously, when the plant falls onto a polygon, it is that polygon which is assigned to the plant since the distance is zero. When the shortest distance is greater than 75 meters we consider that the process is too noisy and we do not assign that polygon to the plant. In addition, to avoid assigning the surface of corridors to plants, we compute for each polygon its number of neigh-

Figure A1: Polygon layer with geocoded establishments overlaid

bors. If an assigned polygon has more than 10 neighors, we consider that the polygon is a corridor or a common space and we do not assign that polygon to the plant.

We then construct a variable corresponding to the combination of assignments pointing in the direction of the polygon the establishment is assigned to. For example if the options "Border" and "Center" assign the plant to the same polygon whereas the "Within" option points to a different polygon for the same plant, then assignment variable for that plant will be "Center-Border". Thus, the assignment variable has the following 7 categories : (1) "Within-Center-Border"; (2) "Within-Center"; (3) "Within-Border"; (4) "Center-Border" (5) "Within"; (6) "Center"; (7) "Border".

Based on this assignment variable, we construct a quality variable as follows: i) we cross-tabulate the assignment variable with the dummy we could build for the observations from Quebec and that identifies those establishments which are assigned to the right polygon (described in section 2.2); ii) for all of our observations, we define as "Excellent" those observations whose assignment category has a high probability of being located on their actual polygon as measured based on observations from Quebec; "Good" is for observations whose assignment category has an intermediate probability of being located on their actual polygon; and "Poor" is for all the categories with a low probability of being located on their actual polygon. Doing so, we implicitly assume that the mapping between the assignment variable and the dummy identifying correct observations in Quebec is representative of the entire country.

For the Parcel-based measure, the process leads to grade as "Excellent" the plants whose assignment category is "Within-Border-Center" or "Within". These plants with an "Excellent" parcel-based measure have a 89% probability of being positioned on their actual polygon. Plants graded as "Good" are those whose assignment category is "Within-Border" or "Within-Center". The plants of "Good" quality have a 60% probability of being positioned on their actual polygon. Finally, "Poor" is the grade for observations whose assignment category is "Border"; "Center-Border" or "Center"; these observations have a 16% probability of being located on their actual polygons.

For the Building-based measure, the category "Excellent" comprises the plants whose assignment category is "Within-Border-Center", "Within-Center", "Within-Border" and "Border-Center". The observations rated as "Excellent" for the Building-based measure have a 78% probability of being positioned on their actual polygon. The quality "Good" is for observations whose assignment category is "Within"; for them, the probability of being positioned on their actual polygon is equal to 66%. The grade "Poor" encompasses observation whose assignment category is "Border" or "Center". These plants have a 36% probability of being positioned on their actual polygon.

Summary: step-by-step data-construction procedure. Below is a summary of the main workflow to construct our dataset.

Step 1. *Creating a unique addresses*. From the Scotts dataset, unique addresses are identified since several plants can share the same location. We create a unique identifier for each address. This step prepares the geocoding process, and will avoid to geocode several times the same address. A dataset of unique addresses is then generated with variables, the detailed address, and the address identifier.

Step 2. *Geocoding unique addresses*. We use the dataset of unique addresses as input for the geocoding process described above. The output file contains geocoded addresses, in addition to the inputs variables, the geographic coordinates of each address, the detailed address as recorded in the database of the geocoder (Google or DMTI) as well as a quality variable indicating the degree of accuracy of the returned coordinates.

Step 3. *Extracting polygon surfaces*. Using a Geographic Information System, the geocoded addresses are overlaid on the polygons featuring parcel or building footprints. Then spatial join techniques are used to associate parcel and/or building polygons to addresses. Three different spatial join approaches are used to associate polygon areas to addresses. The output contains for each address, the associated polygon area from each of the three spatial join approach, as well as the distance between each associated polygon and the geographic coordinates of the address.

Step 4. *Extracting location characteristics*. Using a Geographic Information System, the geocoded addresses are overlaid on shapefiles of dissemination areas, Census Metropolitan Areas (CMAs), zoning restrictions, highways, seaports and airports to compute various location variables : population and surface area of dissemination areas and CMA, distance to closest seaport, freight station, airport and highway junction as well as dummies for zoning categories.

Step 5. *Creating a raw land variable*. This process compares the results of the three different spatial join approaches and finally assign to each address the 'best' result. Quality variables are constructed.

Step 6. *Creating the final dataset*. The Scotts dataset is merged with location characteristics and land measures to obtain the final dataset used in the paper.

B. Quality assessment.

Beyond the measurement challenges mentioned in the previous subsection, geocoding data and assigning them to polygons retrieved from satellite data inherently bring issues regarding the quality of the data and the methodology employed to assign plants to polygons.

First, there can be errors in the polygon datasets. Representing a parcel or a building by a polygon is subject to minor errors. For example, the algorithm used to convert satellite building images into polygon building outlines may fail in some cases to fit exactly the building into its representative polygon. The level of such errors—known as the matching precision—is estimated at 1.3% by the data provider.[23] This type of error only affects the building polygons. Parcel polygons are derived from administrative data and should, therefore, not be subject to measurement error of the type inherent to satellite data.

Second, there can be errors in the plant-to-polygon assignments. Geocoding microdata is an inherently noisy process. Even minor errors in the geocoding of plants can lead to their mis-assignment to polygons. To gauge the scope of false assignments in our dataset, we make use of the subset of data for the province of Quebec (QC). The reason is that the polygon identifiers in the QC dataset are the same as the official identifiers of the polygons as recorded on the governmental website of the land register "Infolot".[24] We can, therefore, randomly draw a set of addresses of plants in QC from our dataset and compare the parcel identifiers from "Infolot" to those obtained by our assignment procedure. Using a sample of 1,667 addresses, we find 1,320 correct

[23]See `https://github.com/Microsoft/CanadianBuildingFootprints` on the GitHub website where the data are released.

[24]On that website, it is possible to recover the identifier of a parcel by providing the address of a location. See `https://appli.mern.gouv.qc.ca/infolot/`.

assignments. Put differently, the probability for a plant in QC to be located exactly on its actual polygon is 79.16%.

Table A2: Assignment quality

Assignment quality	Parcel (PB)		Building (BB)	
	N	%	*N*	%
Excellent	8,782	78.83	22,978	96.43
Good	720	6.46	487	2.04
Poor	1,639	14.71	363	1.52
Total	11,141	100.0	23,828	100.0

Notes: Distribution of geocoded establishments in 2017 by quality categories. The classification includes the quality of the geocoding and the quality of the polygon assignment process. Concerning the geocoding quality, all observations with a less than excellent quality are removed, and the remaining are used to construct the three groups: excellent, good, and poor. The final sample we use is of excellent quality and missing values of covariates used in the regression analysis are removed. We have a sample of 8,708 parcels.

As explained in the part "Assignment to polygons" of the Appendix, the assignment of plants to polygons is based on three options that can potentially point to different polygons. Among the 1,667 addresses that we use for validation, if we restrict ourselves to the subset of observations for which the three options in the assignment procedure point to the same polygon, the share of correct assignments increases to 91.3%. In other words, plants for which the three assignment options point to the same polygons are very likely to be correctly assigned. Making use of that observation, we finally construct a 'quality' variable based on: (i) how accurate the geocoding of the establishment is; and (ii) how likely a correct assignment to a polygon is. This quality variable—which we construct for the whole dataset, not just Quebec—has three categories: excellent, good, and poor (see the part "Assignment to polygons" for more details). Table A2 summarizes the distribution of observations across data-quality categories for parcels and building footprint. In the remainder of the paper, unless noted otherwise, we only keep observations of 'excellent' quality.

C. Identifying city centers

To identify city centers, we identify clusters of densely populated dissemination areas (DAs, equivalent to 'census blocks') within each Canadian Census Metropolitan Areas (CMA) or Census Agglomerations (henceforth we use CMAs to mean either Census Metropolitan Areas or Census Agglomerations). To do so, we follow the procedure of

Behrens et al. (2020), who suggest an algorithm to construct clusters of manufacturing plants based on their spatial concentration. We simply replace plants by densely populated DAs.

Formally, for each CMA, we identify the clusters of densely populated DAs as follows:

- we flag all DAs with population density greater than the third quartile of the population density distribution of the CMA;
- we draw a circle with 500 meters radius around each flagged DA and compute the hypergeometric probability of having the number of flagged DAs in that circle, given the overall number of flagged DAs in the CMA. We also compare the total population of the flagged DAs within the circle to the total population of the flagged DAs in the CMA;
- A DA is considered a focal point of population concentration if the hypergeometric probability we computed is below 1% and if the ratio of the total population of the flagged DAs in the circle compared to the total population of the flagged DAs in the CMA is greater than the median observed in the CMA;
- we finally construct population clusters by drawing a buffer of 1 kilometer around each DA identified as a focal point and merging together all the overlapping buffers

City centers are the geographic centers of the population clusters identified following this procedure. Based on this method, we identify 223 city centers across Canada. There are six centers in the Toronto CMA, three centers in the Montréal CMA, and one in the Vancouver CMA.

Using these city centers, for each plant, we compute the average of the distance between the plant and the centers of its CMA using as weights the population in a 500 meters radius around each center.

D. Additional tables and results

Table A3: Distribution of plants across industries in the final dataset

	Parcel		Building		Scotts Data	
	N	%	*N*	%	*N*	%
311 Food mfg	745	8.6	590	8.1	2,875	8.9
312 Beverage and tobacco product mfg	77	0.9	52	0.7	340	1.0
313 Textile mills	32	0.4	22	0.3	96	0.3
314 Textile product mills	227	2.6	178	2.5	743	2.3
315 Clothing mfg	307	3.5	212	2.9	712	2.2
316 Leather, allied product mfg	40	0.5	35	0.5	127	0.4
321 Wood product mfg	353	4.1	316	4.4	1,884	5.8
322 Paper mfg	149	1.7	126	1.7	501	1.5
323 Printing, support activities	726	8.3	555	7.6	2,270	7.0
324 Petrol, coal product mfg	20	0.2	15	0.2	134	0.4
325 Chemical mfg	433	5.0	381	5.3	1,539	4.7
326 Plastics, rubber products mfg	545	6.3	481	6.6	1,907	5.9
327 Non-metallic mineral product mfg	387	4.4	341	4.7	1,951	6.0
331 Primary metal mfg	125	1.4	103	1.4	537	1.7
332 Fabricated metal product mfg	1,301	14.9	1,129	15.6	5,226	16.1
333 Machinery mfg	1,061	12.2	955	13.2	4,542	14.0
334 Computer, electronic product mfg	300	3.4	262	3.6	1,032	3.2
335 Electrical, appliance mfg	256	2.9	228	3.1	784	2.4
336 Transportation equipment mfg	297	3.4	254	3.5	1,099	3.4
337 Furniture, related product mfg	421	4.8	351	4.8	1,392	4.3
339 Miscellaneous mfg	906	10.4	670	9.2	2,726	8.4
Total	8,708	100.0	20,443	100.0	32,417	100.0

Notes: This table reports the distributions of the Scottâs database along with our final sample for the two measures in 2017 across the different industries at the NAICS 3-digit level.

Table A4: Distribution of plants across provinces in the final dataset

	Parcel		Building		Scotts Data	
	N	%	N	%	N	%
Alberta	0	0.0	0	0.0	2,735	8.4
British Columbia	2,211	25.4	1,793	24.7	3,812	11.8
Manitoba	419	4.8	366	5.0	983	3.0
New Brunswick	214	2.5	193	2.7	708	2.2
Newfoundland	0	0.0	0	0.0	272	0.8
Nova Scotia	0	0.0	0	0.0	765	2.4
North West Territories	0	0.0	0	0.0	6	0.0
Nunanvut	0	0.0	0	0.0	6	0.0
Ontario	1,881	21.6	1,578	21.7	13,735	42.4
Prince Edward Island	0	0.0	0	0.0	144	0.4
Quebec	3,709	42.6	3,083	42.5	8,430	26.0
Saskatchewan	274	3.1	243	3.3	799	2.5
Yukon	0	0.0	0	0.0	22	0.1
Total	8,708	100.0	7,256	100.0	32,417	100.0

Notes: This table reports the distributions of the Scottâs database along with our final sample for the two measures in 2017 across the Canadian provinces. The three Territories $NorthWestTerritories, Yukon, and Nunavut$ have been removed because they contain few observations.

Table A5: Testing for selection on observable plant characteristics

Dep. var.:	$\mathbb{1}$ in sample			
	(1)	(2)	(3)	(4)
Ln Employment			-0.009	0.005
			(0.014)	(0.018)
Headquarter			0.046	-0.014
			(0.031)	(0.033)
Exporter			0.003	-0.011
			(0.053)	(0.041)
$\mathbb{1}$ Residential zoning		-0.152^a	-0.156^a	-0.114^a
		(0.057)	(0.058)	(0.042
$\mathbb{1}$ Recreational zoning		-0.605^a	-0.605^a	-0.329^a
		(0.091)	(0.092)	(0.079)
Ln City population				0.253^a
				(0.058)
Ln Population density 500m				0.085
				(0.053)
Ln Distance to closest major airport				0.025
				(0.076)
Ln Distance to closest major seaport				-0.043
				(0.093)
Ln Distance to closest freight station				0.071
				(0.058)
Ln Distance to closest highway junction				-0.035^a
				(0.021)
Fixed effects:				
4-digit industry	Yes	Yes	Yes	Yes
Province	No	Yes	Yes	Yes
Observations	24,457	24,457	24,457	24,457
Pseudo R^2	0.016	0.272	0.272	0.326

Notes: This table reports the estimates of a probit model where the dependent variable equals 1 if the establishment is in the estimation sample. $\mathbb{1}$ denotes $\{0,1\}$ dummy variables. Standard errors clustered at the city-level in parentheses. ^{c} $p < 0.10$, ^{b} $p < 0.05$, ^{a} $p < 0.01$.

Table A6: Plant-level parcel size by industry

	Parcel size			
	N	Mean	Median	CV
311 Food mfg	745	14,906.3	5,711.1	2.6
312 Beverage and tobacco product mfg	77	29,819.0	8,312.4	2.9
313 Textile mills	32	11,587.3	3,524.2	1.9
314 Textile product mills	227	7,509.2	4,045.0	1.4
315 Clothing mfg	307	6,649.3	3,853.6	1.3
316 Leather, allied product mfg	40	6,261.0	2,615.9	1.4
321 Wood product mfg	353	20,172.9	6,918.5	5.0
322 Paper mfg	149	23,797.7	13,359.4	1.9
323 Printing, support activities	726	7,996.7	3,486.8	2.1
324 Petrol, coal product mfg	20	22,928.1	9,315.0	1.5
325 Chemical mfg	433	18,558.6	8,213.5	2.7
326 Plastics, rubber products mfg	545	13,352.8	7,914.5	1.6
327 Non-metallic mineral product mfg	387	16,901.6	8,058.6	2.5
331 Primary metal mfg	125	39,747.3	6,253.9	5.3
332 Fabricated metal product mfg	1,301	11,494.9	5,565.2	3.1
333 Machinery mfg	1,061	12,131.6	6,599.6	2.1
334 Computer, electronic product mfg	300	10,580.8	6,467.8	1.5
335 Electrical, appliance mfg	256	13,318.3	8,416.8	1.4
336 Transportation equipment mfg	297	28,172.2	6,806.6	4.1
337 Furniture, related product mfg	421	9,238.9	4,546.8	2.9
339 Miscellaneous manufacturing	906	8,939.5	2,710.3	3.3
Total	8,708	13,354.2	5,757.6	2.7

Notes: This table reports descriptive statistics for parcel sizes across 3-digit industries for our final dataset.

Table A7: Plant-level parcel size per worker by industry

	Parcel size per worker			
	N	Mean	Median	CV
311 Food mfg	745	1,120.9	172.0	7.2
312 Beverage and tobacco product mfg	77	3,598.2	208.4	3.2
313 Textile mills	32	2,602.3	313.3	4.0
314 Textile product mills	227	1,503.2	351.6	2.2
315 Clothing mfg	307	661.5	218.8	1.8
316 Leather, allied product mfg	40	1,109.2	232.7	2.1
321 Wood product mfg	353	1,297.0	387.9	2.4
322 Paper mfg	149	967.9	389.4	2.0
323 Printing, support activities	726	1,512.4	333.2	3.6
324 Petrol, coal product mfg	20	1,249.1	693.3	1.3
325 Chemical mfg	433	1,204.2	368.4	3.3
326 Plastics, rubber products mfg	545	990.3	297.2	4.6
327 Non-metallic mineral product mfg	387	1,951.0	390.4	8.4
331 Primary metal mfg	125	787.7	295.3	2.0
332 Fabricated metal product mfg	1,301	1,134.9	316.7	3.9
333 Machinery mfg	1,061	1,063.1	327.1	2.9
334 Computer, electronic product mfg	300	1,060.5	302.9	3.3
335 Electrical, appliance mfg	256	955.8	297.7	2.2
336 Transportation equipment mfg	297	2,145.5	300.1	8.8
337 Furniture, related product mfg	421	1,564.4	316.4	7.8
339 Miscellaneous mfg	906	1,399.3	352.1	3.4
Total	8,708	1,281.0	310.3	4.3

Notes: This table reports descriptive statistics for parcel size per worker for our two land measures across 3-digit industrie. The sample is our final dataset.

Table A8: Building-to-parcel ratio by industry

	Building-to-parcel ratio			
	N	Mean	Median	CV
311 Food mfg	590	0.37	0.37	0.61
312 Beverage and tobacco product mfg	52	0.29	0.29	0.77
313 Textile mills	22	0.50	0.49	0.37
314 Textile product mills	178	0.35	0.35	0.60
315 Clothing mfg	212	0.43	0.42	0.51
316 Leather, allied product mfg	35	0.37	0.37	0.59
321 Wood product mfg	316	0.26	0.22	0.79
322 Paper mfg	126	0.38	0.38	0.52
323 Printing, support activities	555	0.40	0.38	0.54
324 Petrol, coal product mfg	15	0.09	0.07	0.77
325 Chemical mfg	381	0.32	0.31	0.63
326 Plastics, rubber products mfg	481	0.35	0.36	0.57
327 Non-metallic mineral product mfg	341	0.27	0.25	0.74
331 Primary metal mfg	103	0.30	0.31	0.60
332 Fabricated metal product mfg	1,129	0.34	0.33	0.59
333 Machinery mfg	955	0.31	0.28	0.60
334 Computer, electronic product mfg	262	0.34	0.31	0.62
335 Electrical, appliance mfg	228	0.34	0.34	0.52
336 Transportation equipment mfg	254	0.31	0.29	0.66
337 Furniture, related product mfg	351	0.37	0.37	0.55
339 Miscellaneous mfg	670	0.37	0.34	0.59
Total	7,256	0.34	0.33	0.60

Notes: This table reports descriptive statistics for the building-to-parcel-ratio across 3-digit industry. We have constrained the sample such that the parcel size exceeds the building footprint.