# AGGLOMERATION ECONOMIES: MICRODATA PANEL ESTIMATES FROM CANADIAN MANUFACTURING*

**John R. Baldwin**

*Economic Analysis Division, Statistics Canada, R.H. Coats Building, Ottawa, Ontario, Canada K1A 0T6. E-mail: john.baldwin@statcan.gc.ca*

**W. Mark Brown**

*Economic Analysis Division, Statistics Canada, R.H. Coats Building, Ottawa, Ontario, Canada K1A 0T6. E-mail: mark.brown@statcan.gc.ca*

**David L. Rigby**

*Department of Geography, University of California Los Angeles, Los Angeles, CA 90095. E-mail: rigby@geog.ucla.edu*

**ABSTRACT.** This paper identifies the main sources of urban increasing returns, after Marshall. The geographical distance across which externalities flow is also examined. We bring to bear on these questions plant-level data organized in the form of a panel across the years 1989 and 1999. Plant-level production functions are estimated across the Canadian manufacturing sector as a whole and for five broad industry groups, each characterized by the nature of its output. The panel data overcome selection bias resulting from unobserved plant-level heterogeneity that is constant over time. A related set of estimates using instrumental variables allay persistent concerns with endogeneity. Results provide strong support for Marshall's claims about the importance of buyer-supplier networks, labor market matching and spillovers. We show that spillovers enhance plant productivity within industries rather than between them and that these spillovers are highly localized.

## 1. INTRODUCTION

How much does geography matter to the performance of firms? Are the benefits of some locations as important to competitive advantage as the individual characteristics of business establishments themselves? What are the sources of increasing returns found in specific locations, and do these location-specific economies accrue to all types of establishments or only to some? How far do the benefits of colocation extend over space?

These questions have deep roots within urban and regional economics, extending back well over 100 years to the work of Marshall (1920). Prompted by his evocative discussion of the local "industrial atmosphere," our primary concern is identification of the sources of agglomeration and evaluation of their significance. In this endeavor, we follow a recent resurgence of interest in agglomeration, comprehensively reviewed by Rosenthal and Strange (2004) and Duranton (2007). This resurgence is driven, in part, by recognition of the continued importance of cities as centers of economic activity, in spite of recent innovations in transportation, information and communication technologies,

*This research was supported by a Canadian Studies Faculty Research Grant and by matching funds from the Microeconomic Analysis Division of Statistics Canada. We are indebted to Desmond Beckstead and Robert Gibson for research support both with regards to the longitudinal files used herein and the census and input/output calculations used to capture agglomeration effects.

Received: August 2008; revised: July 2009; accepted: August 2009.

 DOI: 10.1111/j.1467-9787.2010.00675.x

and the fragmentation of much production around the globe (Scott, 2001). Interest in agglomeration also reflects the increasing availability of georeferenced, microdata that allow researchers to interrogate new arguments about urban increasing returns in novel ways.

Standard theoretical claims about agglomeration hinge on geographical proximity. By locating in cities, firms have access to established pools of labor that are both specialized and deep, thus minimizing costs associated with search and training (Mincer, 1984; Barron et al., 1989; Becker and Murphy, 1992). In addition, the creativity and diversity of talent, associated with larger cities in particular, is seen to play a critical role in the generation and incubation of new businesses (Jacobs, 1969; Duranton and Puga, 2001). Cities, as centers of dense economic activity, also provide individual firms with abundant opportunities for the local sourcing of inputs and the distribution of output, thus reducing transportation costs (Chinitz, 1961). This same urban concentration of firms is thought to enhance the production of knowledge and its localized spillover, either through face-to-face exchange of tacit information, or through the interfirm mobility of human capital (Lucas, 1988; Jaffe et al., 1993; Rauch, 1993; Almeida, and Kogut, 1999).

Much of the early empirical analysis of agglomeration examined the general relationship between city size and productivity (Sveikauskas, 1975; Moomaw, 1981; Beeson and Husted, 1989). Moomaw (1983) and Gerking (1994) provide reviews of this work. Subsequent efforts have become increasingly refined and largely focused on discriminating between localization (own-industry) economies and urbanization (cross-industry) economies, or in a dynamic context between Marshall-Arrow-Romer (MAR) externalities and Jacobs externalities (Glaeser et al., 1992). Examining industry employment growth in U.S. cities between 1956 and 1987, Glaeser et al. (1992) find that regional specialization, a proxy for MAR externalities, has little effect, while city diversity, a proxy for Jacobs' externalities, boosts growth. They conclude that spillovers across industries are much more important at explaining employment growth than within industry spillovers, especially in mature cities. Focusing on eight manufacturing sectors across U.S. metropolitan areas between 1970 and 1987, Henderson et al. (1995) find that MAR externalities increase growth in mature capital goods industries, but that Jacob's externalities have no significant impact. They go on to show that in relatively immature, high technology sectors both types of externalities are present, but that with maturity and movement of these industries out of large, diversified cities, MAR economies predominate, consistent with the arguments of Duranton and Puga (2001). In more sophisticated efforts to control for unobserved heterogeneity, Black and Henderson (1999) and Henderson (2003) explore the nature of agglomeration economies using panel methods on plant-level data drawn from the U.S. Census Bureau's Longitudinal Research Database (LRD). Black and Henderson (1999) report no evidence of agglomeration economies of any kind in capital goods industries, though MAR economies operate within high technology sectors. Henderson (2003) reports similar findings.

A series of other empirical papers has focused on identification of the sources of agglomeration economies, after Marshall (1920). Thus, Dumais et al. (1997) show that industry employment growth in metropolitan areas is dependent upon the industry's demand for labor by occupation and the local distribution of workers across occupations. Little evidence was found to support arguments linking employment growth with knowledge spillovers or local buyer-supplier networks. Using similar data from the LRD, a recent extension of this work across more detailed industry groupings in U.S. cities finds much stronger support for all three Marshallian agglomeration factors—the labor mix, the local density of the buyer-supplier network and spillovers (Rigby and Essletzbichler, 2002). Rosenthal and Strange (2001) exploit Dunn and Bradstreet data in their exploration of the determinants of agglomeration. They regress an index of the spatial concentration of industries at different geographical scales on proxies for knowledge spillovers, labor pooling, input sharing and other factors. They find that labor pooling is significantly linked to industry concentration across scales while the influence of spillovers is limited to the zipcode (local) scale. In a subsequent paper, they show more carefully that localization economies attenuate rapidly with distance (Rosenthal and Strange, 2003).

Baldwin et al. (2008) link the arguments of a number of the papers discussed above using cross-sectional data on Canadian manufacturing establishments in 1999. Adapting measures of labor market matching and the buyer-supplier network from Rigby and Essletzbichler (2002) to the Canadian economy, they find broadly similar results. In their search for spillovers, they employ own-industry plant-counts to capture MAR economies, after Henderson (2003), and population size as a proxy for Jacobs' economies. Across the Canadian economy as a whole they find evidence of both types of externalities operating, though the localization-MAR effects are stronger. Like Rosenthal and Strange (2003) they show that the geographical range of own-industry externalities is limited, extending no more than 10 km.

Analysis in this paper also focuses on Canadian manufacturing plants. Canadian cities are generally smaller than those in the United States and therefore the sample offers the advantage of testing for the existence of externalities in smaller markets that have not yet exhausted agglomeration economies. Across these plants, we estimate production functions that embody two sets of arguments, one capturing plant characteristics, as well as those of parent firms, if applicable, the second capturing place-specific characteristics originally linked to economic performance by Marshall (1920).

The production functions are estimated using panel methods to overcome omitted variable bias common in cross-sectional analysis. The panel techniques require observations of individual manufacturing plants over time. We examine manufacturing establishments in 1989 and 1999. Only plants that were in business in these two years comprise our sample. The longitudinal panel also permits us to examine whether the relationships found in the cross-section, the result of developments that have occurred over many years, also exist in the short-run. That is, have changes in locational characteristics during the 1990s influenced changes in the productivity of plants. While earlier papers by Henderson (2003) and Rosenthal and Strange (2003) use panel techniques applied at the plant-level, those papers restricted their scope to only a few sectors of the economy. In this paper we present results for the Canadian manufacturing sector as a whole and for all major industry groups. The novelty of our analysis extends to the presentation of results from analysis using instrumental variables to address concerns about potential problems of endogeneity (see Henderson, 2003; Duranton, 2007). Finally, our investigation of the spatial extent of spillovers employs plant-level locational coordinates, geographic data of higher resolution than that used to date.

At the aggregate level, our results show that plant productivity is significantly influenced by the occupational and geographical distribution of workers, the density of the buyer-supplier network and the count of own-industry establishments (a proxy for the existence of intraindustry spillovers) within the region in which the plant is located. Indeed, the first two of these three agglomeration economies have elasticities that are larger than those for some plant/firm characteristics. These results are broadly consistent with earlier findings from cross-sectional investigation in the United States (Rigby and Essletzbichler, 2002) and Canada (Baldwin et al., 2008).

Following Rosenthal and Strange (2003) and Henderson (2003), we explore the geographical extent of the benefits that derive from the colocation of plants. For each establishment, we count the number of own-industry plants located within concentric circles of different radii. Across our full sample of Canadian manufacturing plants, there is a positive relationship between own-industry plant counts, our measure of localization

TABLE 1: Variable Description

| Variable Names | Description |
|---|---|
| *Plant Characteristics* | |
| Labor productivity | Value added divided by the number of production workers in the plant |
| Profit to value added ratio | Value added minus wages divided by value added |
| Production workers | Number of production workers |
| Nonproduction to production worker ratio | Nonproduction workers divided by production workers |
| *Place Characteristics* | |
| Labor mix | Defined in subsection 2.2 |
| Local density of upstream suppliers | Defined in subsection 2.2 |
| Plants within 5 km | Number of plants within 5 km in the same 2-digit industry |
| Plants within 10 km | Number of plants between 0 and 10 km in the same 2-digit industry |
| Plants within 50 km | Number of plants between 0 and 50 km in the same 2-digit industry |
| Plants within 200 km | Number of plants between 0 and 200 km in the same 2-digit industry |
| Population | Population of the census metropolitan area or census agglomeration where the plant is located |

economies, and productivity. The relationship is statistically significant for short distances (0 to 5 km), but not for longer distances. This result is consistent with Rosenthal and Strange (2003) who find the benefits of own-industry colocation attenuate rapidly with distance, and with Funderburg and Boarnet (2008) who detail the spatial concentration of employment across industrial clusters in Southern California. We find no productivity benefit of locating in regions with a larger population, our proxy for urbanization economies (Jacobs, 1969). In fact, the independent effect of city size tends to be negative, suggesting congestion effects.

Extending the analysis to five separate industry groups reveals that the different sources of agglomeration economies do not operate uniformly across the economy. The local density of buyer-supplier networks, the local labor mix, and intraindustry spillovers exerted a significant influence on productivity in three of five industry clusters. The significance of the count of own-industry plants varied across distance bands depending on the industry in question.

## 2. DATA AND MODEL

The variables used in our econometric models are readily separated into two groups, characteristics of individual business establishments or firms, and characteristics of particular locations. Table 1 lists the variables in our models and provides brief descriptions. The plant-level information is developed from the Canadian Annual Survey of Manufactures (ASM) for 1989 and 1999. The panel techniques we employ require observations on individual establishments for at least two years.

Our place-specific data are derived from the ASM, from the Household Census in 1991 and 2001 and from Canadian input-output accounts. All data were geocoded to a constant 2001 census geography for census metropolitan areas (CMAs) and census agglomerations (CAs). In 2001, there were 141 CMAs/CAs in Canada ranging in size from Kitimat, BC with a population of about 10,000 to the Toronto CMA with a population of about 4.6 million. The 141 regions contained approximately 80 percent of the Canadian population in 2001 and the same percentage of Canadian manufacturing establishments in 1999.

*Plant-Firm-Specific Characteristics*

The dependent variable in our analysis is labor productivity, measured as value added divided by the number of production workers. For each plant, we measure value added and production workers at their mean across three years. For 1989, these are the two adjacent years. Owing to the fact that 1999 is the last year on the longitudinal file, we take the mean level of value added and production workers for 1999 and the two previous years. Value added is measured in constant dollar terms using an industry-level deflator. We are interested in real not nominal growth in productivity because it is the former that tracks the volume component of productivity. In our analysis we remove industry fixed-effects by first-differencing productivity values over time. If we did not account for differences across industries in average inflation rates we would have reintroduced industry fixed effects that are related to the inflationary environment of each industry. We focus on manufacturing value added and production workers rather than the related measures of total value-added and total employment because the manufacturing-based measures of productivity are more closely tied to the location of a specific establishment. In measures of overall productivity (total value added divided by total employment), especially in multiplant firms, the origins and location of nonmanufacturing value added can be difficult to identify.

We utilize three-year means for value added and the number of production workers per plant, as well as all other plant-level characteristics, in order to reduce the year-over-year variability inherent to microdata of this kind. Plants often encounter shocks that may obscure the relationship between plant-level inputs and output (e.g., because of labor hoarding). Using three-year means helps to reduce the effect of this variability on our estimates.

Labor productivity is expected to depend on several plant-level characteristics. These include plant size, capital intensity, and the ratio of nonproduction to production workers. We anticipate that labor productivity will be higher in plants that are larger in size because they are able to take advantage of various forms of scale economies (e.g., those that result from longer production runs). Plant size is measured by the number of production workers.

The productivity of production workers is also expected to rise as the amount of machinery and equipment with which they work increases. We would like to capture the effect of mechanization with a variable measuring the capital to labor ratio. Unfortunately, capital stock data are unavailable at the plant level and so we use a proxy variable, the ratio of profits to value added, to represent the capital-labor ratio. In a competitive economy, profits just cover capital costs and therefore offer a good proxy for the relative capital embedded in different plants, especially when averaged across several years. Other studies that have used this proxy to estimate a production function using the Canadian manufacturing survey microdata have generated coefficients on capital that are very close to those derived from industry-level data on capital stock (see Baldwin and Gu, 2003). Profits are measured as value added minus wages.

Production workers tend to generate higher levels of output if more nonproduction workers are contributing to the production process. For instance, more input from management and engineering functions can help to improve the organization of the production process. Hence, we expect labor productivity to be positively associated with the ratio of nonproduction to production workers.

*Model*

The relationships between value added, plant size, and capital intensity noted above can be formally derived from a production function using Cobb-Douglas technology where value added (VA) is expressed as

$$VA = AK^{\alpha}L_{pw}^{\beta}L_{npw}^{\sigma}, \tag{1}$$

where $K$ is a measure of capital input, $L_{pw}$ is the number of production workers employed by the plant and $L_{npw}$ is the number of nonproduction workers. Henderson (2003) reports little difference in parameter estimation between Cobb-Douglas or more flexible forms of production function. We prefer the Cobb-Douglas form for it dampens potential problems of collinearity. Equation (1) may be rewritten such that labor productivity ($LP$) is a function of capital and labor inputs

$$LP = \frac{VA}{L_{pw}} = A\left(\frac{K}{L_{pw}}\right)^{\alpha}\left(\frac{L_{npw}}{L_{pw}}\right)^{\sigma}L_{pw}^{\beta+\alpha+\sigma-1}. \tag{2}$$

Hence, labor productivity is a positive function of the amount of capital employed per production worker, the number of nonproduction workers for each production worker, and the size of plant, as measured by the number of production workers. The Annual Survey of Manufacturers (ASM) provides measures of value added and the number of production and nonproduction workers. However, as already noted, it does not provide estimates of capital and therefore we need to develop a proxy($\hat{K}$). We estimate $\hat{K}$ from the following expression for profit($\pi$):

$$\pi = VA - wages = r\hat{K}, \tag{3}$$

where $r$ is the rate of return on capital. The profit to labor ratio $r\hat{K}/L_{pw}$ can be substituted into Equation (2), and if we assume the rate of return is equalized across plants, then

$$LP = Ar^{\alpha}\left(\frac{\hat{K}}{L_{pw}}\right)^{\alpha}\left(\frac{L_{npw}}{L_{pw}}\right)^{\sigma}L_{pw}^{\beta+\alpha+\sigma-1}. \tag{4}$$

Given this formulation, variation in profits across industries and provinces can be accounted for by industry and province fixed effects.

One of the practical issues with Equation (4) is that our proxy of the capital to labor ratio and our measure of productivity are very highly correlated because both contain value added in their numerator and labor in their denominator. To address this problem, we estimate a slightly different model that utilizes a capital to value added ratio rather than capital to labor

$$LP = \tilde{A}\tilde{r}\left(\frac{\hat{K}}{VA}\right)^{\frac{\alpha}{1-\alpha}}\left(\frac{L_{npw}}{L_{pw}}\right)^{\frac{\sigma}{1-\alpha}}L_{pw}^{\frac{\beta+\alpha+\sigma-1}{1-\alpha}}, \tag{5}$$

where $\tilde{A} = A^{1/(1-\alpha)}$ and $\tilde{r} = r^{\alpha/(1-\alpha)}$. Equation (5) can be used to solve for the values of $\alpha$, $\beta$, and $\sigma$.

In order to estimate Equation (5) we include a multiplicative error term $\varepsilon$ and use its logarithmic transformation

$$\ln LP_{ijk} = \ln\tilde{A} + \ln\tilde{r} + \delta_1\ln\frac{\hat{K}_i}{VA_i} + \delta_2\ln\frac{L_{npw\,i}}{L_{pw\,i}} + \delta_3\ln L_{pw\,i} + \ln\varepsilon_i, \tag{6}$$

where $\delta_1 = \frac{\alpha}{1-\alpha}$, $\delta_2 = \frac{\sigma}{1-\alpha}$, and $\delta_3 = \frac{\beta+\alpha+\sigma-1}{1-\alpha}$. Note also that $i$ indexes plants, $j$ indexes firms and $k$ indexes geographic locations.

Throughout the analysis, we assume that other characteristics of the firm and the characteristics of the location of the firm are transmitted through the multifactor productivity term $\tilde{A}$. Hence,

$$\ln \tilde{A} = a + \phi' \mathbf{lnX}_j + \theta' \mathbf{lnG}_k + \gamma_i + \eta_j + \lambda_k, \tag{7}$$

where $\mathbf{X}$ is a vector of characteristics related to the firm that controls plant $i$ and $\mathbf{G}$ is a vector of characteristics that are associated with location $k$. These locational characteristics are related either to the metropolitan area associated with $k$ or are calculated based on a set distance from $k$, where $k$ can be thought of as a point in space.

We measure two types of firm characteristics in the model. First, we identify whether the plant is part of a multiestablishment firm. This is a binary variable where the reference group is single-plant firms. Our expectation is that multiplant firms will be more productive than single plant firms. Multiestablishment status brings the benefit of firm-wide economies to the plant. For instance, multiestablishment firms may be better able to collect and analyze information that can improve management practices and thus raise productivity. Second, we identify whether plants are foreign controlled. Foreign-controlled plants are expected to have higher level of productivity because they have access to a broader range of experiences and technologies (Baldwin and Gu, 2005). Foreign control is also a binary categorical variable where the reference group is domestically controlled plants.

*Place-Specific Characteristics*

The agglomeration variables that we develop in our productivity model, the local density of buyer-supplier networks, labor pooling, and knowledge spillovers, can all be traced back to Marshall (1920). We outline below the variables employed to measure these Marshallian economies, along with indicators used to capture other types of agglomeration economies.

An area's labor pool supports the needs of a particular industry if the occupational distribution of an area corresponds to the distribution required by that industry. The labor mix for an industry within a metropolitan area is defined after Dumais et al. (1997) as

$$LABMIX_i^m = \sum_o \left( L_{io} - \sum_{j \neq i} \frac{E_j^m}{E^m - E_i^m} L_{jo} \right)^2,$$

where $o$ represents an occupation, $i$ and $j$ index industries and $k$ refers to the metropolitan area. $L$ measures the proportion of workers in a particular industry and occupation, while $E$ measures the number of workers in a single industry or in all industries within a metropolitan area. This index is a sum of squared deviations that measures the degree to which the occupational distribution of employment in an industry is matched by the occupational distribution of the workforce in the metropolitan area as a whole, excluding the specified industry. The occupational distribution of industry workers is calculated at the national level and covers some 47 occupations at the two-digit level using the 1991 Standard Occupational Classification, which is used for the 1991 and 2001 Censuses. We anticipate that a better match between the occupational distribution (demand) in an industry and the occupational distribution of the entire workforce of a metro area (supply) will boost productivity. Improved matches reduce the value of the squared term. Thus, we expect a negative coefficient on this variable in the regressions.

We calculate the benefits of the local density of buyer-supplier networks using national input-output data and indicators of the local concentration of production within specific sectors of the economy. These networks might convey additional benefits in the form of interindustry spillovers embodied in material flows between industrial sectors. High correlation between estimates of the geographic concentration of upstream producers and downstream customers led us to focus on upstream activity only. To measure local variation in the density of upstream connections for each four-digit industry and for each census metropolitan area in Canada, we identify an upstream supplier-weighted location quotient:

$$USXLQ_j^m = \sum_{i,i \neq j} w_{ij}^n \left( \frac{TVS_i^m / \sum_i TVS_i^m}{TVS_i^n / \sum_i TVS_i^n} \right).$$

The term in the parentheses is a location quotient for each industry $i$ in metro area $m$. The location quotients are calculated using the total value of shipments (*TVS*) of each industry and measure the degree to which a particular city is specialized in an industry. A value less than one would indicate an industry is underrepresented, while a value greater than one would indicate the industry was overrepresented. The terms $w_{ij}$ represents the weight of industry $i$ as a supplier of industry $j$, that is, the proportion of all manufactured input purchases by industry $j$ supplied by industry $i$. Supplier weights are estimated from interindustry transactions and are derived from the Canadian national input-output tables. The subscripts $i$ and $j$ refer to each of the 236 four-digit SIC manufacturing industries, $m$ refers to one of 140 or so metropolitan areas in Canada and $n$ refers to the nation. Note that we also removed the influence of the own-industry in these measures, by dropping the principal diagonal from the input-output direct coefficients matrix. Metropolitan areas whose economies are specialized in industries that are significant suppliers to industry $j$ will have a relatively high *USXLQ* and this is expected to have a positive effect on labor productivity in plants within industry $j$ within those areas. Rigby and Essletzbichler (2002) and Baldwin et al. (2008) have employed a similar measure in cross-sectional models of the sources of agglomeration, finding results broadly comparable with those reported below.

Note that because the labor mix and buyer-supplier network measures are defined at the metropolitan level, the values for these variables for a given industry are constant for all plants in that industry and metropolitan area. As we have noted above, this necessitates adjustment of the standard errors in our model, for as Moulton (1990) demonstrates, they can be biased when merging aggregate variables across microunits of observation.

The third agglomeration effect arises from knowledge spillovers that are generated by the close proximity of producers in the same industry in the same urban area—intraindustry spillovers. Measuring knowledge spillovers is notoriously difficult, even impossible as Krugman (1991) claims, for they do not leave a paper trail. Jaffe et al. (1993) disagree. They argue patent citations can be used to track the spatial limits of knowledge flows. Nevertheless, the linking of patent information to the plant-level data that are increasingly used to study agglomeration is surprisingly underdeveloped. Rigby and Essletzbichler (2002) show that flows of knowledge embodied in intermediate goods enhance the productivity of agglomerated plants, but that sheds little light on the role of disembodied information flows. We spent some time examining the influence of local own- and cross-industry patents, in industries of use and make, on plant labor productivity, but were discouraged by the results that were broadly insignificant. Our measures all used simple counts of patents within metropolitan areas and industries linked to the patent classification rather than citations. Raw patent counts for 1999, earlier years, or groups of years were not significantly related to productivity.

As a result, we follow Henderson (2003) and Rosenthal and Strange (2003) and use counts/densities of plants in specific geographical areas as a proxy for intraindustry knowledge spillovers. For geographical areas, we exploited data on the latitude and longitude of individual plants to define concentric circles of varying distances around each. The circles employed had radii of 0–5 km, 0–10 km, 0–50 km, and 0–200 km. We admit that these distances were chosen in an arbitrary fashion, since we do not have much theory to suggest over precisely what distances particular kinds of information actually flow. For each plant, we counted establishments within the same two-digit (SIC) industry. We anticipate that as the number of plants increases within a concentric circle of a given size, so too does the potential flow of knowledge and that is expected to boost plant productivity. However, we are unsure over what distances such economies are likely to flow and thus we experiment with circles of different size.

We add metropolitan population size to our model as a proxy for urbanization economies that are not captured elsewhere in our model. The benefits of urban size are many. Large urban economies bring with them greater industrial and occupational diversity that facilitate the transfer of new innovations across industries (Jacobs, 1969). Large population centers also create the demand for infrastructure that can enhance the productivity of all industries (e.g., highways, airports, ports, and communications networks). While metropolitan size is not a direct measure of diversity, our earlier work (Baldwin et al., 2008) using a Herfindahl index of urban employment diversity across industrial sectors produced no significant results.

*Sample Characteristics*

We limited our sample in several ways. By construction, plants in rural areas are excluded from the study. While rural plants account for almost 25 percent of manufacturing employment in Canada, it is very difficult to define labor markets and associated place-specific characteristics for these plants. Brown and Baldwin (2003) explore some of the differences between urban and rural manufacturing establishments in Canada. The sample is also restricted to plants with positive value added and positive returns to capital. For the later, this implies value added minus wages is greater than zero. As a practical matter these restrictions are imposed because logarithmically transformed variables with a value of zero or less are mathematically undefined. They are also imposed because plants with negative value added or negative returns to capital are likely undergoing significant economic shocks. This may blur the relationship between inputs and output. We also dropped from our sample a relatively small number of plants that changed location and/or industry between 1989 and 1999. Plant mobility can produce very large swings in the values of our independent variables, including our measures of agglomeration, especially for those establishments that move between labor markets with quite different characteristics. These movements exert a great deal of leverage on our parameter estimates, particularly in our panel specification where there are large numbers of relatively small changes in the values of our agglomeration variables. Limited analysis on the sample of plants that changed location and/or industry after 1989 did not reveal much consistency in terms of the model parameters estimated. For this reason we do not believe that excluding these plants biases our remaining sample in any particular way.

Due to the longitudinal nature of the analysis, the most significant restriction to our set of plants is that they must have lasted at least 10 years. In 1999, this restriction, plus all of the others noted above, reduced the number of plants in the sample from about 29,000 to 11,300. Omitting plants that do not remain in business at least 10 years significantly reduces the number of observations in our sample and raises questions about sampling bias. However, the results reported below are very similar to those published earlier on a much larger cross-section of plants from 1999 and we have found that they are robust to broad changes in sample characteristics (see Baldwin et al., 2008). Descriptive statistics

TABLE 2: Descriptive Statistics: Plants Present in 1989 and 1999

| | 1989 | | | | 1999 | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Median | Std. Dev. | Obs. | Mean | Median | Std. Dev. | Obs. |
| *Plant Characteristics* | | | | | | | | |
| Labor productivity | 82,775 | 57,910 | 113,862 | 11,323 | 87,298 | 55,644 | 112,083 | 11,323 |
| Profit to value added ratio | 0.58 | 0.58 | 0.16 | 11,323 | 0.58 | 0.58 | 0.18 | 11,323 |
| Production workers | 53 | 15 | 230 | 11,323 | 59 | 19 | 198 | 11,323 |
| Nonproduction to production worker ratio | 0.46 | 0.37 | 0.52 | 11,323 | 0.42 | 0.33 | 0.53 | 11,323 |
| *Place Characteristics* | | | | | | | | |
| Labor mix | 5.1 | 4.3 | 2.4 | 3,204 | 5.5 | 4.8 | 2.5 | 3,204 |
| Local density of upstream suppliers | 6.0 | 1.2 | 24.5 | 3,204 | 6.9 | 1.2 | 29.0 | 3,204 |
| Plants within 5 km | 41 | 17 | 74 | 11,323 | 31 | 13 | 54 | 11,323 |
| Plants within 10 km | 92 | 41 | 130 | 11,323 | 71 | 32 | 100 | 11,323 |
| Plants within 50 km | 371 | 191 | 449 | 11,323 | 275 | 162 | 323 | 11,323 |
| Plants within 200 km | 731 | 520 | 735 | 11,323 | 544 | 399 | 559 | 11,323 |
| Population | 159,220 | 37,932 | 463,249 | 138 | 178,011 | 39,992 | 535,224 | 138 |

*Source*: Annual Survey of Manufacturers (1989 and 1999).

and a correlation matrix for the dependent and independent variables are presented in Tables 2 and 3.

## 3. ECONOMETRIC STRATEGY

The primary econometric issue associated with the estimation of Equation (6) is the potential correlation of the error term with the left-hand-side variables. This correlation may stem from the presence of unobserved fixed effects and/or the endogeneity of one or more of the independent variables stemming from reverse causality. Unobserved fixed effects associated with plant $i$, its related firm $j$, and location $k$ may be correlated with our vector of geographical characteristics $\mathbf{G}$. These are represented in Equation (7) by $\gamma_i$, $\eta_j$, and $\lambda_k$, respectively. To illustrate the econometric problem, consider the case of a leading firm that was established by chance in a location several decades in the past. Over the ensuing decades, let us assume that the firm grew because of its superior production processes and product development (an unobserved characteristic). Often successful firms generate spin-offs as employees who developed technical and management expertise start their own businesses producing related products (see Klepper, 2002). With the development of the firm, and its spin-offs, local input suppliers emerge and the workforce of the local community is transformed, increasingly matching that of this geographic cluster of firms. In this case, we would observe a positive association between the level of labor productivity of the firms found in this cluster (the original firm and its spin-offs) and the mix of labor, the presence of upstream suppliers, and the number of firms in the

TABLE 3: Correlation Matrix

| Change In | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| (1) Labor productivity | 1 | | | | | | | | |
| (2) Profit to value added ratio | 0.61 | 1 | | | | | | | |
| | (0.00) | | | | | | | | |
| (3) Production workers | −0.26 | −0.14 | 1 | | | | | | |
| | (0.00) | (0.00) | | | | | | | |
| (4) Nonproduction to production worker ratio | 0.35 | 0.27 | −0.39 | 1 | | | | | |
| | (0.00) | (0.00) | (0.00) | | | | | | |
| (5) Labor mix | −0.21 | −0.02 | 0.00 | 0.02 | 1 | | | | |
| | (0.00) | (0.09) | (0.60) | (0.04) | | | | | |
| (6) Local density of upstream suppliers | 0.06 | 0.03 | 0.11 | 0.00 | −0.01 | 1 | | | |
| | (0.00) | (0.00) | (0.00) | (0.75) | (0.25) | | | | |
| (7) Plants within 5 km | 0.06 | 0.01 | 0.05 | 0.00 | −0.11 | 0.00 | 1 | | |
| | (0.00) | (0.20) | (0.00) | (0.62) | (0.00) | (0.99) | | | |
| (8) Plants within 200 km | 0.05 | 0.01 | 0.04 | −0.02 | −0.16 | 0.02 | 0.31 | 1 | |
| | (0.00) | (0.25) | (0.00) | (0.01) | (0.00) | (0.02) | (0.00) | | |
| (9) Population | −0.04 | −0.01 | 0.06 | 0.02 | −0.02 | −0.01 | −0.09 | 0.09 | 1 |
| | (0.00) | (0.28) | (0.00) | (0.03) | (0.00) | (0.40) | (0.00) | (0.00) | |

*Notes*: All variables were logged for each year 1989 and 1999 and then differenced. Terms in parentheses are $P$-values.

*Source*: Annual Survey of Manufacturers, 1989 and 1999.

same industry. This result is traceable not just to localization economies—labor matching, buyer-supplier links, and knowledge spillovers—but to the special nature of the progenitor firm. The same logic applies to geographic locations. The concentration of firms may be related to natural features (e.g., access to a resource stock) rather than any form of localization economy.

To address these issues, we substitute Equation (7) into Equation (6) and take the first difference across periods

$$\ln \Delta LP_{ijk} = \Delta a + \delta_1 \Delta \ln \frac{\hat{K}_i}{VA_i} + \delta_2 \Delta \ln \frac{L_{npw\,i}}{L_{pw\,i}} + \delta_3 \Delta \ln L_{pw\,i}. + \phi^t \Delta \mathbf{lnX_j} + \theta^t \Delta \mathbf{lnG_k} + \Delta \ln \varepsilon_i \tag{8}$$

In so doing, we eliminate the firm- and location-level fixed effects that might be correlated with our Marshallian localization economies. But of course we now give ourselves a harder task in isolating the influence of these fixed effects because there may be little change in the variables of interest and this will increase the standard errors of the estimates. For simplicity, we assume that the rate of return on capital is constant within plants across our two time periods and so this term is dropped in Equation (8).

Although differencing the data eliminates the correlation between the error term and the right-hand-side variables due to unobserved fixed effects, it does not eliminate correlation induced by reverse causality. That is, changes in our measures of Marshallian localization economies may be driven by a generalized productivity shock to plants within a particular location. For instance, falling barriers to trade may advantage some locations because of better access to external markets. The resulting generalized gains in productivity may attract more investment, raising the number of plants, inducing the development of more upstream production and, perhaps, improving the mix of labor as workers invest in skills related to this broad set of growing industries.

TABLE 4: General Model Estimates: Change in Plant-Level Productivity

| | Model 1 | Model 2 |
|---|---|---|
| *Change in Plant Characteristics* | | |
| Profit to value added ratio | 0.76 (<.001) | 0.75 (<001) |
| Production workers | −0.10 (<001) | −0.11 (<001) |
| Nonproduction to production worker ratio | 0.38 (<001) | 0.38 (<001) |
| Multiplant status (reference = single plant) | 0.10 (0.079) | 0.09 (0.002) |
| Foreign control status (reference = domestic control) | 0.10 (<001) | 0.09 (<001) |
| *Change in Place Characteristics* | | |
| Labor mix | | −0.51 (<001) |
| Local density of upstream suppliers | | 0.10 (<001) |
| Plants within 5 km | | 0.02 (0.001) |
| Plants within 200 km | | 0.02 (0.403) |
| Population | | −0.15 (0.045) |
| Constant | 0.04 (<001) | 0.05 (<001) |
| No. of observations | 11,323 | 11,323 |
| $F$ | 829 | 637 |
| Prob.> $F$ | <.001 | <.001 |
| $R^2$ | 0.42 | 0.47 |
| Root MSE | 0.45 | 0.43 |

*Notes*: All variables are log transformed, with the exception of binary variables, and differenced between the years 1989 and 1999. In all regressions, standard errors are corrected for heteroskedasticity and potential correlation of errors within CMA/CAs. *P*-values are presented in parentheses.

*Source*: Annual Survey of Manufacturers, 1989 and 1999.

We develop a set of instruments to account for the potential correlation of the error term with our measures of Marshallian localization economies. This is done primarily by lagging our Marshallian variables, in level form, back to 1979. These are measured as means for these variables across plants present in 1979 in each geographic unit. These lagged levels are unlikely to be correlated with contemporaneous shocks to labor productivity. However, they may be correlated with changes to our suspect independent variables if there is a tendency for regression to the mean. Depending on the specification, we also use as instruments lagged levels of labor productivity and changes in labor productivity over the previous period (1979 to 1989).

## 4. PANEL MODEL ESTIMATES

### *First Difference Estimates*

We estimate different forms of Equation (8). Our key results are reported in Tables 4–6. Table 4 presents output from two first difference models. These models were estimated using ordinary least squares after differencing between years. All standard errors are robust and corrections have been made for potential correlation of errors between manufacturing establishments found in the same region (Moulton, 1990). Model 1 shows the relationship between labor productivity and plant characteristics alone. As expected, labor productivity tends to be significantly higher in plants where the profit/value added ratio, our proxy for the capital-labor ratio, is high. This variable consistently displays the largest elasticity of all independent variables, typically raising productivity over 8 percent for every 10 percent increase in the profit-value added ratio. As the ratio of nonproduction workers to production workers rises across plants, so productivity also tends to increase. The elasticity of this variable is typically less than half that of the profit/value added ratio. The coefficients on these plant characteristic variables yield sensible estimates of the coefficients of the production function for the manufacturing sector—the implied labor share is 0.51 ($\beta = 0.29$ and $\sigma = 0.22$), capital share ($\alpha = 0.43$) and there are near constant returns to scale ($\alpha + \beta + \sigma = 0.94$).

Plants that belong to multiestablishment firms also display higher productivity values than single establishment firms, and foreign-owned plants are more productive than domestic plants. In both cases, these effects are significant and the elasticity on foreign-owned plants is about the same as that for multiplant firms.

Model 2 adds our agglomeration measures. We examine three mechanisms of agglomeration after Marshall, a measure of labor pooling, a measure of the local density of buyer-supplier (input-output) networks, and a measure of spillovers. In addition, we use metropolitan population size as an indicator of more general urbanization economies.

Following Henderson (2003), we measure spillovers using own-industry plant counts. Because we have latitude and longitude data for each establishment, we can count the number of neighboring plants, own-industry or not, at varying distances around a given plant location. We attempt to estimate the distance across which spillovers flow by counting plants within circles of progressively greater radii and relating those counts to productivity. Table 4 shows that the productivity of an individual establishment increases with the number of plants in the same industry that are located within 5 km. When we add to the model, plant counts for distances up to 10 km, up to 20 km, up to 50 km, and up to 200 km, no additional significant results are found. Hence, like others, we find a strong distance gradient with respect to intraindustry spillovers. Earlier experiments with cross-industry plant counts and with measures of neighboring own-industry and cross-industry employment produced no interesting findings. Similarly, analysis of patent counts by industry and metropolitan area, as a measure of localized knowledge production, showed no significant relationship to productivity. The elasticity for own-industry plant counts within 5 km that we report is broadly consistent with those obtained by Henderson (2003) using own-industry plant counts by U.S. county, though it is considerably larger than those reported by Rosenthal and Strange (2003) using own-industry employment across zip code centroids. This difference might result from our separation of the different forms of agglomeration in analysis of the microgeography of spillovers.

Our measure of the local density of the buyer-supplier network also exerts a positive and significant impact on the labor productivity of individual establishments, with an elasticity of 0.1. Thus a doubling of the density of the metropolitan input-output network would increase plant productivity by about 10 percent on average. This elasticity is about an order of magnitude larger than that we reported earlier based on a cross-section of plants for 1999 (Baldwin et al., 2008).

By far the most significant agglomeration measure in terms of impact on plant productivity is our labor-matching variable. This is consistent with earlier results. Indeed, the partial regression coefficient on the labor mix variable shows that plants located in metropolitan areas where the occupational distribution of workers is closely related to the occupational distribution of their own workforce have significantly higher labor productivity. A 10 percent improvement in this occupational match raises plant productivity approximately 5 percent. We note here that labor-matching has the second largest influence of all variables in our model, including plant and firm characteristics as well as place-characteristics. This is a significant finding and speaks to the potential significance of agglomeration economies. Dumais et al. (1997) also report that labor pooling is the most significant of Marshall's agglomeration economies driving colocation.

Finally, our measure of urbanization economies, population size, exerts a significant, negative influence on plant productivity, perhaps reflecting congestion costs. The elasticity on this measure of urbanization economies is relatively large at 0.15.

This negative influence of metropolitan size on productivity has not generally been found in the literature, though we caution that, in analysis across individual industries, our measure of urbanization economies is insignificant. Indeed, Glaeser et al. (1992) report a positive impact of urbanization economies on employment growth, and mixed evidence on the importance of urbanization economies are reported by Rosenthal and Strange (2003) and Henderson (2003).

*Instrumental Variable Estimates*

While first differencing accounts for the impact of unobserved fixed effects on our estimates, it does not account for the influence of reverse causality. To address this problem, we utilize two kinds of instrumental variables (IV) estimators—two-stage least squares (2SLS) and generalized methods of moments (GMM). We utilize the GMM estimator because it is more efficient than 2SLS when errors are not i.i.d. (Woolridge, 2001). Since we have a large sample, our estimates should not be affected by the poor small sample properties of the GMM estimator.

Table 5 shows the IV results for two models, with both the 2SLS and GMM estimates reported for each. Our instruments include lags of the Marshallian agglomeration variables, the mean level of labor productivity in 1979 and the change in labor productivity between 1979 and 1989 at the two-digit industry level across metropolitan areas. The two models presented only differ in the first stage, according to which instruments are included. Model 1 includes the change in mean labor productivity between 1979 and 1989 and the mean level of labor productivity in 1979. These two variables were excluded from Model 2 because overidentification tests (Hansen's $J$ statistic) suggested (weakly) that one or both of them were correlated with the error term. This leads us to place more confidence in the estimates from Model 2. Overall, partial-$R^2$ statistics suggest that our instruments are relatively weak, though for all of them $F$-statistics are above 10, a general indicator of their usefulness.

The estimates provided in the second stage of our instrumental variables analysis, and focusing on Model 2, show that there is little change in the coefficients for plant characteristics from those presented in Table 4. The same is also true of the coefficient on population, a place characteristic. For the instrumented variables, there was a general tendency for the estimates here to reinforce those derived from the original panel model (see Table 4, Model 2). That is, the estimates are consistent in terms of sign and statistical significance. However, the absolute values of the IV elasticities were larger than those reported in our panel model. Henderson (2003) reports similar results. Larger coefficients from our IV estimators might be attributed to the power of our instruments. However, experimentation with the LIML estimator, that is less susceptible to problems associated with weak instruments (Stock and Yogo, 2005), provided similar point estimates. The exception to the pattern of consistency between the first difference and IV estimates was the number of plants within 200 km. Here the value of the coefficient switched from positive and insignificant, to negative and significant.

In general, the IV estimates are broadly consistent with those provided by the first difference estimator, though they are larger in value. Our interest in the IV results largely stems from concerns with potential endogeneity in attempts to isolate the influence of agglomeration on economic performance (see Duranton, 2007). Given the results presented in Tables 4 and 5, we feel confident that endogeneity bias is not compromising our general findings. That said, we place more emphasis on the results reported in Table 4, because of remaining concerns over the quality of our instruments.

TABLE 5: Instrumental Variable Estimates: Change in Plant-Level Productivity

| | Model 1 | | Model 2 | |
|---|---|---|---|---|
| | (1) | (2) | (1) | (2) |
| *Change in Plant Characteristics* | | | | |
| Profit to value added ratio | 0.73 (<.001) | 0.66 (<.001) | 0.73 (<.001) | 0.74 (<.001) |
| Production workers | −0.13 (<.001) | −0.17 (<.001) | −0.12 (<.001) | −0.12 (<.001) |
| Nonproduction to production worker ratio | 0.38 (<.001) | 0.32 (<.001) | 0.40 (<.001) | 0.40 (<.001) |
| Multiplant status (reference = single plant) | 0.06 (0.050) | 0.05 (0.022) | 0.06 (0.041) | 0.09 (<.001) |
| Foreign control status (reference = domestic control) | 0.06 (0.002) | 0.04 (0.009) | 0.07 (0.001) | 0.05 (0.003) |
| *Change in Place Characteristics* | | | | |
| Labor mix[1] | −0.91 (<.001) | −0.99 (<.001) | −1.19 (<.001) | −1.31 (<.001) |
| Local density of upstream suppliers[1] | 0.63 (<.001) | 0.75 (<.001) | 0.38 (0.049) | 0.42 (0.003) |
| Plants within 5 km[1] | 0.15 (0.019) | 0.18 (<.001) | 0.10 (0.032) | 0.10 (0.003) |
| Plants within 200 km[1] | −0.21 (0.106) | −0.35 (<.001) | −0.24 (0.077) | −0.23 (0.023) |
| Population | −0.13 (0.349) | −0.13 (0.029) | −0.20 (0.055) | −0.14 (0.090) |
| Constant | 0.000 (0.986) | −0.017 (0.402) | −0.014 (0.656) | −0.013 (0.566) |
| Estimator | 2SLS | GMM | 2SLS | GMM |
| No. of observations | 10,615 | 10,615 | 10,615 | 10,615 |
| $R^2$ | 0.34 | 0.28 | 0.36 | 0.33 |
| $F$ (Prob.>$F$) | 602 (<.001) | 903 (<.001) | 578 (<.001) | 757 (<.001) |
| Root MSE | 0.47 | 0.50 | 0.47 | 0.48 |
| Hansen $J$ statistic ($\chi^2$ $P$-value) | 18.1 (0.113) | 18.1 (0.113) | 9.4 (0.498) | 9.4 (0.498) |

| | Model 1: First-Stage | | | Model 2: First-Stage | | |
|---|---|---|---|---|---|---|
| *First-stage* | Partial $R^2$ | $F$ | Prob. > $F$ | Partial $R^2$ | $F$ | Prob. > $F$ |
| Labor mix | 0.21 | 85.9 | <.001 | 0.13 | 35.7 | <.001 |
| Local density of upstream suppliers | 0.03 | 9.6 | <.001 | 0.03 | 10.0 | <.001 |
| Plants within 5 km | 0.05 | 31.7 | <.001 | 0.03 | 33.8 | <.001 |
| Plants within 200 km | 0.25 | 55.0 | <.001 | 0.25 | 67.4 | <.001 |
| *Included Instruments* | | | | | | |
| Change in labor productivity 1979 to 1989 | Yes | Yes | | No | No | |
| Labor productivity in 1979 | Yes | Yes | | No | No | |
| Labor mix in 1979 | Yes | Yes | | Yes | Yes | |
| Local density of upstream suppliers in 1979 | Yes | Yes | | Yes | Yes | |
| Plants within 5 km in 1979 | Yes | Yes | | Yes | Yes | |
| Plants within 200 km in 1979 | Yes | Yes | | Yes | Yes | |
| Provincial fixed effects | Yes | Yes | | Yes | Yes | |

[1]Instrumented variables.

*Notes*: With the exception of provincial fixed effects, all included instruments are measured for geographic units at their mean values in levels for plants present in 1979 or changes in mean levels for plants existing 1979 and those existing in 1989. All variables are log transformed, with the exception of binary variables, and differenced between the years 1989 and 1999. In all regressions, standard errors are corrected for heteroskedasticity and potential correlation of errors within CMA/CAs. $P$-values are presented in parentheses.

*Source*: Annual survey of manufacturers, 1979, 1989 and 1999.

*Sectoral Estimates*

The results from our general model, presented in Table 4, show that, on average, agglomeration economies raise the productivity of individual producers. However, there is no guarantee that the benefits of colocation are equally important for all businesses in all industries. One simple way of exploring this question is to examine how the different sources of agglomeration economies operate across manufacturing industries. In the Canadian context, this approach is problematic for two reasons. First, the number of plants within most three or four digit (SIC) industries in Canada is quite small and so it is difficult to obtain statistically significant results. Second, it would also be difficult to make sense of results that stretch across hundreds of sectors. To overcome this problem, we follow a different tack, aggregating individual manufacturing industries together into five broad sectors and then estimating our first-difference model across each of those sectors. In another paper, Brown and Rigby (2009) explore how agglomeration economies interact with a variety of establishment and firm characteristics.

The five sectors are taken from the Organization for Economic Cooperation and Development (OECD, 1987). They are defined as natural resource-based, labor-intensive, scale-based, product-differentiated, and science-based. The original OECD classification was tailored for use with Canadian manufacturing data. Baldwin and Rafiquzzaman (1994) list the four-digit (SIC) industries assigned to each of the OECD sectors. Each sector is defined primarily on the basis of the factors that influence the process of competition. For resource-based industries, the primary determinant of competitive success is access to abundant natural resources. For the labor-intensive sector, it is labor costs. For scale-based industries, competition hinges on the length of production runs. In the product-differentiated group, competition depends on an ability to target production to the demands of various markets. Finally, competition in science-based sectors depends on the application of scientific knowledge.

Table 6 shows the results from estimating our fixed-effects panel model for each of the five OECD sectors. Overall, plant characteristics affect labor productivity in a consistent way across these five industry groupings, though the sizes of the partial regression coefficients are variable. Productivity increases with the capital-labor ratio, with plant size, and with the ratio of nonproduction to production workers. Our firm measures, multiplant status and domestic/foreign ownership status, have the same positive sign across all OECD sectors, though the coefficients vary in size markedly and are not uniformly significant. Plants that are part of multiestablishment firms have higher productivity than single-establishment firms, though the productivity differential is statistically significant only in natural resource-based and science-based OECD sectors. Similarly, while foreign-owned plants have higher productivity than domestically owned plants, the difference is significant only in scale-based and science-based sectors.

Our three key Marshallian agglomeration economies, the labor mix, the density of upstream suppliers, and the own-industry plant count within 5 km, our proxy for spillovers, all have signs consistent with theoretical expectations across the five OECD economic sectors. The coefficient on our labor mix, or labor matching variable, suggests that an improved match between the supply and demand of workers by occupation improves productivity, though that effect is only statistically significant in scale-based, product-differentiated, and science-based sectors. In these three industries, the labor mix variable is the second most important influence on plant productivity of any variable measuring plant, firm or place characteristics. In the science-based sector, the labor mix variable has the highest elasticity confirming the importance of key labor inputs in knowledge intensive production. Our earlier cross-sectional analysis produced broadly similar results, though the patterns of significance of the labor mix variable differed across OECD

TABLE 6: Sectoral Model Estimates: Change in Plant-Level Productivity

| | Natural Resource-Based | Labor Intensive | Scale-Based | Product-Differentiated | Science-Based |
|---|---|---|---|---|---|
| *Change in Plant Characteristics* | | | | | |
| Profit to value added ratio | 0.85 (<.001) | 0.69 (<.001) | 0.65 (<.001) | 0.82 (<.001) | 0.91 (<.001) |
| Production workers | −0.11 (<.001) | −0.12 (<.001) | −0.13 (<.001) | −0.09 (0.002) | −0.11 (<.001) |
| Nonproduction to production worker ratio | 0.31 (<.001) | 0.32 (<.001) | 0.46 (<.001) | 0.41 (<.001) | 0.49 (<.001) |
| Multiplant status (reference = single plant) | 0.12 (<.001) | 0.10 (0.118) | 0.02 (0.331) | 0.03 (0.558) | 0.13 (0.005) |
| Foreign control status (reference = domestic control) | 0.04 (0.158) | 0.06 (0.131) | 0.09 (0.019) | 0.16 (0.106) | 0.19 (0.017) |
| *Change in Place Characteristics* | | | | | |
| Labor mix | −0.09 (0.096) | −0.17 (0.112) | −0.53 (<.001) | −0.58 (<.001) | −0.66 (<.001) |
| Density of upstream suppliers | 0.10 (<.001) | 0.06 (0.040) | 0.19 (<.001) | 0.06 (0.309) | 0.07 (0.219) |
| Plants within 5 km | 0.011 (0.455) | 0.001 (0.927) | 0.027 (0.014) | 0.014 (0.393) | 0.034 (0.040) |
| Plants within 200 km | −0.20 (0.022) | 0.15 (0.017) | 0.15 (0.015) | −0.18 (0.069) | 0.22 (0.045) |
| Population | −0.10 (0.328) | −0.11 (0.171) | −0.31 (0.101) | −0.16 (0.269) | 0.30 (0.175) |
| Constant | 0.08 (0.002) | 0.08 (0.001) | 0.10 (0.001) | 0.02 (0.310) | 0.03 (0.456) |
| No. of observations | 3,028 | 2,933 | 2,545 | 2,012 | 805 |
| $F$ | 202 | 415 | 198 | 141 | 97 |
| Prob.> $F$ | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| $R^2$ | 0.47 | 0.50 | 0.49 | 0.41 | 0.52 |
| Root MSE | 0.42 | 0.37 | 0.42 | 0.47 | 0.47 |

*Notes*: All variables are log transformed, with the exception of binary variables, and differenced between the years 1989 and 1999. In all regressions, standard errors are corrected for heteroskedasticity and potential correlation of errors within CMA/CAs. *P*-values are presented in parentheses.

*Source*: Annual Survey of Manufacturers, 1989 and 1999.

sectors. We have greater faith in the results reported here because omitted variables in the cross-sectional analysis may have generated biased and inconsistent estimators.

Across all OECD sectors, the density of the regional buyer-supplier network exerts a positive, albeit not consistently significant, influence on plant productivity. The density of upstream suppliers is significant for all but the product-differentiated and science-based industries. The elasticity on the local buyer-supplier network variable is about twice as high in the scale-based sector than in the natural resource sector and it is about twice as high in the resource sector as the labor intensive sectors of the economy.

Results on the own-industry plant count variables are somewhat more mixed. We believe that the plant count within the 5 km distance band is a measure of intraindustry spillovers, though it is less clear what the plant count within the 200 km distance band is capturing. At no other distances did plant counts generate significant impacts on productivity. We focus here on our measure of spillovers. Intraindustry spillovers raise productivity significantly in the scale-based and science-based sectors. It is in the science-based industries where this effect is most prominent, though the elasticity on our measure of spillovers is relatively modest. Across the five different industry groups, our measure of urbanization economies, metropolitan population size, was not significantly related to plant productivity.

## 5. CONCLUSION

Over the last decade or so, a number of researchers have employed plant-level, microdata to examine the existence and the veracity of agglomeration economies. The research community has moved well beyond identification of a generic class of increasing returns using crude proxies such as population size. Estimating models of economic performance based on plant-level observations has allowed investigators to control for individual plant characteristics and to assess the extent to which locational attributes influence plant performance. We extend this line of research estimating a panel model using 1989 and 1999 data to overcome problems associated with omitted variable bias. A remaining concern regarding endogeneity prompted use of instrumental variables techniques. We find no evidence that endogeneity compromises our attempts to understand the relationship between plant performance and the characteristics of the places where those plants are located.

Regardless of estimation strategy, from the cross-sectional analysis of an earlier paper (Baldwin et al., 2008) through the instrumental variables estimates reported above, our results provide strong support that different types of agglomeration economies improve the performance of manufacturing establishments, independent of plant and firm characteristics. In general, we find broad support for the positive impacts of Marshall's three mechanisms of agglomeration—labor pooling, the local input-output network, and spillovers. At an aggregate level we find that population size is negatively related to productivity, a finding at odds with some past work. When we move to the level of individual industries, urbanization economies have no significant influence on plant performance.

By far the most important of the agglomeration variables appears to be the labor mix, the consistency of the match between the local supply and demand for labor across occupations. In general, this agglomeration variable was more important to plant performance than all plant and firm characteristics except our proxy for the capital-labor ratio. This finding is clear affirmation that geography matters. We also find that the local density of the buyer-supplier network exerts as much influence on plant productivity as establishment size, multi- and foreign-plant status, though within individual industrial sectors this argument must be tempered.

Exploration of the distance across which economies accrue from local concentrations of own-industry plant counts was performed by counting numbers of plants within distance bands of varying radii around each individual plant observation. The results here suggest that the benefits of intraindustry colocation extend no more than 5 km. While we are hesitant to argue that 5 km denotes a precise distance threshold above which interplant economies are suddenly truncated, these results are consistent with our earlier claims, and those of Rosenthal and Strange (2003), that such benefits are highly localized.

Two further questions related to the influence of agglomeration economies on plant performance are identified here. The first has to do with the interaction between agglomeration economies and plant and firm characteristics. Dumais et al. (1997), Henderson (2003) and Rosenthal and Strange (2003) begin this exploration, while Brown and Rigby (2009) push a little further exploring how external economies are related to the varying capacity of individual establishments to generate competitive advantage internally. A second question asks how our understanding of the relationship between performance and the characteristics of plants and locations might change if we had better information regarding the quality of the labor force employed within individual businesses. The rapid development of matched employer-employee data should soon permit analysis of this issue, along with the influence of labor mobility on plant productivity.

## REFERENCES


Almeida, P. and B. Kogut. 1999. "Localization of Knowledge and the Mobility of Engineers in Regional Networks," *Management Science*, 45, 905–917.

Baldwin, J. R., D. Beckstead, W. M. Brown, and D. L. Rigby. 2008. "Agglomeration and the Geography of Localization Economies in Canada," *Regional Studies*, 42, 117–132.

Baldwin, J. R. and W. Gu. 2003. "Participation in Export Markets and Productivity Performance in Canadian Manufacturing," *Canadian Journal of Economics*, 36, 634–657.

———. 2005. "Global Links: Multinationals, Foreign Ownership and Productivity Growth in Canadian Manufacturing," *The Canadian Economy in Transition*, No. 9. Ottawa: Statistics Canada.

Baldwin, J. R. and M. Rafiquzzaman. 1994. "Structural Change in the Canadian Manufacturing Sector (1970–1990)," Research Paper No. 61, Analytical Studies Branch, Research Paper Series. Ottawa: Statistics Canada.

Barron, J. M., D. A. Black, and M. A. Lowenstein. 1989. "Job Matching and On-The-Job Training," *Journal of Labor Economics*, 7, 1–19.

Becker, G. S. and K. M. Murphy. 1992. "The Division of Labor, Coordination Costs and Knowledge," *Quarterly Journal of Economics*, 107, 1137–1160.

Beeson, P. E. and S. Husted. 1989. "Patterns and Determinants of Productive Efficiency in Manufacturing," *Journal of Regional Science*, 29, 15–28.

Black, D. and V. Henderson. 1999. "Spatial Evolution of Population and Industry in the United States," *American Economic Review*, 89, 321–327.

Brown, W. M. and J. R. Baldwin. 2003. "The Changing Geography of the Canadian Manufacturing Sector in Metropolitan and Rural Regions, 1976–1997," *The Canadian Geographer*, 47, 116–134.

Brown, W. M. and D. L. Rigby. 2009. "Marshallian Localization Economies: Where Do They Come From and To Whom Do They Flow?" California Center for Population Research, Working Paper # 2009-020. University of California, Los Angeles.

Chinitz, B. 1961. "Contrasts in Agglomeration: New York and Pittsburgh," *American Economic Review*, 51, 279–289.

Dumais, G., G. Ellison, and E. L. Glaeser. 1997. "Geographic Concentration as a Dynamic Process," Center for Economic Studies Working Paper # 98-3, US Bureau of the Census. Washington, DC.

Duranton, G. 2007. "California Dreamin': The Feeble Case for Cluster Policies," Unpublished manuscript, Department of Economics, University of Toronto.

Duranton, G. and D. Puga. 2001. "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products," *American Economic Review*, 91, 1454–1477.

Funderburg, R. G. and M. G. Boarnet. 2008. "Agglomeration Potential: The Spatial Scale of Industry Linkages in the Southern California Economy," *Growth and Change*, 39, 24–57.

Gerking, S. 1994. "Measuring Productivity Growth in US Regions: A Survey," *International Regional Science Review*, 16, 155–185.

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer. 1992. "Growth in Cities," *Journal of Political Economy*, 100, 1126–1152.

Henderson, V. 2003. "Marshall's Scale Economies," *Journal of Urban Economics*, 53, 1–28.

Henderson, V., A. Kuncoro, and M. Turner. 1995. "Industrial Development in Cities," *Journal of Political Economy*, 103, 1067–1090.

Jacobs, J. 1969. *The Economy of Cities*. New York: Vintage.

Jaffe, A. B., M. Trajtenberg, and R. Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 108, 577–598.

Klepper, S. 2002. "The Capabilities of New Firms and the Evolution of the U.S. Automobile Industry," *Industrial and Corporate Change*, 11, 645–666.

Krugman, P. R. 1991. "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99, 483–499.

Lucas, R. E. 1988. "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22, 3–42.

Marshall, A. 1920. *Principles of Economics* (8th ed.). London: Macmillan.

Mincer, J. 1984. "Labor Mobility, Wages, and Job Training," *Mimeograph*. New York: Columbia University.

Moomaw, R. L. 1981. "Productivity and City-Size: A Critique of the Evidence," *Quarterly Journal of Economics*, 95, 675–688.

———. 1983. "Spatial Productivity Variations in Manufacturing: A Critical Survey of Cross-Sectional Analyses," *International Regional Science Review*, 8, 1–22.

Moulton, B. R. 1990. "An Illustration of the Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *Review of Economics and Statistics*, 72, 334–338.

Organization for Economic Co-operation and Development. 1987. *Structural Adjustment and Economic Performance*. Paris.

Rauch, J. E. 1993. "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities," *Journal of Urban Economics*, 34, 3–33.

Rigby, D. L. and J. Essletzbichler. 2002. "Agglomeration and Productivity Differences in US Cities," *Journal of Economic Geography*, 2, 407–432.

Rosenthal, S. S. and W. C. Strange. 2001. "The Determinants of Agglomeration," *Journal of Urban Economics*, 50, 191–229.

———. 2003. "Geography, Industrial Organization and Agglomeration," *Review of Economics and Statistics*, 85, 377–393.

———. 2004. "Evidence on the Nature and Sources of Agglomeration Economies," in V. Henderson and J. Thisse (eds.), *Handbook of Regional and Urban Economics*, vol. 4. Amsterdam: North-Holland, pp. 2119–2171.

Scott, A. J. 2001. *Global City-Regions: Trends, Theory and Policy*. Oxford: Oxford University Press.

Stock, J. H. and M. Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression," in D. W. K. Andrews and J. H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor or Thomas Rothenberg*. Cambridge: Cambridge University Press, pp. 80–108.

Sveikauskas, L. 1975. "The Productivity of Cities," *Quarterly Journal of Economics*, 89, 392–413.

Wooldridge, J. M. 2001. "Application of Generalized Method of Moments Estimation," *The Journal of Economic Perspectives*, 15, 87–100.