Competitive advantage, industry clusters and the magnetic attraction of investment in new plant and equipment from outside the region

A paper submitted for the 57th ERSA Congress
29 August - 1 September 2017
Groningen, The Netherlands

Timothy F. Slaper, Ph.D.
Research Director
Indiana Business Research Center
Indiana University Kelley School of Business

Ping "Claire" Zheng, Ph.D.
Economic Research Analyst
Indiana Business Research Center
Indiana University Kelley School of Business

Abstract

While the literature on agglomeration externalities has emphasized the competitive and productivity benefits associated with the concentration and co-location of related industries – i.e., industry clusters – the research is sparse on whether regions with specialized industry clusters magnetically attract investment from firms outside the region. Agglomeration externalities create benefits for related industries to co-locate, but to what degree do these externalities attract similar or complementary industries?

In this paper, we address whether, and to what degree, agglomeration externalities magnetically attract new operations and employment into a region. Using greenfield foreign direct investment data at the U.S. county level, we conclude that firms are more likely to invest in new or expanded facilities in regions that have a high absolute concentration of employment in their specific industry. Whether this magnetic attraction occurs for complementary industries within an industry cluster, the data suggest that there is a difference between high-tech and not high-tech industries.

We also find that that several regional characteristics that are considered important by site selectors – those informing the FDI location decisions – are more salient than other regional characteristics and attributes, for example, the availability of labor. We also find that certain state-level characteristics are also positively associated with greenfield FDI flows, such as lower electricity costs and good state governance. These results are largely similar and robust across statistical methods – OLS, logit, negative binomial and pseudo-panel – as well as dependent variables.

Keywords:

Agglomeration externalities, FDI, industry clusters, regional development

The literature on agglomeration externalities has largely been silent on whether specialized or diversified production structures attract external economic investment. Much has been written on the degree to which agglomeration externalities create benefits to like kinds of industries co-locating. Yet, the degree to which these externalities attract similar or complementary industries has largely gone unstudied.

In this paper, we address whether and the degree to which agglomeration externalities provide something of a magnetic attraction. Using greenfield foreign direct investment for the U.S., we conclude that firms are more likely to invest in new or expanded facilities in regions that have a relatively high concentration of employment in their particular industry or among complementary industries.

The contribution to the literature is threefold. First, the paper establishes that at a granular U.S. geographic level, firms tend to be attracted to regions – counties – that have an absolute concentration of employment in their industry cluster. Second, we also find that high-tech industries have a different FDI attraction profile than non-high-tech industry clusters, an important consideration for economic development practitioners to consider as they develop their development strategies. Third, we find that several regional characteristics that are considered important by site selectors – those informing the FDI location decisions – are more salient than other regional characteristics and attributes.

By way of an introduction, the paper first presents an overview of the theory and literature associated with the arguments for industry specialization, diversification and agglomeration externalities. In Section 2, we describe the data and measures. Our empirical method and results are presented in Section 3. We conclude with a brief discussion and conclusion in Section 4.

Introduction

The importance of industrial clusters to boost regional economic development has widely gained scholars' attention. Several researchers – Michael Porter and Christian Ketels, among others – have developed the study of cluster-based economic development and touted the employment and competitive benefits of cluster-based development strategies (Porter, 2003; Porter, 2008; Ketels, 2013; Ketels and Memedovic, 2009). The benefits revolve around employment growth, productivity and competitiveness. Although much of the empirical work focuses on the benefits of clusters on industrial employment, innovation and productivity, less systematic empirical attention has paid to identifying strong regional clusters and the regional characteristics that attend cluster formation and growth.

Clusters form due to the benefits of agglomeration externalities in a region, or, said differently, industries that develop in clusters tend to have a competitive advantage (Porter, 1998). It has sometimes been said that clusters form "because there is something in the air." More practically, what is in the air may be that firms in the same or related industries in a cluster benefit from being in close proximity, often called "localization externalities." Long established firms grow and new firms, start-ups, can also take advantage of a well-developed regional labor force and supply chain. One might say that these firms grow based on the region's economic metabolism, that is, they expand using the resources, labor and know-how in the region, as well as technology from outside the region – combined with increasing demand for the cluster's goods and services from outside the region. In this way, regional clusters growth metabolically.

On the other hand, there may be significant benefits to close geographic proximity for already established firms, either young or mature, from outside the region to move into the region to take advantage of these agglomerative externalities. In other words, clusters can also grow "magnetically," that is, a region can attract firms to take advantage of that region's competitive advantage in resources, supply networks and human talent. An example of magnetic growth is greenfield foreign direct investment (FDI).

The empirical question is then: do strong, established, growing clusters tend to attract incoming firms in the form of "foreign" direct investment? (Foreign is in quotes because any investment from outside the region – international or domestic – can take advantage of a region's competitive advantages to colocation.)

Industry cluster strength can be viewed as the relative concentration of an industry cluster, without regard to the balance or concentration of industries within that cluster, or as the absolute concentration of employment in a cluster. Kemeny and Storper (2015) make the case that specialization (or concentration) is better measured based on absolute size and look upon industry strength, or specialization, as measured as measured by a location quotient with a wary eye.

Industry cluster strength aligns with the notion of related variety discussed by Frenken, Van Oort, and Verburg (2007), who categorized industries based on their technological and material requirements. In a similar way, Delgado and colleagues (2016) used industry input-output relationships to categorize industries into industry clusters. The agglomerative benefits of such related variety are often called "localization economies" (Frenken et al., 2007) and were first conceptualized by Marshall over a century ago (1890/1966), and since refined by Arrow (1962) and Romer (1986). The agglomeration byproduct of related variety is often referred to as MAR externalities. MAR externalities are within related industries, usually broadly defined, but in this case MAR externalities would be in evidence within a cluster. "MAR cluster" is hereafter the term used for related industry concentration and its attendant benefits of agglomeration, in contrast to "industry cluster" which is a definition term based on related industry classification as classified by Porter's cluster mapping project (CMP). An industry cluster, therefore, may or may not be concentrated in a region.

The forces of agglomeration as expressed in the formation of regional industry clusters is partially based on the diversification of industries. As a result, it may be helpful to contrast the Jacobian urbanization diversity benefits (Jacobs, 1969) of the unrelated variety of economic sectors, with how industry cluster diversification will be operationalized here. The findings of Delgado et al. (2014) point to cluster variety – multiple clusters in a region that are related – as having a positive effect on innovation as measured by patenting rates. Frenken and colleagues (2007) hypothesize that Jacobian, or unrelated variety, externalities are positively related to employment, but they also discuss an additional dimension to unrelated variety, namely resilience.

In our case, a diversity of industries has both an industry cluster dimension, that is, the portfolio of industry clusters in a region, as well as the portfolio of related industries within an industry cluster. Both are akin to the diversification of stocks in a portfolio, but one is regional diversification of industry clusters while the other is within industry cluster diversification of industries that, by classification, are aggregated into an industry cluster. In other words, cluster diversity used here is not a measure of how, and in what ways, unrelated clusters are different from each other, rather diversity is more synonymous with balance within a region or within an industry cluster. Following Frenken, and colleagues (2007), we use an entropy index to measure cluster industry diversity/balance in a region.

This empirical study investigates the role of magnetic cluster growth in U.S. regions, in the form of greenfield and expansionary investment flows, i.e., FDI. Expanding upon the work of Delgado et al. (2014), we explore the degree to which the agglomeration externalities motivate a firm's decision to move into a region. In the spirit of their work, we examine whether a high concentration of related industries, or strong clusters, tend to attract additional investment inflows and thus additional employment within that cluster. In addition, we are particularly interested in whether a more diversified, or balanced, set of industries within a strong, or highly concentrated cluster – an MAR cluster – tends to attract new greenfield investment or additional expansionary investment among firms already operating in the region.

In the spirit of Ellison and Glaeser (1997), we are investigating whether the location patterns of new investment in plant and equipment, and the concomitant employment, follow a random, throw-a-dart approach or reflect decisions that may be motivated by seeking the competitive benefits of industrial colocation and concentration. For this reason, we can identify investment moving into a region and assess whether a region has relative strength, or specialization, in the cluster associated with the investment. In addition, we can also assess whether the receiving regional industry cluster has a diverse set of industries or simply dominated by one or two industries within that industry cluster.

One can also hypothesize about the level of the associated technological sophistication for the new employment. The investment in non-high-technology clusters is well in evidence after the Great Recession. Initial analysis also shows that the clusters in the top ten list in terms of the number of incoming jobs tend to be more diversified – that is, having higher average Shannon index scores. An initial examination of scatter plots of FDI and industry cluster strength/specialization for the top ten employment receiving clusters show that more diversified clusters tend to garner a greater volume of FDI-related employment. This may signal the importance of a well-developed labor force as well as supply chains and material linkages among co-located firms.

We also attempt to explore additional dimensions that an investor may consider important to a location decision. In their 2016 report, A New Standard, the International Economic Development Council, a Washington, DC-based association for advancing regional economic development, reported data on the most important factors in business location decisions. In addition to site availability, the leading location's characteristics included: infrastructure, workforce characteristics, wages, labor market, demographics, higher education, labor regulations, taxes and incentives. While many of these characteristics are available at the county level, the latter three items are more closely aligned with state policy and practices. If these regional or county characteristics do influence the decision to invest in one region as opposed to another, then these are appropriate control variables. To control for state policy effects, we also included state-based proxy data that investors may consider as indicators of good state governance.

Data Sources and Measures

To examine region-cluster employment growth, we draw on studies of regional economic growth as a function of the level of economic activity and attributes of the region (Combes, 2000; Delgado et al., 2014). The econometric model regresses announced investment on plant and equipment (and employment), both binary and level, on a number of factors that characterize cluster strength, such as employment location quotients (LQ), a measure frequently used to quantify the concentration of a particular population in a region as compared to the relative concentration of the nation (or some other geographic peer). We also use an entropy index, such as the Herfindahl or Shannon, to assess the diversity or evenness of industry clusters in a region as well as diversity of industries within an industry cluster.

We use employment by industry data from QCEW-complete employment estimates and aggregated by industry cluster definitions from the U.S. Cluster Mapping Project (CMP). Thus, all the industry data are bundled into "industry clusters," of which there are 70. The proprietary dataset, fDiMarkets is the source for greenfield and expansion employment and the number of investment projects associated with investment announcements. We use a concordance developed by fDiMarkets to map their industry definitions to NAICS-based industries and, following that, assigning FDI to the cluster aggregations from the CMP. While the CMP used NAICS, production-based industry definitions to build their cluster aggregations, fDiMarkets industry definitions are more of a hybrid between production-based and consumer-based classifications. This resulted in the need to aggregate two or three CMP clusters in order to correspond to fDiMarkets industry aggregations. There were also a couple instances for which there was only one industry in an fDiMarkets-CMP harmonized cluster, for example, tobacco. In these cases, these industries/clusters were removed from the analysis.

There are several potential weakness associated with the FDI announcement data. One, the jobs realized once the plant and equipment are in operation may be different than the number of jobs reported in the press releases. Two, there is no way to verify how many new, incoming magnetic jobs, were created, because of the disclosure constraints associated with record-level QCEW establishment data. In other words, one cannot link an FDI announcement record in 2012 with subsequent establishment data. Three, there is no fixed time between an FDI press release and realized jobs. The latter can vary greatly depending on the industry, the scale of investment, market demand conditions for the firms, etc. That said, firms have been known to spend several years and millions of dollars in site selection and negotiating with local and state officials before making an announcement; thus, we consider the FDI announcements as an appropriate signal for a region's relative attractiveness in terms of agglomeration externalities.

The American Community Survey is the source for several demographic characteristics: the population share of less than high school, some college, bachelor's degree or above, share of prime working ages (25 - 44) as well as mean travel time to workplace. The number of STEM graduates from the county is sourced from IPEDS at the National Center for Education Statistics. The Bureau of Labor Statistics is the source for county unemployment rate and the average hourly wage in manufacturing. County-level 2016 cost of living index – COLI and the transportation cost of living – is sourced from the Council For Community and Economic Research. Interstate lane miles by county – for 2015 – is from the U.S. Department of Transportation. Cost of electricity for industrial use – 2007-2014 – is from the U.S. Department of Energy. The Tax Foundation is the source for many “good governance” indicators: state business tax climate index (2015), state & local taxes per capita (2012), state & local tax as a percent of state income (2012) and the percentage of public pension plans that are funded (2014). Standard & Poors is the source for state credit ratings (2004-2016), another state good governance proxy. Venture capital data was obtained from Thompson Reuters. The source for university knowledge spillovers is Zheng and Slaper (2016).

The unit of analysis is U.S. county and industry cluster. While some FDI projects could be considered “local” – for example real estate development or consumer banking branches – in contrast to traded industries for which the market generally extends beyond the region, the vast majority of FDI announcements are for traded industries and we consider only traded industry clusters. We collected nine years of FDI data from 2007 to 2015 and grouped them into three time intervals of equal length. FDI data by county tends to be sparse. This is especially true of fDiMarkets data in the early to mid-part of the first decade of the 2000s when the company was refining its method and data collection technologies. Thus providing the rationale for using a pseudo-panel model and data for three periods. Of the 30,774 FDI announced projects, or events, in the U.S. over this time period, 20,632 were in the relevant traded industries.

There are 243,698 county by industry cluster observations for the 2007 to 2015 time period, implying that the average county has about 27 traded industry clusters. Around 40 percent of those county-industry clusters are sufficiently concentrated to be considered an MAR cluster in the “cluster development strategy” sense. That is, an above national average concentration of related industries that tend to benefit from economies of agglomeration. In other words, there are about 40 percent of county-by-industry-cluster observations that are, on a relative basis, “true” MAR clusters with LQs above 1.2. Of these MAR clusters, around 14 percent are high-technology industries.

County-by-industry-clusters can have multiple projects or FDI attraction events over the time period. As a result, the number of county-by-industry-clusters that recorded FDI employment announcements is whittled down to 8,194. The average number of jobs per FDI announcement is 190, but each FDI project or event can range from one new job to over 8,000. For an example of the latter, the IT sector in Travis County, Texas, as in Austin, attracted 8,000 new workers based on an FDI announcement.

While the main data source for the explanatory variables is QCEW data by industry, how these data are operationalized to provide measures of regional agglomeration and industry structure warrant discussion. It is to how cluster concentration/specialization and balance/evenness/diversity are measured we turn our attention.

There is a wide variation in terms for cluster concentration/specialization. For example, farming regions and regions endowed with natural resources tend to have very high employment concentration in specialized sectors. On the other hand, high-tech clusters, especially the ones associated with FDI – for example, Travis County, Texas – tend to have a more diversified economy and industrial profile.

We use a common entropy measure – The Shannon Evenness Index – to assess the degree to which a region's industry clusters are even/diversified or uneven. Expressed differently, how balanced or diversified is the portfolio of industry clusters in the region. The variable, *sei_clst*, is expressed as a zero, perfect unevenness, if there is only one industry in the region whereas 1 denotes a perfect balance among industry clusters. For region *g*, the SEI for an industry cluster is:

$$sei\_clst_g = \frac{\sum_{cl=1}^{n} p_g^{cl} * \ln\left(p_g^{cl}\right)}{\ln\left(n_g\right)}$$

Where *cl* is the industry cluster and *n* is the number of industry clusters present in the region.

Industry cluster strength, or specialization, is operationalized using two variables derived from the location quotient for region *g* , namely, *lq_clst*, and the binary threshold, and, lq_bin, to signal that the region has a substantial specialization in industry cluster *cl*.

$$lq\_clst = \frac{P_g^{cl}}{p_{US}^{cl}}; \ \ lq\_bin = lq\_clst \geq 1.2$$

The three measures discussed above are related to how even/diverse industry clusters are among themselves. These are measures "outside" a particular industry cluster. The following two measures address the balance, evenness or industry specialization within a cluster. As discussed above, the concern is the degree to which an industry cluster in region *g* has the same relative concentration of industries as the nation. Does an industry cluster that is dominated by one particular industry in that cluster yield the same magnetic attraction as an industry cluster that is more balanced? Does a wider complement of industries influence investment decisions?

We first measured within cluster evenness in the same fashion as outside cluster evenness, which is to say, applying the Shannon Evenness Index to industries within a cluster. The measure *sei_ind* is operationalized in a similar fashion for industries within cluster *cl* in region *g*:

$$sei\_ind_g^{cl} = \frac{\sum_{i=1}^{n^{g,cl}} p_i^{g,cl} * \ln\left(p_i^{g,cl}\right)}{\ln(n^{g,cl})}$$

While the SEI can provide a general sense of the balance of industries as a whole, even as a whole compared to the national average SEI, it does not provide an indication as to the nature of the imbalance: one hyper-specialized industry or two strongly specialized industries, etc. As result, we dropped this measure of evenness from the analysis and created a new measure in its place.

We developed a specialization, or imbalance, measure that provides a glimpse into the how the industry concentrations deviate from the national averages. The measure uses location quotients and sums up the absolute difference between the region's industries and the national LQ for those industries, which is, by definition, 1. The measure for within-cluster imbalance, or specialization, *clst_bal1*, is:

$$clst\_bal1 = \frac{\sum_{i=1}^{n^{g,cl}} \left| LQ_i^{g,cl} - 1 \right|}{n^{g,cl}}$$

Finally, to conclude the discussion about data and measures, we reference the source for non-obvious measures that may have required special treatment or explanation. What, for example, constitutes a high-tech industry, high-tech employment or high-tech occupations? These definitions are based upon research conducted by the Bureau of Labor Statistics and which serves as the foundation that the Indiana Business Research Center used for the Innovation Index 2.0 as documented in *Driving Regional Innovation* (2016).

The summary statistics for the dependent and explanatory variables are presented in Table 1.

- -

TABLE 1 ABOUT HERE

- -

Empirical Method and Results

Our empirical strategy is straightforward. We first adopted a simple regression model using pooled data from all years that regresses announced FDI-related employment on industry cluster employment and a number of factors that we considered, as well as the interactions of those factors. In addition to FDI employment, we used the same set of explanatory variables to examine how they affect the likelihood of receiving FDI as well as the number of FDI projects. A logit model was used to assess the likelihood of receiving investment and a negative binomial model was used to assess the frequency of investment counts. There are many more zeros, or null events, in the FDI data and as a result the data are still highly skewed. This motivated the authors to use the negative binomial model to complement the other models.

Finally, we pursued an pseudo-panel model to identify the source of variation – that is, is it the change of cluster characteristics over time or the regional differential in clusters – that contributes to the variation in FDI level. We constructed the panel in two ways. One way is in the dimension of space × time ((*i*,*t*), where *i* is a county-industry-cluster and *t* is a time period) and the other is in space × cluster ((*i*,*j*), where *i* is a county and *j* is an industry cluster).

For such a panel study, two types models are commonly examined: fixed effects (FE) and random effects (RE) models. We also estimated the FE model via both within-estimator (WE) and between-estimator (BE). This allows one to evaluate the association between explanatory and dependent variables along each dimensional space.

As noted above, the authors explored several modeling strategies, including OLS, pseudo-panel, logit and negative binomial procedures. The results, explained variation, the coefficients and significance of the explanatory variables, are all similar, but vary because the different modeling strategies use different dependent variables. First, we discuss the results from the OLS model using industry cluster employment as the dependent variable. Following that, we briefly present the results from the logit and negative binomial models. The results of the preferred models for OLS, logit and negative binomial are presented in Table 2 in columns 1, 2 and 3 respectively. Table 3 reports the pseudo-panel models.

\- -

TABLE 2 ABOUT HERE

\- -

The main focus for the OLS model – Model 1 – is attracting FDI investment and its concomitant employment. The OLS model used the log form of the FDI employment level for an industry cluster for all U.S. counties. Our first hypothesis is confirmed. Investment in a region (county) as measured by additional employment attributed to greenfield or expansion FDI is strongly positively associated with total cluster employment (*ln(clstemp))*. The larger the absolute size of an industrial cluster, the greater the magnetic attraction of FDI related employment. The coefficient estimate 0.256 suggests that a 1 percent increase in industrial cluster employment is associated with nearly 30 percent increase in FDI employment, i.e., (exp(0.256)-1). That said, the marginal effect is higher for non-high-tech clusters, in comparison to high-tech clusters, as evidenced by the negative coefficient estimate of the interaction term of high-tech and industry cluster employment *(htflag x ln(clstemp)).*

Also in evidence is that MAR clusters," that is, the presence of relatively high concentration of an industry cluster – recall that "industry cluster" is simply a definitional term similar to "industry sector" and does not indicate the presence of a high concentration of cluster employment – is also positively associated with the binary variable industry clusters with an LQ of greater than 1.2 (*lq_bin*). Low concentration industry clusters, that is, a relatively weak presence of an industry cluster aggregate employment in a region, can have a negative effect on attracting FDI employment, as implied by the negative coefficient for the log of an industry cluster LQ (*lnlq_clst*). Considering that a vast majority of counties have an LQ of below one (1) for most a county' industry clusters, this result is not surprising. Moreover, the interaction term for an MAR cluster (*lq_bin*) and industry cluster LQ (*lnlq_clst*) is also positive, corroborating the evidence that specialization in an industry cluster serves as a magnet for FDI related employment.

Industry cluster diversity, or evenness, in a county is negatively associated with an increase in FDI related employment. That is, those regions that specialize in one or two industry clusters are less diversified and tend to receive more FDI related employment. This effect, however, is mainly driven by high-tech clusters, based on high-tech and SEI interaction (*htflag x sei_clst*). The positive coefficient for the binary variable for high-tech clusters would indicate that, in general, high-tech industry clusters tend to gain more FDI related employment than non-high-tech clusters (*htflag*).

Within industry cluster diversity, as measured by *ln(clst_bal1)*, a measure of imbalance within the industries that make up an industry cluster, relative to the U.S., would indicate that specialization within a cluster is positive, as the coefficient for imbalance in positive and statistically significant. However, this effect is offset by the negative effect from the interaction term for high-tech and imbalance (*htflag x ln(clst_bal1)*, suggesting that industries in high-tech industry clusters appear to benefit from a more balanced industrial profile within a cluster for attracting FDI employment. Conversely, industry clusters that are not high-tech would not be penalized for the lack of within cluster diversity or evenness.

Leaving the magnetic benefits of agglomeration aside, there are several regional characteristics that may also influence attracting FDI employment, irrespective of whether site selectors and prospective investors consider them important. The cost of living (for transportation), for example, changes sign and loses statistical significance once state-level characteristics are considered. The results for level of educational attainment is also ambiguous. The presence of universities graduating STEM degree holders is somewhat positive, indicating that the presence of a robust educational system may positively influence FDI decisions. The presence of prime working age adults is strongly positive and statistically significant. That, combined with higher unemployment rates, would indicate that FDI decision makers are interested in locations with abundant labor. (Higher average wage for manufacturing employment, a signal for a tight labor market, was not significant.) The presence of any venture capital flowing into the region was negative, yet high levels of venture capital are positive. Interpreting the mean travel time to work is difficult. Increases in mean travel time may indicate congestion in cities, a negative. On the other hand, long commutes from one rural county to another or from one exurb to another may indicate a degree of labor flexibility and a larger labor shed from which to draw talent.

We also included state cost and policy characteristics, such as electricity cost and tax burden measures, into the models, in the consideration that these would influence the decision making of whether to invest in a region. It appears that only electricity cost has a strong association with FDI employment – 1 percent increase in electricity cost is associated with 26 percent decrease in the employment. Of the several measures for good state governance, business conditions and taxes, only state and local taxes appear to have an influence.

The above variables taken together weakly explain the variation of greenfield and expansion FDI-related employment. The overall fit of the OLS model suggests that 16 percent of variation is explained. This result is not entirely different from the pseudo R-square results from the logit and negative binomial models of 0.220 and 0.173 respectively. We now turn the attention to the explanatory variables in those models.

Using a logit model, we estimated the odds of attracting FDI projects – all projects, whether large or small are counted the same – based on the same variables in the OLS model. The signs and statistical significance for the absolute size of cluster employment, within cluster specialization and high-tech are similar to that of the OLS. Where the results diverge are the relative measures of cluster strength, or specialization. Neither the variable indicating the presence of an MAR cluster (*ln_bin*) nor the relative concentration of a cluster (*lnlq_clst*) are statistically significant. The odds of attracting FDI projects also does not seem to depend on the interplay between the size of the industry cluster and whether the cluster in the region is high-tech or not. For the logit model, educational attainment emerges as a factor in increasing the chances of attracting FDI projects. A higher proportion of the population without a high school degree reduces the chances while the greater proportion of a region's population with a bachelor's degree or higher increases the odds of attracting FDI projects.

The scale of a region as measured by total county employment emerges as increasing the odds of attracting FDI projects, as does relative proximity to university research and development expenditures, as measured by university knowledge spillovers within 50 miles of universities. Mean travel time reduces a region's chances, while one measure for infrastructure availability, Interstate lane miles per capita, increases a region's attractiveness. Finally, in terms of good governance measures, a state's credit rating positively influences a region's chances for attracting FDI projects.

In addition to FDI employment and the likelihood of receiving FDI projects, the number of FDI projects using a negative binomial model were considered, Model 3 in Table 2. Overall, these results were more congruous to the OLS results in both coefficient, sign and significance. The high-tech and cluster employment interaction term was not statistically significant, unlike the OLS employment model, but then, the number of projects and the number of new workers associated with those projects can deviate considerably. In contrast to the OLS model, the high-tech and county industry cluster evenness/diversity interaction term lost its significance. The cost of living is positively related to the number of project, contrary to expectations. The only educational attainment measure with statistical significance is for some college, and then it has a negative sign, contrary to expectations. Also contrary to expectations is the sign for the proximity to university R&D, while the number of STEM degrees graduates from intuitions in the region is positively related to the number of projects a region attracted. Interstate lane miles per capita is, oddly, negative while the mean travel time is positively related to the number of FDI projects attracted. As for state characteristics, higher electricity costs are associated with fewer FDI projects, but, contrary to expectations, both state and local taxes per capita and the ratio of state pension plans that are funded are marginally negative (and significant). That said, a good business climate is positively associated with attracting FDI projects.

Attention is turned to the results of the pseudo-panel model, reported in Table 3.

- -

TABLE 3 ABOUT HERE

- -

The results of WE in panel I shows that the across-time variation of most of the explanatory variables are not significantly associated with FDI employment, with a few exceptions. The sectors that have higher rates of unemployment (*unempr*) would attract more FDI employment, but so would regions that have grown more highly educated workers (*somecllg* and *baab*). On the other hand, large high-tech clusters (*htflag x lnclsemp*) and regions with more prime working age populations (*prime*) may reduce the inflow of FDI employment over time. That said, the overall fit of the WE model is rather poor, suggesting that time variation is an insignificant source in explaining FDI employment attraction. This leads to the conclusion that the explanatory power of the characteristics associated with attracting FDI employment is not associated with the time dimension, but rather is explained by the cross-sectional variation across the industry cluster space. The results of BE in panel I confirms this conclusion. Although the over-identifying restriction test shows that the RE model might be a better choice over FE models, the coefficient estimates between BE and RE are very close to each other, and their directions – the sign and significance of the coefficients – are also consistent with the OLS.

Panel II eliminates the time dimension in order to assess the differential in geographic and industry cluster space. It is not surprising that the marginal effects from industry cluster-specific characteristics are much stronger (highly significant and larger) than those from region-specific characteristics in WE model that eliminates regional characteristics. The results from BE model are similar to the BE and RE models in panel I. One then concludes that industry cluster effects are more dominate than regional characteristics for FDI location decisions.

In summary, the results of the models are largely consistent both in terms of explanatory variables and explained variation.

Discussion and Conclusion

Whether site selectors or the corporate decision makers are aware of it or not – and based on surveys of site selectors the benefits of the economies of agglomeration are not "important considerations" – the location of FDI projects decision align with the benefits of MAR clusters.

Our findings comport with the regional economic development strategies to focus on industry cluster-based programs to emphasize the competitive benefits of industry specialization, and the associated aspects of established supply chains and a workforce aligned with the needs of related industries. Indeed, FDI decision makers appear to be aware that "there is something in the air" and place their investment dollars accordingly.

While the variation in FDI location is not greatly explained by industry cluster specialization, we have found that MAR clusters do magnetically attract incoming investment into counties and that the type of industry cluster – high-tech versus not high-tech – may dictate whether a region needs to have a balanced, diversified cluster or if specialization in one particular industry within a cluster can be sufficient to attract FDI.

Our finding also indicate that certain regional or state characteristics may be important in attracting FDI. Electricity costs appear to weigh heavily on the scale of the FDI decision. Educational attainment registered as somewhat ambiguous, but FDI decisions in this regard may relate to the scale and nature of the facility activity—small high-tech may not consider this an important consideration but for a 2,000 person manufacturing plant, it is critical. A “flexible” labor market may be an important consideration when considering the scale of a facility: higher unemployment, a larger share of prime working age population and longer travel time to work may indicate sufficient slack in the labor shed to induce larger facilities to locate in a region. A state’s higher credit score may indicate the ability of a state government to negotiate favorable terms to the firm to negotiate tax breaks or worker training and retention incentives. Without more granular, case specific information, these statements are nothing more than hypotheses, but based on our findings, credible research paths to explore.

Absent consistent data for site availability or deal specific details on the incentives – tax reduction benefits or worker hiring and training inducements – to locate in a particular region/county, the data on FDI location decisions would indicate that agglomeration economies, lower electricity costs and good state governance conditions drive where greenfield investment and expansions occur.

In summary, we have tested and found valid the claims that the economies of agglomeration serve as one inducement, among several considerations, for firms to locate in one region as opposed to another. Moreover, these economies of agglomeration are associated with the absolute scale of the agglomerated firms, rather than the relative concentration of those firms within a region. We have also found that a firm within an industrial class may be attracted to regions with other firms of just the same industrial class (not high-tech, i.e.), in contrast to a type of firm seeking a region with several complementary, related firms within a cluster. Finally, we have found several important regional and state characteristics that appear to have motivated FDI decisions. Economic development practitioners, take note.

References

Andersson, F., Burgess, S., & Lane, J. I. (2007). Cities, matching and the productivity gains of agglomeration. *Journal of Urban Economics, 61*(1), 112-128.

Arrow, J. (1962). The economic implications of learning by doing. *Review of Economic Studies, 29*(3), 155–173.

Carlino, G. A., Chatterjee, S., & Hunt, R. M. (2007). Urban density and the rate of invention. *Journal of Urban Economics, 61*(3), 389-419.

Combes, P. P. (2000). Economic structure and local growth: France, 1984–1993. *Journal of Urban Economics*, *47*(3), 329-355.

De Groot, H. L., Poot, J., & Smit, M. (2010). Agglomeration, innovation and regional development: Theoretical perspectives and meta-analysis. In R. Capello & P. Nijkamp (Eds.), *Handbook of regional growth and development theories* (pp. 256-281). Northampton, MA: Edward Elgar Publishing.

Delgado, M., Porter, M. E., & Stern, S. (2014). Clusters, convergence, and economic performance. *Research Policy*, *43*(10), 1785-1799.

Delgado, M., Porter, M. E., & Stern, S. (2016). Defining clusters of related industries. *Journal of Economic Geography*, *16*(1), 1-38.

Delgado, M., & Zeuli, K. (2016). Clusters and regional performance: Implications for inner cities. *Economic Development Quarterly*, *30*(2), 117-136.

Duranton, G., & Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, *72*(4), 1077-1106.

Ellison, G., & Glaeser, E. L. (1997). Geographic concentration in US manufacturing industries: a dartboard approach. *Journal of Political Economy*, *105*(5), 889-927.

Fang, L. (2015). Do clusters encourage innovation? A meta-analysis. *Journal of Planning Literature, 30*(3), 239-260.

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional studies, 41*(5), 685-697.

Henderson, V. (1997). Externalities and industrial development. *Journal of Urban Economics, 42*(3), 449-470.

Indiana Business Research Center. (2016). *Driving regional innovation: The Innovation Index 2.0.* Bloomington, IN: Author. Retrieved from http://www.statsamerica.org/ii2/reports/Driving-Regional-Innovation.pdf

International Economic Development Council. (2016). *A new standard: Achieving data excellence in economic development.* Washington, DC: Author. Retrieved from

Jacobs, J. (1969). *The economy of cities*. New York, NY: Vintage.

Kemeny, T., & Storper, M. (2015). Is specialization good for regional economic development? *Regional Studies, 49*(6), 1003-1018.

Ketels, C. H., & Memedovic, O. (2008). From clusters to cluster-based economic development. *International Journal of Technological Learning, Innovation and Development*, *1*(3), 375-392.

Ketels, C. (2013). Cluster policy: A guide to the state of the debate. In P. Meusburger, J. Glückler, & M. el Meskioui (Eds.), *Knowledge and the economy* (pp. 249-269). Berlin: Springer.

Marshall, A. (1966). *Principles of economics* (8th ed.). London: Macmillan.

Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography, 87*(3), 237-265.

Porter, M. E. (1998). Clusters and competition: New agendas for companies, governments, and institutions. In M. E. Porter (Ed.), *On competition* (pp. 197–299). Boston, MA: Harvard Business School Press.

Porter, M. E. (2003). The economic performance of regions. *Regional Studies 37*, 549–578.

Porter, M. E. (2008). *On competition* (updated and expanded ed.). Boston, MA: Harvard Business School Press.

Porter, M. E., & Ketels, C. (2009). Clusters and industrial districts: Common roots, different perspectives. In G. Becattini, M. Bellandi, & L. De Propris (Eds.), *Handbook of industrial districts* (pp. 172–186). Northampton, MA: Edward Elgar Publishing.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy, 94*, 1002–1037.

Zheng, P., & Slaper, T. F. (2016). University knowledge spillovers, geographic proximity and innovation: An analysis of patent filings across US counties. [Presentation]. *Paper presented at the 63th Annual North American Meetings of the Regional Science Association International. Minneapolis*, MN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2857130

Table 1: Summary statistics

Description	Variable	N	Mean	S.D.	Median	Min	Max
Dependent variables:							
Log of total FDI employment	lnjobs_tot	8,172	4.366	1.298	4.317	0	9.018
Receiving FDI (0 or 1)	fdi_bin	243,698	0.034	0.180	0	0	1
Total no. of FDI projects (discrete)	projn_tot	8,172	2.509	6.296	1	1	234
Cluster characteristics:							
Log of total cluster employment	lnclstemp	243,698	4.425	1.764	4.202	1.609	12.362
County specific cluster diversity (Shannon Evenness Index)	sei_clst	243,698	0.797	0.047	0.798	0.238	1.221
Within industry cluster imbalance (log of county industry cluster score)	lnclst_bal1	243,698	1.126	1.156	1.020	-5.657	7.435
LQ>=1.2 (0 or 1)	lq_bin	243,698	0.378	0.485	0	0	1
Log of industry cluster LQ	lnlq_clst	243,698	-0.157	1.473	-0.239	-7.059	7.735
High-tech cluster indentifier (0 or 1)	htflag	243,698	0.192	0.394	0	0	1
Human capital:							
Population share of Less than high school (in percentage)	nohs	243,698	2.214	0.957	2.049	0	14.612
Population share of some college education (in percentage)	somecllg	243,698	29.137	4.978	29.146	10.169	49.020
Population share of BA or above degrees (in percentage)	baab	243,698	21.164	9.365	18.734	3.876	72.186
Log of population share of STEM graduates (log of percentage)	lnstem	89,608	-2.537	1.420	-2.463	-9.107	2.032
Others county characteristics:							
Transportation COLI	coli_trsp	243,569	100.737	7.274	99.400	79.200	145.900

Table 1: Summary statistics, continued

Population share of prime working ages (25 - 44) (in percentage)	prime	243,688	24.242	3.086	24.097	0	44.320
Log of county total employment	lnctyemp	243,698	9.847	1.565	9.691	3.504	15.238
Employment share in high-tech industries (in percentage)	htempr	243,667	3.564	3.060	2.771	0	76.449
Unemployment rate (in percentage)	unempr	243,667	7.249	2.517	6.914	1.466	28.350
Average hourly wages in manufacturing	ap	243,698	22.463	6.837	21.217	6.158	81.750
Receiving VC (0 or 1)	vc_bin	243,698	0.176	0.380	0	0	1
Log of VC investment	lnvc	243,698	2.947	6.464	0	0	23.451
Log of university knowledge spillovers at 50-mile cutoff	lnkspl_50	128,958	3.154	0.770	3.169	0.757	5.456
Mean travel time	meantravel	243,698	22.153	4.919	21.827	2.603	41.473
Log of lane mile per capita	lnlnmil	243,698	7.908	0.641	7.873	4.570	10.939
State characteristics:							
Log of electricity cost (cents/kilowatthour)	lnelectr	243,698	1.883	0.221	1.838	1.411	3.322
State business tax climate index for 2015	buzclmt15	243,698	5.103	0.769	4.990	3.490	7.790
Right to work identifier (0 or 1)	r2w_flag	243,698	0.474	0.499	0	0	1
State & local tax per capita	tb1	243,698	1321.343	503.666	1335.308	522.013	3138.619
State & loccal tax as state income (in percentage)	tb2	243,698	94.076	12.312	93	65	127
Funded ratio of public pension plans (in percentage)	ppf	243,698	77.481	15.875	79	41	107
State credit rating scores	credit	243,698	6.639	1.315	7	2	8

Table 2: Summary of empirical results of pooled sample

	FDI employment	FDI received or not	No. of projects
	(1)	(2)	(3)
lnclstemp	0.256***	0.434***	0.319***
	(0.02)	(0.01)	(0.02)
sei_clst	0.609	1.341	0.378
	(0.78)	(0.91)	(0.61)
lnclst_bal1	0.211***	0.224***	0.080***
	(0.03)	(0.03)	(0.02)
htflag	2.202**	2.141***	1.496*
	(0.89)	(0.66)	(0.84)
lq_bin	0.151**	0.025	0.165**
	(0.07)	(0.06)	(0.07)
lnlq_clst	-0.123***	-0.017	-0.096***
	(0.04)	(0.03)	(0.04)
lq_bin × lnlq_clst	0.265***		
	(0.06)		
htflag × lnclstemp	-0.061***	-0.007	-0.004
	(0.02)	(0.02)	(0.01)
htflag × sei_clst	-2.107*	-1.444*	-1.765
	(1.15)	(0.83)	(1.08)
Htflag × lnclst_bal1	-0.243***	-0.380***	-0.169***
	(0.05)	(0.04)	(0.04)
Coli_trsp	0.001	-0.001	0.013***
	(0.01)	(0.01)	(0.00)
nohs	0.012	-0.125**	-0.046
	(0.05)	(0.06)	(0.04)
somecllg	-0.000	-0.000	-0.031***
	(0.01)	(0.01)	(0.01)
baab	-0.004	0.018***	-0.005
	(0.00)	(0.01)	(0.00)
lnstem	0.034*	-0.023	0.042***
	(0.02)	(0.02)	(0.01)
lnctyemp	0.049	0.272***	0.128***
	(0.05)	(0.07)	(0.04)
prime	0.033***	0.020	0.032***
	(0.01)	(0.02)	(0.01)
htempr	0.003	0.008	-0.008
	(0.01)	(0.01)	(0.01)
unempr	0.026**	0.189***	0.023**
	(0.01)	(0.02)	(0.01)

vc_bin	-0.667***	-0.642**	-0.744***
	(0.21)	(0.30)	(0.20)
lnvc	0.041***	0.046**	0.043***
	(0.01)	(0.02)	(0.01)
lnkspl_50	-0.066	0.169***	-0.168***
	(0.05)	(0.06)	(0.04)
meantravel	0.019**	-0.049***	0.013**
	(0.01)	(0.01)	(0.01)
lnlnmil	0.001	0.287**	-0.108**
	(0.07)	(0.12)	(0.05)
ap	-0.001	0.004	0.004
	(0.00)	(0.00)	(0.00)
lnelectr	-0.256**	-0.140	-0.217**
	(0.13)	(0.18)	(0.11)
Buzclmt15	0.008	0.038	0.058*
	(0.03)	(0.04)	(0.03)
r2w_flag	0.058	0.055	0.058
	(0.06)	(0.09)	(0.04)
tb1	-0.000	-0.000*	-0.000**
	(0.00)	(0.00)	(0.00)
tb2	-0.005*	-0.000	-0.000
	(0.00)	(0.00)	(0.00)
ppf	-0.002	-0.004	-0.002*
	(0.00)	(0.00)	(0.00)
credit	0.030	0.151***	0.037*
	(0.02)	(0.03)	(0.02)
Constant	0.943	-14.242***	-3.238***
	(0.93)	(1.33)	(0.72)
$\ln(\alpha)$			-1.234***
			(0.09)
Observations	5,032	55,635	5,033
Adj. R^2	0.159		
Pseudo R^2		0.220	0.173

Notes: FDI employment was estimated by OLS, the binary FDI logit model and no. of projects negative Binomial model. The significance of α coefficient suggests that the data is highly skewed and in support of using the negative Binomial model. Note that the sample size is greatly reduced, because for employment and project counts, we only restrict our sample to positive values. Significance: * 10%, ** 5%, *** 1%.

Table 3: Summary of empirical results of panel models for FDI employment

	Panel I			Panel II	
	WE	BE	RE	WE	BE
lnclstemp	0.287	0.149***	0.180***	0.201***	0.191***
	(0.31)	(0.02)	(0.02)	(0.04)	(0.04)
sei_clst	-5.186	-0.309	-0.407	*o.m.*	-1.521
	(3.62)	(0.54)	(0.51)		(1.01)
lnclst_bal1	-0.009	0.122***	0.133***	0.290***	0.118***
	(0.11)	(0.02)	(0.02)	(0.05)	(0.04)
htflag	*o.m.*	0.667	1.142*	2.990**	-0.275
		(0.68)	(0.62)	(1.44)	(1.83)
lq_bin	-0.054	0.112**	0.119**	0.077	-0.033
	(0.14)	(0.05)	(0.05)	(0.11)	(0.14)
lnlq_clst	-0.136	-0.048*	-0.060**	0.014	-0.041
	(0.33)	(0.03)	(0.03)	(0.06)	(0.08)
lq_bin lnlq_clst	-0.037	0.136***	0.154***	0.240**	0.174*
	(0.24)	(0.04)	(0.04)	(0.09)	(0.09)
htflag x lnclstemp	-0.333**	-0.037**	-0.047***	-0.098***	-0.170***
	(0.16)	(0.02)	(0.02)	(0.03)	(0.06)
htflag x sei_clst	-6.376	-0.436	-0.931	-2.749	1.680
	(5.31)	(0.85)	(0.78)	(1.83)	(2.20)
htflag x lnclst_bal1	0.031	-0.145***	-0.159***	-0.383***	-0.262**
	(0.19)	(0.04)	(0.04)	(0.09)	(0.11)
coli_trsp	*o.m.*	-0.005*	-0.002	*o.m.*	-0.017**
		(0.00)	(0.01)		(0.01)
nohs	-0.187	-0.021	-0.038	*o.m.*	-0.066
	(0.12)	(0.03)	(0.03)		(0.06)
somecllg	0.082**	-0.002	-0.004	*o.m.*	0.008
	(0.04)	(0.00)	(0.01)		(0.01)
baab	0.061*	-0.002	-0.001	*o.m.*	-0.007
	(0.04)	(0.00)	(0.00)		(0.01)
lnstem	0.079	-0.007	-0.007	*o.m.*	0.002
	(0.06)	(0.01)	(0.01)		(0.01)
prime	-0.114***	0.030***	0.031***	*o.m.*	0.027*
	(0.04)	(0.01)	(0.01)		(0.01)
htempr	-0.039	-0.001	-0.005	*o.m.*	0.013
	(0.06)	(0.01)	(0.01)		(0.01)
unempr	0.110***	0.024***	0.042***	*o.m.*	0.074***
	(0.02)	(0.01)	(0.01)		(0.03)
vc_bin	-0.106	-0.807***	-0.580***	*o.m.*	-1.371***
	(0.41)	(0.20)	(0.17)		(0.50)

lnvc	0.008	0.050***	0.033***	*o.m.*	0.087***
	(0.03)	(0.01)	(0.01)		(0.03)
lnkspl_50					0.134**
					(0.07)
meantravel	-0.016	0.012***	0.011**	*o.m.*	-0.003
	(0.06)	(0.00)	(0.00)		(0.01)
lnlnmil		0.078**	0.092**	*o.m.*	0.076
		(0.03)	(0.05)		(0.08)
ap	0.023	-0.001	0.004	*o.m.*	-0.001
	(0.02)	(0.00)	(0.00)		(0.01)
lnelectr	-0.368	-0.185**	-0.220**	*o.m.*	-0.074
	(0.31)	(0.09)	(0.11)		(0.19)
buzclmt15		0.003	-0.002	*o.m.*	0.019
		(0.03)	(0.03)		(0.05)
r2w_flag	*o.m.*	0.085**	0.099**	*o.m.*	0.063
		(0.04)	(0.05)		(0.08)
tb1	*o.m.*	-0.000	-0.000	*o.m.*	-0.000
		(0.00)	(0.00)		(0.00)
tb2	*o.m.*	-0.003	-0.005**	*o.m.*	-0.010**
		(0.00)	(0.00)		(0.00)
ppf	*o.m.*	0.001	0.001	*o.m.*	0.002
		(0.00)	(0.00)		(0.00)
credit	0.013	0.007	0.015	*o.m.*	-0.018
	(0.05)	(0.01)	(0.01)		(0.03)
Constant	8.515**	2.601***	2.102***	2.715***	4.797***
	(4.27)	(0.70)	(0.68)	(0.27)	(1.40)
Observations	8,171	8,171	8,171	2,434	2,434
Within R^2	0.061	0.000	0.005	0.097	0.040
Between R^2	0.000	0.099	0.097	0.086	0.195
Overall R^2	0.001	0.126	0.128	0.089	0.112

Notes: The test statistic of over-identifying restriction on the RE model in Panel I is $\chi^2(18) = 129.89$ that rejects the null that is in favor of the FE models. Panel II (the cluster panel) used only the sample of last time period (2012 – 2015) and estimated FE models, due to unbalanced panel. Statistical significance: *10%, ** 5% and *** 1%.