Are There Urbanization Economies in a Post-Socialist City?

Evidence from Ukrainian Firm-Level Data

December 29 2010

By Volodymyr Vakhitov
Kyiv School of Economics
Kyiv Economic Institute

13, Yakira St., Room 319
Kyiv 04119 Ukraine

Email: vakhitov@eerc.kiev.ua

Are There Urbanization Economies in a Post-Socialist City? Evidence from Ukrainian Firm-Level Data

Abstract

In this work I test a set of hypotheses broadly stated by Jacobs (1969) that diversity of firms, individual innovation activities, inter-industry knowledge spillovers and developed urban amenities are essential for a successful development of a city and critical for an individual firm's productivity. The estimation framework is summarized in Rosenthal and Strange (2004) and generally is based on estimating the total factor productivity. Even though the idea has been around for quite a while, and many researchers attempted various estimation techniques to test the stated hypotheses, two factors tell my paper apart. First, we managed to collect a very rich firm-level panel which includes not only the standard variables for the TPF estimation like output, capital and labor measures, but also innovation and FDI activities of firms, individual firm demands for intermediate outputs from other sectors, as well firms' ownership structure. I believe this additional information will not only help us precisely measure the standard diversity variables, but also provide additional insights into interactions between firms, which are always stressed in the related literature. Second, recent advances in econometrics analysis (such as Olley and Pakes (1996) semi-parametric procedure for production function estimation or Blundell-Bond (1998) type estimators of dynamic effects) might help to control for specific micro data issues such as endogeneity and simultaneity bias. Additional attention is be devoted to particularities of a post-socialist urban landscape, which had determined the initial location of industries in cities, and effects of such location for firms' productivity.

Are There Urbanization Economies in a Post-Socialist City? Evidence from Ukrainian Firm-Level Data

Introduction

One of the questions which has long interested both economists and economic geographers was "Does the location of economic activities matter and how?" The earliest papers may be attributed to as far back as Alfred Marshall (1890) who observed geographic concentration of several industries in England and tried to explain this phenomenon by such factors as knowledge spillovers, as well as sharing the common markets for resources and the final output. He also noticed that firms in the same industry not only tend to concentrate, but seem to perform better than stand-alone firms. Marshall's theory was amended by Arrow (1962) and Romer (1986) and after seminal formalization in Glaeser et al. (1992) became widely recognized as Marshall-Arrow-Romer model of dynamic externalities within an industry. On the other hand, Jacobs (1969) argued that the most important factor for a successful urban development is the sharing of ideas not within, but rather between the industries, since only through a constant act of imitating and adding improvements to the existing works innovation and new activities emerge, which together provide sustainable economic growth of a city. The author supported her thesis through numerous empirical conjectures and historical facts when cities with one or a few dominated industries thrived in some periods in time, but eventually faded after their leading industries became obsolete, whereas more diversified cities continued to develop, sometimes slowly though steadily.

The empirical literature distinguishes between Marshall and Jacobs types of externalities based on the "own industry involvement" into the process of knowledge spillovers. Marshall externalities, also known as localization economies, are considered to be external to the firm, but internal to the industry, whereas Jacobs externalities, or "urbanization economies" are said to be external both to the firm and to the industry. It is important to note also that both types of agglomeration economies are assumed to be localized in the same city, or a region in a broader context, where the firm in question is located.

The dichotomy between urbanization and localization economies has been generally accepted in the empirical literature. The major empirical question is then 'How to measure the agglomeration economies in the most "correct" way?' To tackle this issue, three major concerns should be addressed: industry definition, region definition, and measuring issues.

One of the most important issues is to define the frontiers of the "industry-location" cell. What do we imply by "an industry"? What is the most appropriate level of industrial aggregation? Where are the borders of the industry? How the industry should be determined – by the similarity of the production process between firms or by the similarity of their output? How should one treat multiproduct firms, multinational firms or, as a particular case, firms having branches all over the country? These issues are very difficult to deal with empirically, because the data are notoriously scarce. At the same time, there is no a single operationable measure of a "distance" between industries. Conley and Dupor (2003) suggested one of a few attempts to describe interactions between industries based on the detailed input-output matrices for U.S. economy. The authors developed their "distance" measure and found that sectors which share suppliers have higher correlation of growth rates. Ellison et al. (2007) used a simpler measure and found that input-output relationships between firms are important for co-agglomeration, or mutual location of firms from similar industries. Apparently, the input-output analysis is the most feasible for this aim. Unfortunately, IO matrices should be disaggregated at a very fine level (for example, Ellison et al. (2007) used six-digit industry codes). Currently, such data are available for only a few countries. Also, interactions between industries are not necessarily based on the input-output relations. Ellison et al. (2007) also used cross-industry patent citations to control for knowledge spillovers. To account for the common labor pool, the authors also looked at the switches of the specialists trained for a particular task between various industries, using the CPS data.

On the other hand, the geographic scope of interfirm relations seems more straightforward to deal with since the physical distances are much easier to measure. The data availability remains an issue, though. For example, Duranton and Overman (2008) managed to build a distribution of exact distances between British firms and came to a conclusion that agglomeration effects are indeed present, but they tend to attenuate at a rather quick pace and disappear yet after 30 km. Rosental and Strange (2003) applied a concentric-circles approach trying to capture the agglomeration effects for newly established U.S. firms at various distances and also found that the effects attenuate rather quickly. Both studies used advanced GIS software to measure the precise distances and, while proposing interesting results, can hardly be mimicked by most researchers due to high data requirements. National nondisclosure regulations usually prevent from identifying an exact geographical location of a firm.

The most pertinent issue for any student of agglomeration economies is the correct measuring of the effects per se. In the most recent review, Beaudry and Schiffauerova (2009) analyzed a number of empirical papers devoted to the topic. They differentiated the agglomeration measures into three broad groups: describing *share*, *size* and *diversity* of an industry. The "size" variables in the localization case measure the total employment or a number of plants in the same region-industry cell or in the related industries. The "diversity" measures included indices based on "technological closeness", or concentration indices. Standard location quotients (as the ratio of the share of an industry in the region relative to the share of the region in the economy) were used to describe the "share" variables. For estimation of urbanization economies, the authors point to a greater variety of measures. These include Hirschman- Herfindahl index, Gini index, technological similarity index and "related variety" as "diversity" measures, and the number of plants or employment in the "rest of the region", or at "other industries", or at "service industries" as the "size" measures. The choice of a particular measure of the agglomeration effects may be determined first of all by the data availability and the aggregation levels.

Beaudry and Schiffauerova (2009) draw an important methodological conclusion from their analysis: The actual estimates of the agglomeration effects are generally not robust to the choice of a particular measure or the level of aggregation. The estimation results seem to vary between the studies where different measures were used. For example, when the urbanization effects are measured with the "size" variables, the effects seem to be statistically insignificant more often compared to the case when "diversity" measures are used. On the other hand, "other industry employment" most often shows negative effect for Jacobs type externalities, whereas "diversity measures" suggest positive effects. At the same time, the urbanization effects revealed themselves most often in the studies where the data were geographically aggregated at the level of a municipality or even zip-code areas and at the level of a three-digit industry or lower across the industrial dimension. The first finding apparently follows the micro-geographic papers (such as Duranton and Overman (2007) about the extreme locality of the effects. Presence of urbanization externalities for finer industrial aggregation may also be explained by a rapid increasing of the set outside of the "own industry-region cell" where Marshallian effects were initially present. In other words, the firm's environment which under greater aggregation levels is treated as "own industry" becomes "external" when we slide down the aggregation scale. Methodologically it looks like the "localization" effects turn into "urbanization" ones. Therefore it is difficult even theoretically to distinguish between localization and urbanization effects. At the same time the knowledge spillovers between firms are difficult to measure. Hopefully, the variety of different measures will bring researchers closer to understanding the nature of agglomeration effects.

The discussion of agglomeration economies estimation will not be complete without a consideration of the ownership issues. The underlying theory of urbanization economies presumes that land plots are allocated between industrial and dwelling use, and further distributed among the individual firms and households based on a market price of the land and the expected productivity or utility that the land plot may bring to the owner. If the urban land market is absent, as it was the case in the former post-socialist countries, the true value of land is not taken into account during the allocation process. As a result, the long-term allocation of land has brought about multiple anomalies (compared to a market-based case) which may not necessarily be soon corrected by the market forces. In one of the first studies of the "post-socialist city" phenomenon, Bertaud and Renaud (1997) describe such features as virtually absent recycling of the land which cannot be used in the initial capacity (rather, new "rings" of land are developed), too large share of urban land used for manufacturing, or a reverse population density gradient. Once a large firm gets established in such a city it tends to develop a local labor market around itself through providing housing and public infrastructure to its workers. This mono-industrial cluster even if initially located at the outskirts of a city soon "moves into a core" as the city grows. As a result, large adjacent blocks of the city become immensely specialized in a certain industry. Many company towns in the former Soviet Union were formed along this path. However, the process of crowding out old firms with new firms does not seem to work due to the size issues and very expensive relocation costs of the labor force. This phenomenon is known as the "lock in effect" when both firms and the labor force becomes too specialized and tied to the major firm, and hence less mobile. As a result, if an industry becomes obsolete, there are no incentives to relocate it outside of the city or to replace with a newer process due to severe social effects and low mobility of labor. Such a setup is interesting for urban economists in a sense that in post-socialist cities the urbanization effects may appear to be lower, whereas localization effects may be greater compared to cities in market economies. Another testable hypothesis states that such cities either die out completely with time or manage to diversify away from the major industry, thus showing greater urbanization effects for their industries over time.

From the perspective given above, measuring urbanization effects seems a very tricky task. As Rosenthal and Strange (2004) propose, one has to take into account interactions of a firm with other firms in the geographical and industrial space. Ideally, these interactions should also be studied over an extended period of time. The authors conjecture that keeping other things equal, firms which are closer to each other from the geographical, industrial or temporal points of view should demonstrate a greater degree of spillovers and hence greater reliance on such spillovers for own productivity.

In my work I will follow the general methodology proposed by Rosenthal and Strange (2004) and attempt to estimate the urbanization effects within the production function framework. I hope to find modest but increasing urbanization effects for locked-in industries (those with high level of geographical concentration) and greater effects for industries in more diversified cities. On the other hand I will try to stick to the initial idea of Jacobs (1969) who particularly stressed innovation activities and replacing old works with new ones for urbanization economies to reveal themselves. For this aim, I will take into account not only the standard size or diversity measures, such as the employment outside the own industry or various share indicators, but also try to incorporate such variables as the innovative activity in the city, the share of the foreign ownership, and the share of the inputs a firm uses in the production process which are also produced in the same city.

I also compare the estimates of the effects for various levels of industrial and geographical aggregation. The data in hand will let me estimate the effects separately for several groups of similar industries, or industrial sectors. This will distinguish my work from most other studies where pooled estimations were used.

The rest of the paper builds as follows. In the next section I present the data I use. In section three I present a formal model and estimation strategy. The last section concludes.

Data description

For this paper I use the firm-level data of Ukrainian firms for the period of 2001-2007. The choice of this particular period is determined by the data availability and also by the fact that 2001 was the first year when the Ukrainian economy demonstrated a steady growth after a long slump in the nineties. The data come from annual statistical reports, which firms submit to the National Office of Statistics (Derzhkomstat). Each firm has the following basic statistics the production function estimation: a firm's ID, territory and industry codes, output (measures as total sales net of excise taxes), employment (measured as the equivalent of the full-time workers over the year), use of materials and intermediary inputs, and capital as the value of the total assets. The firm's age as the time since the first record of the firm's registration is also available. Statements on FDI will let construct the share of foreign ownership. Annual innovation expenditures are available at the level of a firm and can be used to estimate the innovation activity in the firm's industry and in the location overall. Data from the State Property Fund and current ownership status may provide valuable insights about the former Soviet ownership of the firms and the share of Soviet ownership in the industry to control for the lock-in effect. Each firm reports the values of intermediate outputs acquired from other sectors. Even though these sectors are defined rather broadly it is still possible to estimate the share of inputs available from the own industry vs. imported from other locations. Hopefully this measure will also help to capture the degree of relationship between firms in the same city.

The total number of firms in the database varies between 350 and 450 thousand firms per year representing all commercial firms. Banking institutions and budgetary establishments, such as hospitals, theaters, public administration, etc. are excluded from the database. Only up to 50 thousand firms are big, established enterprises, whereas the rest are small to medium ones (SMEs thereafter). The attrition level of the SME sample is much higher compared to the big firms sample, whereas the life expectancy of small firms is usually shorter. On the one hand, this hampers the estimation of the long-term relationships between firms, but on the other hand it helps better understand the churning process. Therefore it might be instructive to incorporate the size effect into estimation of agglomeration effects. Lafourcade and Mion (2007) noticed a difference in agglomeration effects for Italian firms of different size.

I will perform the analysis only for the manufacturing firms (KVED/NACE sector "D"), using other sectors for constructing necessary controls. This choice is determined by the fact that it is usually quite difficult to determine the output for services to estimate their productivity properly. The total number of manufacturing firms is approximately 15% of the total firms in the economy, whereas they employ almost one third of the labor force. Over 70% of all manufacturing firms in Ukraine are located in cities. I restrict my analysis only to the manufacturing firms in urban locations.

I present the sample composition in Table 1.

Estimation issues

As a working econometric model I use the approach presented in Rosenthal and Strange (2004). The authors describe a standard methodology which is based on the production function estimation and broadly used in the literature. The agglomeration effects enter the estimated equation as a Hicks-neutral technological shift:

$$y_j = g(A_j)f(x_j)$$

In their benchmark model the authors put additional structure on the agglomeration component $g(A_j)$, $A_j = \sum_{k \in K} s(x_k, x_j)\theta(d_{ij}^G, d_{ij}^I, d_{ij}^T)$, which accounts for distances between firms in the geographical, industrial, and temporal space.

Early agglomeration studies could only use the data of rather large degree of aggregation. For example, Nakamura (1985) used the industry data aggregated at the level of two-digit subsectors and large territorial areas (Japanese prefectures). Svejkauskas (1975) explored two-digit industries in the US SMSA areas. Since only cross-section data were available, most authors resorted to the estimation of trans-log or CES production function specification. Both authors reported significant urbanization effects at these levels of data aggregation, but also pointed out that the effects differ for various industries. Other things equal, the economy of scale seems to be an important factor: larger, older and more capital intensive industries seem to benefit more from localization effects, whereas lighter industries which are presumably more mobile and quicker in implementing the innovations enjoy the urbanization effects.

More recent studies, such as Henderson (1986) used firm-level data. The aggregation levels used were three- to four digits across the industrial dimension and a city (or MSA) in the geographical aggregation. The author had acquired an access to several periods of observations which made it possible applying fixed-effects. The findings contradict those of many earlier studies: the urbanization effects seem to be negligible for most sectors (except for such sectors as printing, nonmetallic minerals, and, to some extent, textile), whereas localization effects appear to be significant, especially in cities specialized in those industries. In his later paper, Henderson (2003) compared high tech industry group with machinery (treating the group as a composition of several three-digit industries). He also found no evidence of urbanization effects. However, the author explained these findings by a possible lack of data on business services inputs. He refered to Ciccone and Hall's (1996) suggestion about the omitted variable bias due to implicit character of relationships between firms and possible outsourcing of certain business services to other firms in the same city, which increases with the city size. On the other hand, the author conjectured that industry mobility issues and scale effects may interfere the results.

One important issue daunts practically all estimations of agglomeration effects. The choice of a particular firms' location may affect its productivity, which creates significant endogeneity problem. As Henderson (2003) puts it, the plant inputs and environment variables must me strictly exogeneous to the estimation error term. A standard way to tackle this problem has long been using instrumental variables. However, the issue of good instruments remains vital. Henderson (2003) applied the lagged values of environment variables (size measures of agglomeration effects) based on the assumption that they were uncorrelated with the error terms in time *t*. Also he controlled for plant and location fixed effects as well as for national time-industry fixed effects. In the recent paper with a similar setup, Mayer et al. (2008) also indicate a possibility of unobserved heterogeneity and endogeneity and apply both IV and GMM approaches to address such issues.

I use two techniques: simple fixed effects with clustered robust standard errors to control for possible heterogeneity issues, and Olley-Pakes methodology. As a robustness check, I also a standard within estimator with location fixed effects rather than firm-level fixed effect. The main reason for this is the fact that firms seem to change their production sector during the observation sample, and do this quite often. This is especially true for the small firms, which constitute about 90% of the sample. For our estimation sample, there were 5900 cases of manufacturing firms changing their major output sector (KVED) at least once between 2001 and 2005. Small firms constituted 89% of all firms changing their major KVED code. On the one hand, when this happens it violates initial assumption of having "fixed effects" within a firm. On the other hand, such firms' behavior implicitly supports churning processes Jacobs (1969) mentioned: firms constantly experiment, choose the best they can and thereby increase their productivity (or, at least, their chances of survival). In principle, the share of firms changing their sector can also be considered as one of possible measures of the urbanization processes within a city. In several specifications I also use Blundell-Bond GMM-SYS estimator.

Estimation Results

In a spirit of Henderson (2003) I estimate a standard production function for a given establishment *i* from an industry *j* in a location *m* at year *t*:

$$\ln(y_{jt}) = \alpha \ln(\mathbf{X}_{jt}) + \beta \ln \mathbf{E}_{jt} + \gamma_1 \ln \mathbf{I}_{jt} + \phi_{mj} + \delta_t \times \eta_{jt} + \varepsilon_{mjt}$$

where y_{jt} is the firm's sales net of excise and value added taxes, **X** is the vector of production factors (employment, capital and materials), **E** is a vector of environment variables, such as size or diversity measures of agglomeration, and **I** is a vector of institutional variables, such as ownership. This equation incorporates the cross product of time (δ_t) and industry (η_j) fixed effects. In all specifications I used clustered robust standard errors. The analysis is run initially at the level of two digits industrial aggregation, and then at the level of three digits.

Results of my basic fixed effects specification are presented in Table 2, whereas Olley-Pakes specification results are shown in Table 3. Following Beaudry and Schifforova (2009) taxonomy I used such standard estimators of urban economies as the "size" measures (represented by the employment and total number of firms in the city) and "diversity" measures (employment based and firms count based Hirshman-Herfindahl indices (HHI) were used for this aim).

First of all I would like to comment on my production factors coefficients. In all specifications coefficients on labor, capital and materials are strongly significant and have expected values as, for example, in Henderson (2003). On the other hand, the coefficients also go in line with Brown et al (2006) who used a very similar dataset of Ukrainian firms.

Unlike in Henderson (2003), size effects of urbanization economies turn out to be strongly significant and positive both for urban employment and firms count agglomeration measures. On the other hand, when the agglomeration effects are combined with the firm size, it appears that the larger a firm is, the smaller are the agglomeration effects. This finding is significant and robust across all measures of agglomeration and all econometric specifications. It implies that larger firms are probably benefit from internal scale economies to a greater extent compared to external scale economies.

One of possible reasons for such a result could be an excessive size of firms. The number of firms may increase if the average firm size falls. The average firm size in the estimated sample is 104 employees. However, a closer analysis of the firm size suggests that absolute majority of firms are rather small: over 80% of them have less than 50 employees. There are only 1745 observations when the employment exceeds 1000, which amounts to 350 firms per year on the average. I will further investigate if this is the big firms that drive the "employment" effect negative.

The diversity measures (with HHI = 1 indicating the highest concentration) suggest that diversity is beneficial for the firms' productivity. The diversity in terms of the number of firms indicates slightly greater values than diversity of employment.

Beaudry and Schiffauerova (2009) indicate that in most studies only one of the agglomeration measures was used: the size or the diversity measure. Jacobs (1969) provides a number of illustrative examples when both factors may be important: among cities of the similar size more diverse ones show sustainable growth, whereas in larger cities one can find more opportunities for a separate niche, thereby increasing the general diversity as well. Therefore I decided to incorporate both measures in the same equation, as shown in Table 4. This specification is run with location rather than firm fixed effects, and the agglomeration measures are taken in log form. These changes have almost no effect on validity of production function coefficients, but efficiency of the urban employment coefficient has suffered, without the sign change. Such setting may indicate not only urbanization economies, but also the level ща competitiveness within the city. However, this issue requires further investigation.

One more variable I suggest may be of a special interest. My data let me calculate the index of "intra-firm diversity". Even though a firm is listed in the registries under the category of its major output, it usually produces goods in several industry categories. Those outputs are disaggregated at the level of two digit industry groups. The dataset lets finding the exact share of output for each such group. The sum of the squared shares of output within the firm is the index I use. Inclusion of this index into the estimation suggests that intra-firm diversity is even more important for productivity than diversity within the city. In other words, if a firm is producing only one output, the chances of knowledge spillovers between it and other firms decrease, which is potentially detrimental for the total productivity growth. Since roughly two thirds of the firms produce outputs in the single group, diversification of output may be an important policy implication of this finding.

Finally, I incorporate two more agglomeration variables which may provide useful implications for further policy work. First, I incorporate a pair of indicator variables to account for the major ownership type of the firm. The variables are DO and FO, indicating, respectively, that the major share of the firm is owned by a private domestic or a private foreign entity. If both variables are simultaneously zeros, the firm is predominantly state owned. The results are presented in Table 5. The results suggest that private firms are clearly more productive than state owned, and that majority foreign firms are even more productive than majority domestic ones. In my further analysis I will try to establish the link between a particular agglomeration measure and the ownership variable to study possible connections between ownership and ability to use the agglomeration benefits to the full extent.

In line with Jacobs (1969) I also included innovation measures into the final estimated equation. The last two columns in Table 5 report this result. It turns out that just a fact that there are innovating firms in the city does not affect the firm's productivity. What is really important is the total amount of innovations in the area. Ideally, the proper measure would be not just current innovation flows, but also the past innovation stock, but this variable is not available at present. On the other hand, the lagged values of the innovation variables turned out to be statistically insignificant in all specifications.

I have also repeated this analysis at a different level of industrial aggregation when calculating diversity indices and industry-year fixed effects (the results are not shown since they are basically the same). In the second run I used the aggregation at the level of three digits of KVED. This change mostly kept the results unaffected with one major exception: the effect of the total number of firms in the city has halved in comparison with KVED2 aggregation. This may suggest that KVED2 is probably too broad level of data aggregation, and urbanization effects measured at this level may also incorporate possible localization effects which happen to occur within the same broadly defined industry. However, as Beaudry and Schifforova (2009) assert, there is no consent among the researchers about the best level of industrial aggregation of the data.

Conclusion

The work indicates that urbanization economies do exist in Ukrainian manufacturing, and the size of the effects is comparable to those found in the West. At the same time, since the former Soviet firms tended to be large, and many of them keep the relative size until now, the effect of agglomeration is diminished for larger firms. Most likely, the internal scale economies are "crowded out" by internal processes. The diversification measures provide mixed evidence, but most likely both internal and external diversification at the level of the firm play an important role for productivity increase. This is another important agglomeration results, since diversification is not possible in a monocity environment. In other words, the more diversified a firm is, and the more diversified the industry in the city is, the higher is the productivity of all firms in this industry-city cluster. Finally, ownership plays an important role for agglomeration effects: private domestic firms benefit more from agglomeration, and private foreign firms benefit the most.

List of Tables

Year	Firms with nonzero output and employment			Manufacturing firms	Firms in the sample*
	Small	Big	Total		
2001	145,964	14,996	160,960	25,697	20,168
2002	158,848	14,506	173,354	27,183	21,525
2003	166,581	14,208	180,789	27,606	21,820
2004	170,204	14,853	185,057	27,656	22,414
2005	173,362	15,521	188,883	27,801	22,658
Total	**814,959**	**74,084**	**889,043**	**135,943**	**108,585**

Table 1 (Sample Composition)

	(1)	(2)	(3)	(4)	(5)	(6)
ln(Capital)	0.062***	0.062***	0.062***	0.062***	0.062***	0.062***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
ln(Employment)	0.487***	0.487***	0.505***	0.502***	0.488***	0.489***
	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.008)
ln(Mat. Cost)	0.440***	0.440***	0.440***	0.440***	0.439***	0.439***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Total urban employment	0.116**		0.229***			
	(0.037)		(0.050)			
Total urban firms count		0.003**		0.007***		
		(0.001)		(0.002)		
Urban Empl. x Empl.			-0.045***			
			(0.011)			
Firm Count x Empl.				-0.002***		
				(0.000)		
HHI					-0.013	-0.003
					(0.024)	(0.053)
HHI x Empl						-0.003
						(0.014)
Industry-Year Cross Term	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.63	0.63	0.63	0.63	0.63	0.63
Obs	153759	153759	153759	153759	149398	149398
Firms	42426	42426	42426	42426	41681	41681

* p<0.05, ** p<0.01,*** p<0.001

Table 2 (Basic specification, fixed effects)

	(1)	(2)	(3)	(4)	(5)
ln(Capital)	0.100***	0.100***	0.100***	0.100***	0.097***
	(0.007)	(0.010)	(0.010)	(0.011)	(0.008)
ln(Employment)	0.390***	0.401***	0.390***	0.402***	0.391***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)
ln(Mat. Cost)	0.490***	0.490***	0.490***	0.490***	0.487***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.004)
Total urban employment	0.406***	0.491***			
	(0.060)	(0.057)			
Total urban firms count		-0.030***			
		(0.005)			
Urban Empl. x Empl.			0.011***	0.015***	
			(0.002)	(0.002)	
Firm Count x Empl.				-0.001***	
				(0.000)	
HHI					0.013
					(0.025)
Industry-Year Cross Term	Yes	Yes	Yes	Yes	Yes
Obs	74229	74229	74229	74229	74229
Firms	26942	26942	26942	26942	26942

Table 3 (Olley-Pakes Specification)

	Employment	Firms	Internal diversity
ln (Employment)	0.409***	0.410***	0.405***
	0.012	0.012	0.011
ln(Capital)	0.032**	0.031**	0.026**
	0.013	0.012	0.012
ln(Materials)	0.570***	0.569***	0.568***
	0.02	0.02	0.02
Intrafirm diversity index			-0.515***
			0.025
ln(Urban employment)	0.024		
	0.104		
ln(Urban # of firms)		0.352***	0.355***
		0.089	0.08
Employment - based HHI	-0.108***		
	0.029		
Firm-based HHI		-0.145***	-0.152***
		0.036	0.036
Observations	108723	108723	108585
Number of clusters	178	178	178
R-squared	0.86	0.86	0.86

Table 4 (Combined size and diversity measures)

	Ownership		Ownership + Innovations	
ln (Employment)	0.417***	0.416***	0.417***	0.417***
	0.012	0.012	0.012	0.012
ln(Capital)	0.024**	0.025**	0.026**	0.026**
	0.012	0.012	0.012	0.012
ln(Materials)	0.563***	0.564***	0.561***	0.561***
	0.02	0.02	0.02	0.02
Intrafirm diversity index	-0.529***	-0.528***	-0.533***	-0.534***
	0.025	0.025	0.025	0.025
ln(Urban employment)		0.019		
		0.096		
ln(Urban # of firms)	0.332***		0.363***	0.388***
	0.078		0.064	0.077
Employment - based HHI		-0.107***		
		0.029		
Firm-based HHI	-0.154***		-0.166***	-0.165***
	0.035		0.038	0.038
ln(Number of innovating firms)				-0.01
				0.008
ln(Total innovation expenditures)			0.007*	
			0.004	
Majority private, domestic	0.291***	0.293***	0.285***	0.285***
	0.034	0.035	0.035	0.035
Majority private, foreign	0.598***	0.600***	0.593***	0.593***
	0.058	0.059	0.059	0.059
Observations	108585	108585	104920	105057
Number of clusters	178	178	170	170
R-squared	0.86	0.86	0.86	0.86

Table 5 (Further specifications)

References:

1. Arrow, K., 1962. The economic implications of learning by doing. Reviewof Economic Studies 29, 155–172.
2. Beaudry, Catherine & Schiffauerova, Andrea, 2009. "Who's right, Marshall or Jacobs? The localization versus urbanization debate," Research Policy, Elsevier, vol. 38(2), pages 318-337.
3. Bertaud, Alain & Renaud, Bertrand, 1997. "Socialist Cities without Land Markets," Journal of Urban Economics, Elsevier, vol. 41(1), pages 137-151.
4. Conley, Timothy G. and Bill Dupor, 2003. "A Spatial Analysis of Sectoral Complementarity," Journal of Political Economy, University of Chicago Press, vol. 111(2), pages 311-352.
5. Duranton, Gilles and Diego Puga, 2001."Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products," American Economic Review, American Economic Association, vol. 91(5), pages 1454-1477.
6. Duranton, Gilles and Henry G. Overman (2008), "Exploring the Detailed Location Patterns of UK Manufacturing Industries Using Microgeographic Data," Journal of Regional Science, vol. 48(1), pages 213-243
7. Eberts, Randall W. and McMillen, Daniel P., 1999. "Agglomeration economies and urban public infrastructure," in: P. C. Cheshire and E. S. Mills (ed.), Handbook of Regional and Urban Economics, edition 1, volume 3, chapter 38, pages 1455-1495.
8. Ellison Glenn and Edward L. Glaeser & William Kerr, 2007. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," NBER Working Papers 13068.
9. Ellison, Glenn and Glaeser, Edward L (1997), "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," Journal of Political Economy, vol. 105(5), pages 889-927.
10. Glaeser, Edward L., Hedi D. Kallal and Jose A. Scheinkman and Andrei Shleifer, 1992. "Growth in Cities," Journal of Political Economy, University of Chicago Press, vol. 100(6), pages 1126-52
11. Henderson, Vernon, "Efficiency of Resources Usage and City Size," Journal of Urban Economics, 1986, 19, pp. 47-70.
12. Henderson, Vernon, “Marshall's Scale Economies” Journal of Urban Economics, 2003, 53, pp. 1-28
13. Jacobs, Jane (1969) “The Economy of Cities”, New York, Vintage Books
14. Lafourcade, Miren & Mion, Giordano, 2007. "Concentration, agglomeration and the size of plants," Regional Science and Urban Economics, Elsevier, vol. 37(1), pages 46-68.
15. Levinsohn, James and Amil Petrin, "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 2003, vol. 70(2), pages 317-341.
16. Marshall, Alfred. (1890) “*Principles of Economics”*, New York: Macmillian
17. Nakamura, Riohei. (1985) “Agglomeration Economies in Urban Manufacturing Industries: a Case of Japanese Cities”, *Journal of Urban Economics*, 17: 108–124
18. Romer, P., 1986. Increasing returns and long-run growth. Journal of Political Economy 94, 1002–1037.
19. Rosenthal, Stuart S. and Strange, William C., 2004. "Evidence on the nature and sources of agglomeration economies," in: J. V. Henderson and J. F. Thisse (ed.), Handbook of Regional and Urban Economics, edition 1, volume 4, chapter 49, pages 2119-2171 Elsevier.
20. Rosenthal, Stuart S. and William C. Strange, 2003. "Geography, Industrial Organization, and Agglomeration," The Review of Economics and Statistics, MIT Press, vol. 85(2), pages 377-393.
21. Sveikauskas, Leo (1975) “The Productivity of Cities”, *Quarterly Journal of Economics*, 89: 393–414