Research Policy 43 (2014) 1785–1799

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Clusters, convergence, and economic performance

Mercedes Delgado[a,*], Michael E. Porter[b], Scott Stern[c]

[a] Temple University, Fox School of Business, United States
[b] Harvard University, United States
[c] MIT Sloan, NBER, United States

ARTICLE INFO

Article history:
Received 27 October 2012
Received in revised form 21 May 2014
Accepted 23 May 2014
Available online 28 June 2014

Keywords:
Industry clusters
Dynamic economies of agglomeration
Inter-regional spillovers

ABSTRACT

This paper evaluates the role of regional cluster composition in regional industry performance. On the one hand, diminishing returns to specialization in a location can result in a convergence effect: the growth rate of an industry within a region may be declining in the level of economic activity of that industry. At the same time, positive spillovers across complementary economic activities can provide an impetus for agglomeration: the growth rate of an industry within a region may be increasing in the "strength" (i.e., relative presence) of related industries. Building on Porter (1998, 2003), we develop a systematic empirical framework to analyze the role of regional clusters – groups of closely related industries operating within a particular region – in the growth of regional industries. We exploit data from the US Cluster Mapping Project to examine the effects of agglomeration within regional clusters after controlling for convergence at the region-industry level. Our findings suggest that industries located in a strong cluster register higher employment and patenting growth. Regional industry growth also increases with the strength of related clusters in the region and with the strength of similar clusters in adjacent regions. We also find evidence of the complementarity between employment and innovation performance in regional clusters: both the initial employment and patenting strength of a cluster have a separate positive effect on the employment and patenting growth of the constituent industries. Finally, we find that new regional industries emerge where there is a strong cluster. These findings are consistent with multiple types of externalities arising in clusters, including knowledge, skills, and input–output linkages.

1. Introduction

Significant variation in regional economic performance is a striking feature of the US economy as well as that of other nations. Numerous theories have been proposed to explain why some regions achieve significantly higher growth rates than others in the highly open US economy, with particular emphasis on the role of initial conditions, the potential for innovation and knowledge spillovers, and the composition of economic activity (Porter, 1990; Barro and Sala-i-Martin, 1991; Glaeser et al., 1992; Venables, 1996; Henderson, 1997; Fujita et al., 1999). Policymakers and researchers have focused considerable attention on regions such as Silicon Valley, which appear to have achieved strong economic performance linked to the presence of innovative clusters of related companies and industries (Porter, 1990, 1998; Saxenian, 1994; Swann, 1998; Bresnahan and Gambardella, 2004). The empirical literature examining the impact of the presence of related economic activity on regional performance is small but growing, with most studies focusing on particular dimensions of performance (Baptista and Swann, 1998; Feldman and Audretsch, 1999; Combes, 2000a; Porter, 2003; Glaeser and Kerr, 2009; Delgado et al., 2010).

This paper focuses on the role of clusters – groups of closely related industries co-located in a region – in the employment and innovation growth of the individual industries that constitute each cluster. Empirical investigation of region-industry growth must account for two economic forces: convergence and agglomeration. Convergence arises when the potential for growth is *declining* in the level of economic activity as a result of diminishing returns (Barro and Sala-i-Martin, 1991). While many studies focus on diminishing returns at the regional level, convergence also arises at the region-industry level (Henderson et al., 1995; Dumais et al., 2002). In this case, the region-industry growth rate will be declining in the initial level of economic activity due to mean reversion or diseconomies of agglomeration (e.g., congestion costs that increase the price of inputs).

* Corresponding author. Tel.: +1 215 204 3052.
E-mail addresses: mdelgado@temple.edu (M. Delgado), mporter@hbs.edu (M.E. Porter), sstern@mit.edu (S. Stern).

http://dx.doi.org/10.1016/j.respol.2014.05.007
0048-7333/

Agglomeration exerts a countervailing force on regional performance. In the presence of agglomeration economies, growth is *increasing* in the level of economic activity (Glaeser et al., 1992). Agglomeration arises from interdependencies across complementary activities that give rise to increasing returns. The literature has often contrasted two potential types of agglomeration forces: localization (increasing returns to activities within a single industry) and urbanization (increasing returns to diversity at the overall regional level). The impact of these types of agglomeration is often obscured by the influence of convergence on growth. If both convergence and agglomeration effects are present at the region-industry level, the economic growth of a regional industry will reflect a balancing of the two effects (Henderson et al., 1995).

This paper moves beyond this impasse by focusing on the role of clusters of related industries in region-industry growth. While convergence may prevail at the region-industry level, we examine agglomeration forces that operate across industries within a regional cluster. The presence of complementary activity in a cluster may give rise to externalities that will facilitate the growth of the constituent region-industries.

Our focus on clusters provides three related contributions. First, we are able to move beyond the traditional dichotomy of agglomeration forces: localization of individual industries and urbanization arising from the overall diversity of regional economic activity (Glaeser et al., 1992; Henderson et al., 1995; Combes, 2000a). Instead, building on Porter (1998, 2003), we study agglomeration forces arising among related industries that constitute clusters. By sharing common technologies, knowledge, inputs, outputs, and cluster-specific institutions, industries within a cluster may benefit from numerous complementarities.[1]

We evaluate the role of clusters in the employment and innovation growth of regional industries. Our empirical methodology examines agglomeration within the set of related industries that surround a region-industry, while accounting for convergence at the region-industry level. We expect that a region-industry growth in employment (and innovation) will increase in the "strength," or relative presence of the regional cluster within which that industry operates.

Second, we examine the duality between employment and innovation performance in regional clusters. An active debate has centered on whether co-location of production and innovation is important for employment and innovation outcomes (Dertouzos et al., 1989; Helper et al., 2012; Porter and Rivkin, 2012; Berger, 2013). Motivated by this debate, we assess the role of the initial employment and innovation strength of a cluster in the growth of the industries in the cluster. If innovation concentration matters for employment creation, the innovation strength of the cluster will positively relate to the employment growth of the industries in the cluster.

This analysis also allows examining the types of cluster agglomeration that are associated with regional industry growth. For example, if the initial innovation strength in the regional cluster is the dominant influence on employment growth, this would suggest that employment growth is mainly driven by knowledge spillovers. However, if both the initial employment and innovation strength of the cluster matter for employment growth, this would suggest that a broader set of mechanisms is at work, including input–output linkages and access to demand as well as knowledge spillovers.

Finally, we use the cluster framework to examine the diversification of regional economies. Several studies suggest that the emergence of new industries in a region is affected by the pre-existing industry composition of the region (Swann, 1998; Porter, 1998; Klepper, 2007, 2010; Neffke et al., 2011). Externalities of various sorts may arise within a regional cluster and in related clusters that influence the emergence of industries in the cluster.

We investigate these questions utilizing a novel panel dataset developed by the US Cluster Mapping Project (CMP). This database, drawing on the County Business Patterns data, provides a classification system for mapping clusters within the US economy. The CMP identifies 41 clusters incorporating 589 "traded" industries. Traded industries are those that concentrate in particular regions and sell products or services across regions and countries, in contrast to local industries that primarily serve the local market.

The database includes attributes of cluster composition and economic performance at the region-cluster-industry level between 1990 and 2005, covering 177 mutually exclusive Economic Areas (EAs) in the contiguous United States. We explore several measures of the strength of related industries surrounding a region-industry. We refer to this group of measures as the "cluster environment," which includes the strength of the cluster in the region, the strength of related clusters in the region, and the strength of similar clusters in neighboring regions. For example, motor vehicles and car bodies (SIC-3711) is one of 15 industries in the automotive cluster. We look at the presence in a region of this industry and of the other 14 industries. The automotive cluster can also be linked to as many as six related clusters (such as metal manufacturing) that may be present in the region and to automotive clusters in geographically adjacent regions.

In order to examine the role of cluster agglomeration in the growth of the regional industries operating in a cluster, we must account for bias from unobserved factors. This includes the size of the region or policies associated with certain types of regions or industries that may be correlated with a region's cluster composition and subsequent performance. Our core models specify region-industry growth between 1990 and 2005 as a function of the initial size of the region-industry, the initial strength of the cluster environment around that region-industry, and region and industry fixed effects. By accounting for the overall growth rate of a given region and industry, we are able to examine the relationship between clusters and region-industry growth.

Our findings provide support for the distinct influences of convergence and cluster-based agglomeration. We find that the employment growth rate of a region-industry is *declining* in the initial level of employment at the region-industry, but is *increasing* in the employment strength of the cluster environment to which that industry belongs. This suggests that agglomeration economies arise within a regional cluster, across related clusters, and with the same cluster in neighboring regions. Based on our results, cluster agglomeration seems to matter for the employment growth of various types of industries, including both high-tech and low-tech manufacturing as well as service.

We also find evidence on the duality between employment and innovation performance in clusters. The initial employment and patenting strength of a regional cluster each have a separate positive relationship with region-industry employment growth. The positive effect of the patenting strength of the cluster on employment suggests that innovation in a cluster facilitates employment creation. Since both the employment and patenting strength of the cluster matter for growth, this suggests that multiple types of externalities are at work.

[1] Other studies examine the influence of various industry interdependencies, notably input–output, patenting, and occupational linkages, on the co-agglomeration of manufacturing industries in a region (Ellison and Glaeser, 1997; Ellison et al., 2010). However, they do not address the broader array of interdependencies associated with clusters of related industries in both service and manufacturing, and the role of clusters in regional industry performance. Frenken et al. (2007) move beyond urbanization economies by separating out the role of "related" and "unrelated" economic diversity in regional growth. We discuss these measures in Section 4.

The positive relationship between cluster strength and employment growth does not come at the expense of innovation. We find that patenting growth in a region-industry is increasing in the initial patenting and employment strength of the cluster to which that industry belongs. Multiple types of externalities that arise in clusters thus appear to contribute to innovation creation.

Finally, we find that regional cluster strength is positively associated with the emergence of new industries within the cluster. The strength of the cluster in neighboring regions also contributes to the emergence of new industries in the regional cluster. This suggests that both intra- and inter-regional spillovers play a role in the evolution of regional clusters.

Prior empirical studies (such as Feldman and Audretsch (1999)) have demonstrated the impact of science-based related industries on region-industry innovation performance, but our analysis suggests that the impact of complementarities across related industries is far broader and more pervasive. Rather than being confined to particular types of industries (manufacturing or high technology) or operating through particular channels (university–industry linkages), our results suggest that a broader set of complementarities arising across related industries contributes to the emergence and growth of regional industries. Our findings also generalize and extend the conclusions of more qualitative studies on the role of clusters (Porter, 1990, 1998; Swann, 1992; Bresnahan and Gambardella, 2004).

The remainder of the paper is organized as follows: Section 2 describes the role of clusters in the performance of regional industries. Section 3 presents the empirical framework. Section 4 explains the data and the cluster definitions. Section 5 discusses the findings on the role of the (employment-based) cluster environment in region-industry employment growth. Section 6 examines the duality between employment and innovation in clusters. Section 7 studies the role of the cluster environment in the emergence of new regional industries. A final section concludes.

2. Clusters and economic geography

The agglomeration of related economic activity is a central feature of economic geography (Marshall, 1920; Porter, 1990; Krugman, 1991; Ciccone and Hall, 1996; Ellison and Glaeser, 1997), but its prominence and role have been a puzzle. In a given location, limitations on resources can result in diminishing returns. This can lead to convergence in economic activity (employment, productivity) across regions over time (Barro and Sala-i-Martin, 1991; Higgins et al., 2006). However, the striking geographic concentration of related economic activity, with copious examples ranging from textiles in northern Italy to financial services in New York City, reveals the powerful role of agglomeration. Starting with Marshall (1920), economists have highlighted three distinct drivers of agglomeration: input–output linkages, labor market pooling, and knowledge spillovers. Each of these mechanisms is associated with cost or productivity advantages to firms that result in increasing returns to geographically proximate economic activity. Over time, an extensive literature has incorporated additional agglomeration drivers, including local demand, specialized institutions, and the structure of regional business and social networks (Porter, 1990, 1998; Saxenian, 1994; Markusen, 1996).

As suggested by Glaeser et al. (1992), the relationship between the initial structure of economic activity within a region and region-industry economic growth is subtle. Differing scopes of agglomeration forces may be at work. Agglomeration may arise from the specialization of a region in a particular industry (so-called "localization economies") or from exploiting the overall diversity of industries in an entire regional economy (so-called "urbanization economies").[2] Empirical identification of these agglomeration effects has been hampered because of the role of convergence on regional growth patterns.

Consider the relationship between the growth of economic activity and the initial level of economic activity within a region-industry. At the industry level of analysis, both convergence and agglomeration effects may be present. Region-industries may be subject to convergence effects (i.e., the coefficient on the initial *level* of economic activity is negative), either as the result of mean-reversion or diminishing returns.[3] The returns to economic activity can be diminishing due to cost-based competition and congestion costs. A large presence of firms in an industry relative to the size of the region can intensify local competition for inputs, dampening incentives for entry and business expansion. For example, if the price of specialized inputs is increasing in the number of local firms, there could be diminishing returns as a result of congestion costs (Swann, 1998; Sorenson and Audia, 2000; LaFountain, 2005; Duranton, 2007).

However, region-industries may also be subject to agglomeration effects. There may be externalities across firms within individual industries in learning, innovation, and spawning entrepreneurs (Audretsch, 1998; Henderson, 2003; Gompers et al., 2005; Glaeser and Kerr, 2009; Delgado et al., 2010). The empirical relationship between regional industry specialization and the growth of employment in that industry will be ambiguous, and will depend on the precise nature of competition (cost- or innovation-based) and the pattern of strategic interaction among firms.

While convergence may prevail at the region-industry level, it is possible to examine the agglomeration within the regional cluster that surrounds the particular region-industry (i.e., the set of other related industries co-located in the region). The presence of related economic activity – such as specialized suppliers, a large or advanced local customer base, producers of complementary products and services, and specialized institutions – may increase the pool of available inputs in a location and give rise to externalities of various sorts. This, in turn, may enhance the resources available for firm growth (Porter, 1990, 1998, 2003; Swann, 1998; Audretsch, 1998; Baptista and Swann, 1999).

After controlling for the convergence effect at the region-industry level, an industry participating in a cluster with a larger relative employment size should grow faster than the same industry in a region with limited presence of the cluster. A strong regional cluster (and related clusters) may enable agglomeration economies, including larger pools of skilled employees, knowledge spillovers, specialized suppliers, and sophisticated buyers. Proximity of related economic activity can also reduce transaction costs and induce the growth of specialized local institutions, such as educational programs and trade groups that reinforce the complementarities across related industries. Thus, a strong regional cluster should enhance the employment growth of the industries in the cluster through increasing efficiency, productivity, and/or returns to investment (Saxenian, 1994; Porter, 2003; Bresnahan and Gambardella, 2004; Bönte, 2004; Frenken et al., 2007; Delgado et al., 2010).[4] However, there could be congestion costs taking place

[2] The findings of the papers testing for localization and urbanization economies are mixed. See the reviews of Rosenthal and Strange (2004) and Beaudry and Schiffauerova (2009).

[3] We draw on the convergence concept used by the growth literature to study economic activity across countries, regions, and regional industries (Barro and Sala-i-Martin, 1991; Henderson et al., 1995; Higgins et al., 2006).

[4] The increase in employment in clusters could be partially attributed to the creation of new businesses. Prior work shows that complementarities across related economic activity facilitate entrepreneurship (Glaeser and Kerr, 2009; Feser et al., 2008; Delgado et al., 2010). However, the net effect of business creation on employment growth within a regional cluster will depend on the rate of firm churn and the

at the cluster-level as well. Larger regional clusters can generate congestion costs in related inputs markets (Ciccone and Hall, 1996; Swann, 1998; Sorenson and Audia, 2000). For example, this could occur if the industries in the cluster share inputs, and the costs of inputs are increasing with the relative size of the cluster.

In the empirical work, we also consider inter-regional spillovers, and examine the relationship between region-industry employment growth and the presence of the cluster in neighboring regions. Economic geography theory suggests that neighboring regions play an important role in shaping opportunities for growth (Fujita et al., 1999; Fujita and Thisse, 2002). Clusters co-located in nearby regions may benefit from inter-regional spillovers, which lower the costs of entrepreneurship and business expansion (e.g., by providing access to customers, by allowing firms to leverage proximate inputs, institutions, etc.). However, a strong cluster in a neighboring region can be a source of competition for constrained inputs and demand. The cluster strength in neighboring regions will have an ambiguous effect on the growth of regional industries, depending on the relative salience of inter-regional spillovers versus locational competition.

As discussed earlier, we expect that the employment strength of a cluster will facilitate the employment growth of the industries in the cluster. A related question is whether the innovation strength of the cluster contributes to employment growth. This analysis relates to the debate on whether the co-location of R&D and employment (production) is important for regional performance (Dertouzos et al., 1989; Porter and Rivkin, 2012; Berger, 2013). While innovation and production are each geographically concentrated (Audretsch and Feldman, 1996; Alcacer, 2006; Helper et al., 2012), the complementarities between these two activities are poorly understood.[5]

This literature motivates an empirical question concerning the duality between employment and innovation performance in clusters. The initial innovation strength of a cluster may exert knowledge spillovers that facilitate the employment growth of the region-industries participating in the cluster. Examining this requires to simultaneously account for the employment strength of the cluster. This analysis provides suggestive evidence regarding the types of mechanisms that are associated with employment growth in clusters. For example, if the innovation strength of the regional cluster solely matters for employment growth, knowledge spillovers would be the key driver of growth. If both the employment and innovation strength of the regional cluster matter for employment growth, this suggests that a broader set of mechanisms – such as knowledge spillovers, access to specialized inputs, and sophisticated demand – are at work.

While our analysis focuses on employment growth, we also examine innovation growth to better understand the interplay between these two performance dimensions. The innovation strength of a regional cluster may facilitate opportunities for innovation if knowledge spillovers arise among related industries (Audretsch, 1998; Baptista and Swann, 1998; Feldman and Audretsch, 1999; Aharonson et al., 2008). The employment strength of the cluster may also facilitate opportunities for innovation (e.g., as the result of input–output linkages or through the presence of innovation-seeking local demand). In the empirical analysis, we examine if, after controlling for the convergence effect at the region-industry level, the growth of innovation in a region-industry increases with the initial innovation and employment strength of the regional cluster. If both the employment and innovation strength of the cluster matter for innovation growth, this suggests that an array of mechanisms beyond knowledge spillovers are at work in clusters.

Our final analysis concerns the role of clusters in the diversification of regional economies. Prior studies suggest that the industry composition of a region affects the emergence of new industries in the region (Swann, 1992, 1998; Porter, 1998; Klepper, 2007, 2010; Neffke et al., 2011; Frenken et al., 2011). In a recent study of the evolution of Swedish regions, Neffke et al. (2011) found that industries that were technologically related to the pre-existing industries in a region had a higher probability of entering that region. Industry studies also suggest that the emergence of industries is affected by the presence of related activities in a region. Studies of the automotive industry in Detroit and the tire industry in Akron show that spinoffs from existing firms play an important role in the emergence of these industries, and these spinoffs tend to leverage related industry experiences in the new industries (Klepper, 2007, 2010; Buenstorf and Klepper, 2009). Building on these studies, we examine the role of the initial cluster strength in the subsequent emergence of related industries in the cluster.

3. Econometric model

To examine region-industry employment growth, we draw on studies of convergence that specify economic growth as a function of the level of economic activity and attributes of the region (Barro and Sala-i-Martin, 1991; Combes, 2000a). While convergence forces may prevail at the region-industry level, we argue that agglomeration forces may arise in clusters of related industries (Porter, 1998; Swann, 1998; Feldman and Audretsch, 1999). To test these ideas, we examine region-industry growth between 1990 and 2005 for 177 regions and up to 589 traded industries, totaling 55,083 region-industries with positive employment in 1990. To study the potential for cluster agglomeration in regional industry growth, we distinguish between the level of regional specialization in a particular industry and the strength of the cluster environment around that region-industry, and control for the average growth of the industry and the region. Our core econometric specification for region-industry employment growth is:

$$\begin{aligned}\ln\left(\frac{Employment_{icr,2005}}{Employment_{icr,1990}}\right) &= \alpha_0 + \delta \ln(Industry\quad Spec_{Employ,\, icr,1990})\\ &+ \beta_1 \ln(Cluster\quad Spec^{outside\, i}_{Employ,\, icr,1990})\\ &+ \beta_3 \ln(Related\quad Clusters\quad Spec^{outside\, c}_{Employ,\, cr,1990})\\ &+ \beta_3 \ln(Cluster\quad Spec\quad in\quad Neighbors_{Employ,\, cr,1990}) + \alpha_i + \alpha_r + \varepsilon_{icr}.\end{aligned} \tag{1}$$

The dependent variable is employment growth of the (four-digit SIC) industry i in cluster c in region (Economic Area (EA)) r over the period 1990–2005. The explanatory variables are specified at the initial date, 1990, which allows for the long-term effect of agglomeration economies on regional performance (Henderson, 1997). We compare the growth rates of EA-industries, accounting for the overall growth of the industry and the EA. The source of identification in Eq. (1) is the cross-sectional variation in industry and cluster composition across regions rather than exogenous changes in the cluster environment over time.

To estimate the combined effect of the convergence and agglomeration forces at the region-industry level, we specify an EA-industry growth model that includes (log of) regional

type of entrepreneurship (Rosenthal and Strange, 2009; Haltiwanger et al., 2009; Baptista and Preto, 2010).

[5] Theory predicts benefits from co-locating innovation and production in industries where the manufacturing process is not standardized (Vernon, 1966; Duranton and Puga, 2001). Some industry studies find results consistent with this theory (Pisano, 1997; Fuchs and Kirchain, 2010).

specialization in the industry (ln *Industry Spec$_{Employ}$*). This variable is an employment-based location quotient that captures to what extent industry employment is over-represented in the region (see Section 4 for a precise definition). In our model with industry and EA fixed effects, the variation in this variable is driven exclusively by variation in employment across EA-industries (i.e., the estimated coefficient for (log of) *Industry Spec* would be the same as for (log of) the level of employment in the EA-industry).[6]

To capture the cluster-driven agglomeration forces, we include three measures of the employment-based strength, or relative presence, of the cluster environment surrounding that region-industry (see Section 4 for a precise definition of these measures): *Cluster Spec$_{Employ}$* (a measure of cluster strength in the set of other closely related industries comprising the cluster), *Related Clusters Spec$_{Employ}$* (a measure of the strength of related clusters in the region), and *Cluster Spec in Neighbors$_{Employ}$* (a measure of cluster strength in adjacent geographic regions).

To illustrate the unit of observation and explanatory variables, consider the pharmaceutical preparations industry (SIC-2834) in the biopharmaceutical cluster in the Raleigh-Durham-Cary EA. For this region-industry, we look at the region's specialization in the industry (*Industry Spec$_{Employ}$*), in related industries in biopharmaceuticals excluding industry SIC-2834 (*Cluster Spec$_{Employ}$*), and in other related clusters (e.g., medical devices; *Related Clusters Spec$_{Employ}$*), and the adjacent regions' specialization in biopharmaceuticals (*Cluster Spec in Neighbors$_{Employ}$*).

We include industry (α_i) and region fixed effects (α_r) to control for other differences across industries and regions that affect region-industry growth. Our analysis accounts for unobserved factors – such as the size of the industry, the size of the region (urbanization economies), resource endowments, etc. – which might be correlated both with our explanatory variables and region-industry employment growth.

Our prediction concerning the coefficient on the initial relative size of an industry in the region (*Industry Spec*) is ambiguous, depending on the relative salience of convergence and agglomeration forces at the region-industry level. We expect industries in regions with strong clusters and related clusters to perform better than industries in regions lacking cluster strength ($\beta_1 > 0$ and $\beta_2 > 0$).

Importantly, we test whether the effect of cluster specialization is driven by industry complementarities within clusters or simply results from random aggregation of industries. We implement a Monte Carlo falsification test in which we construct "random" clusters by randomly assigning the 589 industries into clusters (see Section 5 for a detailed description of our procedure). By constructing random clusters, we are able to evaluate whether our results are simply an artifact arising from the inclusion of *any* sets of industries or whether they depend on clusters constructed based on inter-industry complementarities.

To account for correlation of the error terms across industries within a regional cluster, the standard errors are clustered by region-cluster. Finally, since nearby regions tend to specialize in like clusters, there might be unobserved spatial autocorrelation (Anselin, 1988). We account for this in two ways: by including in our specifications the strength of similar clusters in neighboring regions and by directly testing for spatial correlation using spatial econometric techniques.[7]

Parallel to the region-industry employment growth model specified in Eq. (1), we specify the following model to estimate the growth in patenting (a measure of innovation):

$$\ln\left(\frac{Patenting_{icr,2005}}{Patenting_{icr,1990}}\right) = \alpha_0 + \delta \ln(Industry\quad Spec_{Patent,\,icr,1990}) + \beta_1 \ln(Cluster\quad Spec^{outside\,i}_{Patent,\,icr,1990}) + \beta_2 \ln(Related\quad Clusters\quad Spec^{outside\,c}_{Patent,\,cr,1990}) + \beta_3 \ln(Cluster\quad Spec\quad in\quad Neighbors_{Patent,\,cr,1990}) + \alpha_i + \alpha_r + \varepsilon_{icr}. \tag{2}$$

The dependent variable is the growth in new patents granted in industry *i* in region (EA) *r* over the period 1990–2005. The explanatory variables are based on the patents granted in 1990 in the region-industry and in the cluster environment (see Table 1).

4. Data and cluster definitions

Data from the County Business Patterns (CBP) dataset is coded with cluster definitions from Porter (2003). Before turning to the precise variable definitions, it is useful to provide an overview of the data sources and the cluster definitions. The CBP dataset is a publicly available database that provides annual county-level measures of private-sector non-agricultural employment at the level of four-digit SIC codes (which we refer to as industries).[8] The data is aggregated to the region-industry level and the region-cluster level, using four-digit SIC codes as the primary industry unit,[9] and economic areas (EAs as defined by the US Bureau of Economic Analysis) as the main geographic unit.[10] The analysis focuses on the 1990–2005 period.

In addition to employment growth in a given EA-industry, we also examine the growth in patenting. The EA-industry patent data is drawn from the US Patent and Trademark Office (USPTO). This dataset includes detailed information about all utility patents, including inventor location and technology classification. Constructing patenting measures is complicated, both because USPTO patents are assigned to patent classes but are not directly matched to SIC codes, and because the multiple inventors of a given patent may be located in different regions. We utilize a patent-SIC code

[6] We are not trying to separate the convergence effect from agglomeration that occurs at the region-industry level, but rather to estimate the combined effect. Some studies have unsuccessfully tried to disentangle both effects by including in their models the employment level of the regional industry as well as the specialization of the region in the industry (Glaeser et al., 1992; Henderson et al., 1995). However, including both variables in the specification induces interpretation and identification problems (see discussion at Combes, 2000b).

[7] Following Anselin (1988) and the extensions developed by LeSage (1999), we test for spatial autocorrelation using a first-order spatial autoregressive (FAR) model. We estimate our OLS specification (Eq. (1)) and then estimate the FAR model: $\hat{\varepsilon}_{icr} = \rho W \hat{\varepsilon}_{icr} + \mu_{icr}$, where $\hat{\varepsilon}$ are the residuals from the OLS estimation, and W is an $N \times N$ matrix (where N is the total number of region-industries) with elements equal to 1 for adjacent regions and 0 otherwise. Under the null hypothesis of no spatial autocorrelation if $\rho = 0$.

[8] CBP data is made available at the county, state, and US level. One problem of the CBP data is cell suppression to protect the confidentiality of firms in a certain geography-industry with a small presence of firms. When employment data is suppressed, a range is reported. We utilize the mid-point in the range in our data. We use the unsuppressed Longitudinal Business Database (LBD) of the US Census Bureau to replicate the employment growth analysis reported in Table 2, and the results only change trivially.

[9] In order to use EA-industry data back to 1990, the analysis employs the SIC system rather than the NAICS system, which was introduced in 1997. By construction, recent NAICS-based data can be translated (with some noise) into the older SIC system. The use of NAICS or SIC definitions has no meaningful impact on our core empirical results.

[10] There are 179 EAs covering the entirety of the United States. To minimize concerns about differences in transportation costs and the definition of neighboring regions, we exclude the Alaska and Hawaii EAs. The EAs are meaningful economic regions and have been highly stable over time (Johnson and Kort, 2004).

Table 1
Definitions and descriptive statistics of variables ($N = 55{,}083$).

Variables	Definitions	Mean (Std. Dev.)
Employment growth$_{1990-05}$	EA-industry growth in employment $\ln(employment_{2005}/employment_{1990})$	−.320 (1.690)
Patenting growth$_{1990-05}$	EA-industry growth in patenting $\ln(patenting_{2005}/patenting_{1990})$	.243 (1.688)
Emergence$_{2005}$	Dummy equal to 1 if EA-industry exists as of 2005 (but did not as of 1990; $N = 48{,}430$)	.290 (.450)
Industry Spec$_{Employ}$	EA-industry employment-based Location Quotient (LQ) in 1990	2.340 (7.207)
Cluster Spec$_{Employ}$	EA-cluster employment-based LQ (outside the EA-industry) in 1990	1.315 (2.202)
Related Clusters Spec$_{Employ}$	Related clusters' employment-based LQ in EA (outside the EA-cluster) in 1990	1.153 (1.042)
Cluster Spec in Neighbors$_{Employ}$	Neighboring clusters' average employment-based LQ (outside the EA) in 1990	1.278 (1.278)
SIC2 SPEC$_{Employ}$	Two-digit SIC code employment-based LQ in EA (outside the EA-industry) in 1990	1.270 (1.937)
Industry Spec$_{Patent}$	EA-industry patent-based LQ in 1990	1.128 (2.941)
Cluster Spec$_{Patent}$	EA-cluster patent-based LQ (outside the EA-industry) in 1990	1.130 (1.360)
Related Clusters Spec$_{Patent}$	Related clusters' patent-based LQ in EA (outside the EA-cluster) in 1990	1.024 (.547)
Cluster Spec in Neighbors$_{Patent}$	Neighboring clusters' average patent-based LQ (outside the EA) in 1990	1.114 (.680)

Note: The employment and patent indicators are based on CBP and USPTO data, respectively.

concordance algorithm developed by Silverman (1999), in which USPTO patents are assigned, on a fractional basis, to four-digit SIC codes in a consistent (albeit somewhat noisy) manner. Each patent is then assigned, on a fractional basis, among the locations of the inventors.

4.1. Cluster definitions: groupings of related industries

The first step in creating a grouping of industries is to define a "similarity matrix" that captures the relatedness between any pair of industries. In the second step, some numerical method is used based on the similarity matrix to create groups (called "clusters") in a way that industries in the same cluster are more similar among themselves than to those in other clusters (Everitt et al., 2011; Delgado et al., 2013). A set of cluster definitions will provide the industry boundaries for any cluster (e.g., the Automotive cluster will contain 15 industries), and then clusters can be mapped into any regional unit (e.g., EAs in our analysis).

Some studies measure industry relatedness, but do not define clusters. In particular, Ellison and Glaeser (1997) and Ellison et al. (2010) measure the co-location between industries using the conglomeration index, but do not create industry groupings. Only a few studies define clusters, including Feser and Bergman (2000) and Feser (2005) for input–output-based clusters; and Porter (2003) for industries related by various types of linkages in both service and manufacturing.[11]

In this paper, the cluster definitions are drawn from the US Cluster Mapping Project (Porter, 2003). The CMP develops a methodology for grouping four-digit SIC codes into clusters. The methodology first distinguishes three types of industries with distinct competition and locational drivers: local, natural resource-dependent, and traded. Local industries are those that serve primarily the local markets (e.g., utilities) whose employment is roughly evenly distributed across regions. Natural resource-dependent industries are those whose location is tied to local resource availability (e.g., logging). Traded industries are those that tend to be geographically concentrated and produce goods and services that are sold across regions.[12]

Traded industries are the focus of our analysis. To measure the relatedness between any pair of industries *ij*, the CMP computes the pairwise correlation of industry employment across locations *r* (Correlation coefficient ($employment_{ir}$, $employment_{jr}$)) in 1996. This measure is referred to as "locational correlation" of employment, and is based on revealed co-location patterns to capture any type of potential externalities (skills, demand, supply, knowledge, and others) present across industries.[13] Using the locational correlation of employment, the CMP then groups each of the 589 industries to one of 41 mutually exclusive clusters (referred to as "narrow" cluster definitions)[14] Within a cluster such as information technology, for example, there are nine four-digit SIC code industries, including electronic computers (SIC-3571) and software (SIC-7372), reflecting the fact that location of employment in computer hardware and software is highly correlated.[15] The resulting clusters may contain service and manufacturing industries and industries from different parts of the SIC system.[16]

The CMP also develops a "broad" cluster definition in which a given industry may be shared by multiple clusters (as inferred through the locational correlation of employment patterns). While the narrow clusters form our key measures of related industries, we use the broader clusters to develop a measure of the strength of related clusters surrounding a given region-cluster.

4.1.1. Other industry groupings

Delgado et al. (2013) create a new clustering methodology to generate and assess sets of cluster definitions. They validate

[11] Ellison and Glaeser (1997) study the co-agglomeration of pairs of manufacturing industries, building an index that reflects "excess" concentration. In a related paper, Ellison et al. (2010) test various mechanisms (input–output, knowledge, labor pooling) inducing the co-agglomeration of pairs of industries, and conclude that they all seem to matter. This finding shows that there are many types of industry complementarities (Forni and Paba, 2002), and is consistent with the clustering methodology developed in Porter (2003). In contrast, other studies focus on specific types of linkages in terms of input–output (Feser, 2005), and technological, science-based, and market proximity (Scherer, 1982; Jaffe et al., 1993; Feldman and Audretsch, 1999; Neffke et al., 2011; Bloom et al., 2012). Other studies focus on characterizing particular sets of regional industries by examining their types of firms and their networks (Saxenian, 1994; Boschma, 2005; Aharonson et al., 2008; Lorenzen and Mudambi, 2013).

[12] Traded industries account for 30% of US employment and over 87% of US patents.

[13] The locational correlation of employment and the resulting CMP cluster definitions are derived based on 1996 state-industry data and are fixed over time. Pairwise locational correlation of employment change little over time, so we would not expect major changes in the cluster definitions during our time period.

[14] Porter (2003) describes the methodology. The primary classifications are based on locational correlation of employment of region-industry pairs. However, it is possible that industries with high employment co-location may have little economic relationship. Two adjustments are made to eliminate spurious correlations. First, the SIC industry definitions are used to reveal logical links. Second, the National Input–Output accounts are used to identify meaningful cross-industry flows.

[15] See Porter (2003) for a comprehensive list of the 41 traded clusters and some key attributes.

[16] On average, clusters have industries with four different two-digit SIC codes.

the CMP definitions, and show that the individual clusters contain industries that are highly related based on multiple types of linkages. There are other industry groupings that could capture agglomeration that arises among related industries. The simplest grouping will be industries with the same two-digit SIC code used in other studies (Baptista and Swann, 1998; Frenken et al., 2007). This set groups industries that are related based on their list of products/services; but it fails to capture linkages between industries with different two-digit SICs (e.g., manufacturing and services).[17] Thus, groupings based purely on industry definitions may perform less effectively in capturing various types of inter-industry linkages than CMP definitions.

Delgado et al. (2013) propose a new set of cluster definitions – referred to here as the Benchmark Cluster Definition BCD – that incorporates current data and multiple industry similarity matrices based on input–output, labor occupations, and the co-location patterns of employment and establishments. The BCD groups a similar set of industries as the CMP, but it uses the 2007 NAICS code. Although the BCD has a significant overlap with the CMP, there are also relevant differences since new data and clustering method are used.[18]

While the CMP cluster definitions are our key measures of clusters, we use the two-digit SIC groupings and the BCD as alternative industry groupings to examine economies of agglomeration that arise in groups of related industries.

4.2. Sample description and dependent variable definitions

The empirical analysis focuses on explaining the growth of existing regional industries by using the sample of EA-industries that have a positive level of employment in 1990 (55,083 observations). For this sample, we define two dependent variables: EA-industry growth in employment and in patenting over the period 1990–2005 ($\ln(Employment_{ir,2005}/Employment_{ir,1990})$ and $\ln(Patenting_{ir,2005}/Patenting_{ir,1990})$).[19] To compute the employment growth rate for EA-industries where we observe zero employment in 2005, we set employment equal to one for these observations.[20] We implement numerous sensitivity analyses involving alternative approaches to deal with the zeros. We use the sample of regional industries that have positive employment in both the base and the terminal period (sample of 48,921 observations); vary the base and terminal periods (1990–1996 and 1997–2005 versus 1990–2005); and use the whole sample (including EA-industries with zero employment in 1990). Our main findings remain essentially unchanged.

Additionally, we examine the emergence of new-to-the-region industries by using the sample of EA-industries with zero measured employment in 1990 (48,430 observations). The dependent variable is a dummy variable equal to one if there is any positive level of employment as of 2005 ($Emergence_{2005}$; with a mean of 0.29 and a standard deviation of 0.45). These analyses include probit specifications of the probability of having industries new to the region as of 2005 as a function of the initial cluster strength.

4.3. Explanatory variables

To define our variables of the strength of a region-industry and the strength of the cluster environment that surrounds the region-industry, we draw on a body of prior work that uses location quotient (LQ) as a primary measure of regional specialization (Glaeser et al., 1992; Porter, 2003). We compute two sets of explanatory variables – one set based on employment and another set based on new patenting activity in 1990. We explain below the employment-based variables, and the patent-based variables are defined similarly. (Table 1 shows the descriptive statistics for all the variables.)

4.3.1. Industry specialization

Employment-based industry specialization in a region in the base year is measured by the share of regional employment in the regional industry as compared to the share of US employment in the national industry:

$$Industry\ \ Spec_{Employ,i,r,90} = \frac{employ_{i,r}/employ_r}{employ_{i,US}/employ_{US}},$$

where r and i indicate the region (EA) and the industry, respectively. This indicator captures the degree to which the industry is "over-represented" in terms of employment in the EA. In our sample, the $Industry\ Spec_{Employ}$ variable has a mean of 2.34 and a standard deviation of 7.21 (Table 1). As noted earlier, we specify this variable in log form and include EA and industry fixed effects in our model (Eq. (1)); and so, the independent variation is driven exclusively by variation in employment across EA-industries.

4.3.2. Cluster specialization

We use a similar procedure to develop a measure for cluster specialization. For a particular EA-industry, the specialization of the EA in cluster c is measured by the share of regional employment in the regional cluster (*outside the industry*) as compared to the share of US employment in the national cluster (*outside the industry*):

$$Cluster\ \ Spec_{Employ,icr,90} = \frac{employ^{outside\,i}_{c,r}/employ_r}{employ^{outside\,i}_{c,US}/employ_{US}}.$$

The average $Cluster\ Spec_{Employ}$ is 1.31 (and the standard deviation is 2.20; Table 1). Since cluster specialization is measured relative to the overall size of the region, a region may exhibit specialization in a particular cluster even though that region only holds a small share of the national employment of that cluster. While it is not surprising that leading regions in the automotive cluster include Detroit-Warren-Flint (MI) and Cleveland-Akron-Elyria (OH), there are pockets of automotive cluster strength in smaller regions, such as Lexington-Fayette-Frankfort-Richmond (KY) and Louisville-Elizabethtown-Scottsburg (KY-IN).

It is important to note that the correlation between industry specialization and cluster specialization is small (0.20). Some regions can have a high industry specialization, but low strength in other related industries that constitute the cluster, and vice versa. These two variables, while related, seem to capture different agglomeration forces – those that take place within the focal industry versus those that take place within the cluster.

4.3.3. Specialization in related clusters

To measure the strength of related clusters, we use the broad cluster definitions in Porter (2003). We identify as clusters related to a given cluster c those broad clusters that have at least one of c's narrow industries in common. For example, in the case of Automotive, related clusters include Production Technology and Metal Manufacturing, among others. Having identified the set of clusters

[17] For example, Frenken et al. (2007) define *related* diversity in a region by examining the diversity of industries within a given two-digit industry class (i.e., entropy at the five-digit level industries within each two-digit industry class); and find that related diversity enhances regional employment growth.

[18] The BCD will be available at the US Cluster Mapping website (http://clustermapping.us).

[19] $Patenting_{irt}$ are the new patents granted in year t at industry i in EA r.

[20] Similarly, to compute the patenting growth of EA-industries with zero patenting in 1990 or 2005, we set patenting equal to the minimum in the sample.

related to a given cluster (C_c), we then measure the degree of overlap between each pair of clusters (c, b) using the average proportion of narrow industries that are shared in each direction[21]:

$$\omega_{c,b} = Avg\left(\frac{shared\quad industries_{c,b}}{narrow\quad industries_c}, \frac{shared\quad industries_{b,c}}{narrow\quad industries_b}\right).$$

The strength of a region in clusters related to c is then defined by a weighted sum of the location quotients associated with each (narrow) related cluster:

$$Related\quad Clusters\quad Spec_{Employ,\,cr} = \frac{\left(\sum_{b\in C_c}(\omega_{c,b} * employ_{b,r})\right) / \left(\sum_{b\in C_c}(\omega_{c,b} * employ_{b,US})\right)}{employ_r/employ_{US}}.$$

For instance, based on this weighting, which emphasizes the degree of overlap between clusters, our measure of the strength of related clusters for the Automotive cluster weights the presence of the Metal Manufacturing cluster more heavily than that of the Furniture cluster.

4.3.4. Strength in neighboring clusters

Specialization in a particular cluster may be spatially correlated across neighboring regions – for example, the historical strength of the Automotive cluster near Detroit is likely reinforced by cluster specialization in Automotive in neighboring regions.[22] We develop a measure of the presence of similar clusters in neighboring EAs (*Cluster Spec in Neighbors* variable) to explore the role of interregional cluster spillovers in the growth of a given region-industry. The inclusion of this variable allows controlling for the fact that the geographical boundaries of clusters may expand across regions.[23] We compute this variable as the average specialization (location quotient) of adjacent EAs in the cluster.

5. Clusters and employment growth

We now turn to the key findings on the role of clusters in region-industry economic performance. Table 2 provides our main econometric specifications of how the growth rate of EA-industry employment varies with the initial level of industry specialization and the cluster environment in the region. We begin in (2-1) with a simple model relating EA-industry employment growth to the (log of) initial level of employment-based industry specialization (*Industry Spec*$_{Employ}$), and a comprehensive set of industry and EA fixed effects. The estimated coefficient is negative, suggesting that convergence dominates the impact of agglomeration at the region-industry level.

In model 2-2, we introduce a second variable (*Cluster Spec*$_{Employ}$), the initial employment-based specialization of the region in the set of related industries comprising the cluster (excluding the focal region-industry). This is our core specification, relying exclusively on variation in the employment strength of the cluster, conditioning on the overall growth rate for the region and industry.[24] While the coefficient on *Industry Spec* continues to be negative (and roughly of the same magnitude), the coefficient on cluster specialization is positive and significant. The estimated annual rate of convergence in region-industry employment is 3.5%, which is similar to the ∼4% annual rate of convergence in city-sector employment found by Henderson et al. (1995).[25] At the same time, a one standard-deviation increase above the mean in cluster specialization is associated with a 1.3 percentage point increase in the expected *annual* employment growth of region-industries.[26] This positive relationship between cluster specialization and region-industry growth holds for region-industries with either low or high relative size.[27]

While the findings in model 2-2 suggest that agglomeration forces arise in strong clusters, congestion costs could be at work as well. Congestion costs could increase with the relative size of the regional cluster and at some point dominate the agglomeration economies. For example, this could occur if the industries participating in the cluster share inputs and the costs of inputs increase with the size of the cluster. In a sensitivity analysis, we examine non-linearities in the cluster specialization effect in two ways (not reported). We first estimate the baseline model 2-2 using a set of four dummies for the quartile values of cluster specialization (the omitted dummy is the lowest quartile of *Cluster Spec*$_{Employ}$). The estimated coefficients of the quartile dummies are each positive and significant, and increasing for higher quartiles. Furthermore, we estimate model 2-2, allowing the coefficient of ln *Cluster Spec*$_{Employ}$ to vary for each of its four quartile ranges. The four estimated coefficients are positive and significant. The findings suggest that agglomeration effects prevail at the region-cluster level even for the strongest (top quartile) regional clusters.

In model 2-3, we introduce two additional dimensions of the cluster environment: strength in related clusters (*Related Cluster Spec*$_{Employ}$) and the presence of strong clusters in neighboring regions (*Cluster Spec in Neighbors*$_{Employ}$). We find positive and significant coefficients for each of the three measures of the cluster environment surrounding a particular region-industry. A one standard-deviation increase above the mean in each aspect of the cluster environment is associated with a 1.9 percentage point increase in the expected *annual* employment growth (1.0%, 0.4%, and 0.5% for *Cluster Spec*$_{Employ}$, *Related Clusters Spec*$_{Employ}$, and *Cluster Spec in Neighbors*$_{Employ}$, respectively).[28]

Table 2 includes a number of robustness checks by re-estimating the baseline model 2-2 using alternative industry groupings to measure regional clusters. In model 2-4, we run a Monte Carlo falsification test in which we construct random clusters by randomly assigning the 589 industries into 41 clusters (without

[21] For example, Automotive shares five narrow industries (out of 15) with the broad Production Technology, and Production Technology shares seven narrow industries (out of 23) with the broad Automotive. Their overlap is 0.32.

[22] Table A1 in the Online Appendix shows that the specialization of a region in a given cluster is significantly correlated to the average specialization of neighbors in the same cluster (correlation coefficient of 0.50).

[23] Other studies use spatial density analysis to identify the geographical boundaries of individual industries rather than using administrative regional units (Duranton and Overman, 2005).

[24] In our sample of EA-industries with positive employment in 1990, there are observations with zero employment in the cluster (outside the industry), in related clusters, or in neighboring clusters. To use these observations, we replace ln(variable) with its minimum value in the sample. To control for unobserved attributes of these types of observations, we add across all models a dummy equal to one if the particular variable was corrected. For example, model (2-2) includes a dummy equal to one if an EA-industry has zero employment in 1990 in the cluster (outside the industry). All our findings only change trivially when dropping these dummies.

[25] The annual convergence rate is $100 \times (\ln(\delta+1))/-15)$ following Barro and Sala-i-Martin (1991).

[26] The estimated 1.3% magnitude effect is computed as $100 \times (\ln(3.517) - \ln(1.315)) \times 0.2$ divided by 15 years.

[27] To test this, we estimate the baseline model (2-2), allowing the coefficient of *Cluster Spec*$_{Employ}$ to vary for EA-industries with low versus high *Industry Spec* (i.e., below versus above the median value in 1990).

[28] The findings in our core models (2-2 and 2-3) are robust to a number of sensitivity checks in the sample used, such as using the non-zero sub-sample (i.e., EA-industries with positive employment in both 1990 and 2005), using all EA-industries (including those with zero employment in 1990), and using Metropolitan Statistical Areas as the regional unit. The positive relationship between the cluster environment variables and region-industry employment growth is also robust to excluding the industry and region fixed effects. Finally, we also test for spatial correlation running a FAR model on the residuals of models 2-2 and 2-3. We cannot reject the null hypothesis of no spatial autocorrelation.

Table 2
EA-industry employment growth over 1990–2005.

	Employment growth$_{1990–05}$ (N = 55,083)					
	2-1	2-2	2-3	2-4 Random Clusters	2-5 SIC2	2-6 BCD, Growth$_{2003-11}$ N = 80,240
$\ln$ *Industry Spec*$_{Employ}$	**−.355 (.006)**	**−.395 (.006)**	**−.405 (.006)**	**−.355 (.000)**	**−.388 (.006)**	**−.223 (.005)**
$\ln$ *Cluster Spec*$_{employ}$		**.200 (.008)**	**.149 (.009)**			
$\ln$ *Related Clusters Spec*$_{Employ}$			**.091 (.013)**			
$\ln$ *Cluster Spec in Neighbors*$_{Employ}$			**.104 (.012)**			
$\ln$ *Random-Cluster Spec*$_{Employ}$				−.001 (.011)		
$\ln$ *SIC2 Spec*$_{Employ}$					**.197 (.008)**	
$\ln$ *Cluster Spec*$_{BCD,Employ}$						**.176 (.007)**
Region (EA) fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	.369	.380	.382		.378	.195

Notes: Bold and bold-italic numbers refer to coefficients significant at 1% and 5% levels, respectively. Standard errors are clustered by EA-cluster. Model 2-4 uses random clusters and reports the bootstrapped standard errors. In model 2-5, industries are grouped by two-digit SIC code, and the standard errors are clustered by EA-SIC2. Model 2-6 uses the BCD industry grouping. Models 2-2 to 2-6 include a dummy equal to 1 if there was zero employment in 1990 in the cluster; and model 2-3 also includes dummies equal to 1 if there was zero employment in the related clusters or in the neighboring clusters.

replacement). We repeat this process to create 2000 simulated cluster definitions, and select the 200 cluster definitions that have the closest distribution to our original cluster definitions (where "closest" is defined in terms of the distribution of size of clusters). For each of the 200 random cluster definitions, we estimate the exact same specification as model 2-2. Model 2-4 reports the average coefficient from this exercise (with the standard errors based on the empirical distribution from the simulation exercise). The coefficient associated with the random cluster specialization measures are essentially zero and statistically insignificant. In other words, cluster-driven agglomeration depends critically on grouping industries into meaningful groups of related activities (Porter, 2003), and does not simply reflect random industry groupings.

While the CMP definitions are meaningful, there are other industry groupings that could capture relevant inter-industry interdependencies (Delgado et al., 2013). As discussed in Section 4, we use two alternative groupings: two-digit SIC code groupings in model 2-5 and the BCD in model 2-6.

In model 2-5, we define related industries as all four-digit SIC code (traded) industries within the same two-digit SIC. This grouping of industries captures some forms of relatedness, but fails to capture complementarities among service and manufacturing industries. Despite these limitations, the relationship between specialization in same two-digit SIC industries (*SIC2 Spec*) and EA-industry growth is positive and significant, consistent with our earlier findings. To assess the differences between the CMP and SIC2 grouping, we also re-estimate model 2-2 adding the specialization in clusters and in related clusters as well as the specialization in same two-digit SIC industries (not reported). The estimated coefficients of all these variables are positive and statistically significant, with the joint estimated coefficients of *Cluster Spec* and *Related Cluster Spec* significantly higher than that of *SIC2 Spec*.[29] These findings suggest that our baseline cluster variables capture relevant inter-industry linkages beyond those based purely on industry definitions, but may not capture all types.

Finally, in model 2-6 we use the BCD cluster definitions, which rely on six-digit NAICS industries. Since the BCD is based on more recent data, we examine region-industry employment growth in a later time period (2003–2011). The relationship between cluster specialization and EA-industry growth is positive and significant, consistent with the baseline model 2-2.

[29] A one standard-deviation increase above the mean in these variables is associated with a 0.9%, 0.3%, and 0.7% increase in the expected annual employment growth, respectively.

5.1. When do clusters matter more for employment growth?

While our findings highlight a positive relationship between the initial cluster strength and region-industry employment growth, prior work suggests that the impact of agglomeration forces can vary across types of regions, clusters, and industries. For example, some of the complementarities harnessed by strong clusters may be subject to economies of scale. Thus, the size of the local region can matter.

We examine how the effect of cluster specialization on region-industry employment growth varies by region size, US cluster size, and industry types in Table 3. We investigate the role of region size in model 3-1 by introducing a set of interaction effects between industry and cluster specialization and a dummy variable (*Large EA*), which is equal to one for large regions (above the median employment). The main findings hold for large and small EAs – convergence at region-industry level and agglomeration at the cluster level. However, the estimated region-industry convergence effect is higher in larger EAs. Region-industries tend to have higher employment levels in larger EAs and are subject to greater reversion to the mean. Notably, the impact of clusters on employment growth is greater in larger EAs as well. The size of the local market, greater economies of scale in specialized inputs, and the supply of public goods in larger regions, such as universities, could explain this result.

To better examine how region heterogeneity affects our results, we re-estimate the baseline employment growth model (model 2-2 in Table 2) allowing for EA-specific coefficients for industry and cluster specialization. The estimated coefficients of $\ln$ *Cluster Spec*$_{Employ}$ for each of the 177 EAs are plotted in Fig. 1a, separating out large and small regions. The EA-specific industry specialization coefficients are always negative (not reported). In contrast, the EA-specific cluster specialization coefficients are always positive; and larger EAs tend to have greater estimated cluster effects. The findings suggest that cluster agglomeration is not the result of a set of outlier regions.

Next, we examine the sensitivity of our results to cluster heterogeneity. We start in model 3-2 by studying the role of size of the national cluster. We introduce a set of interaction effects between industry and cluster specialization and a dummy variable (*Large Cluster*), which is equal to one for the top-20 US clusters based on total employment as of 1990 (see Table A2 in Online Appendix for a list of large clusters). The main findings hold for large and small US clusters. However, the estimated cluster effect is higher in large US clusters, which have a broader array of industries or larger individual industries, and may better exploit economies of scope and scale.

Table 3
EA-industry employment growth – region size, cluster size, and industry type.

	Employment Growth$_{1990-05}$ (N = 55,083)		
	EA size	Cluster size	Industry type
	3-1	3-2	3-3
ln *Industry Spec*$_{Employ}$	**−.352 (.010)**	**−.322 (.010)**	**−.482 (.010)**
ln *Cluster Spec*$_{Employ}$	**.166 (.011)**	**.170 (.011)**	**.194 (.012)**
Large EA × ln *Industry Spec*$_{Employ}$	**−.069 (.012)**		
Large EA × ln *Cluster Spec*$_{Employ}$	**.064 (.013)**		
Large cluster × ln *Industry Spec*$_{Employ}$		**−.112 (.013)**	
Large cluster × ln *Cluster Spec*$_{Employ}$		**.047 (.014)**	
High-tech × ln *Industry Spec*$_{Employ}$			**.093 (.015)**
High-tech × ln *Cluster Spec*$_{Employ}$			***.038 (.016)***
Low-tech × ln *Industry Spec*$_{Employ}$			**.122 (.013)**
Low-tech × ln *Cluster Spec*$_{Employ}$			−.009 (.015)
EA fixed effects	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes
R-Squared	.381	.382	.381

Notes: Bold and bold-italic numbers refer to coefficients significant at 1% and 5% levels, respectively. See notes in Table 2. In model 3-3, the omitted industry type is services. High-tech and low-tech are dummies for manufacturing industries with high and low patent intensity.

We further investigate cluster heterogeneity by estimating the baseline employment growth model (model 2-2 in Table 2) allowing for cluster-specific coefficients for the industry and cluster specialization variables. The estimated coefficients of ln *Cluster Spec*$_{Employ}$ for each of the 41 clusters are plotted in Fig. 1b, separating out large and small US clusters. The cluster specialization coefficients are positive and statistically significant except for two clusters (distribution services and entertainment clusters; see Table A2), and they are, on average, higher for larger US clusters. The findings suggest that cluster agglomeration is not driven by particular types of clusters.

Finally, we examine how the effect of cluster specialization on region-industry employment growth varies by industry types. Industries are heterogeneous in many dimensions, including the propensity to patent (higher in manufacturing) and the diversity of consumers (higher in service). Industries can then vary in the extent of agglomeration benefits within clusters. We examine these differences in Table 3 in model 3-3 by defining three mutually exclusive types of industries. A total of 146 industries are classified as service industries based on SIC codes. The manufacturing industries are classified into "high-tech" and "low-tech" based on patent intensity in 1990. A total of 147 industries with two or more patents granted by 1000 workers (higher quartile of patent intensity) are classified as high-tech manufacturing; and the other 296 industries are low-tech manufacturing. We find that cluster agglomeration seems to matter for the employment growth of these three industry types. This result adds to the previous literature that focused primarily on the effect of external agglomerations on high-tech manufacturing industries. There are also meaningful differences across industry types. High-tech manufacturing industries experience significantly lower convergence and higher cluster effects on employment growth than do service industries.

Our results suggest that industries within regional clusters with larger (relative) employment size have greater opportunities for employment growth, a finding that is robust across regions, clusters, and industry types. We recognize that this relationship could be correlated with unobserved region-cluster and region-industry specific policies and initiatives, such as the presence of local institutions or policies that support the growth of a particular regional cluster, or lower tax rates for the smaller industries within a larger cluster. While we cannot make a causal assertion in the absence of exogenous variation in the cluster environment, the robustness of our findings across regions and clusters suggests a positive relationship between the initial employment-based cluster strength and the employment growth of industries within the cluster.

6. The duality between employment and innovation performance within clusters

We have shown that the initial employment strength of a cluster seems to contribute to region-industry employment growth. Next, we explore whether the initial innovation strength of a cluster generates knowledge spillovers that facilitate region-industry employment growth.

We examine the duality between employment and innovation performance in clusters in Table 4 by characterizing the cluster strength using both employment and patenting (ln *Cluster Spec*$_{Employ}$ and ln *Cluster Spec*$_{Patent}$). These two variables have a

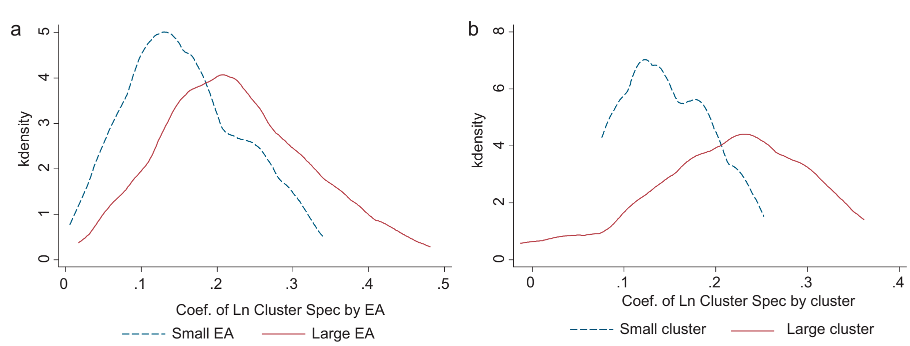

Fig. 1. EA-industry employment growth – cluster effects by region and by cluster. (a) Coef. of ln *Cluster Spec* by region (177 EAs). (b) Coef. of ln *Cluster Spec* by cluster (41 clusters). *Notes*: The graphs plot the Kernel density (kdensity) of the estimated coefficients of ln *Cluster Spec*$_{Employ}$ from estimating model 2-2 (Table 2), allowing the ln *Industry Spec*$_{Employ}$ and ln *Cluster Spec*$_{Employ}$ coefficients to vary for each EA (a) or each cluster (b).

Table 4
EA-industry employment growth – the duality of employment and patenting in clusters.

	Employment Growth$_{1990-05}$ (N = 55,083)	
	4-1	4-2
ln *Industry Spec*$_{Employ}$	−**.397 (.006)**	−**.406 (.006)**
ln *Cluster Spec*$_{Employ}$	**.190 (.008)**	**.143 (.009)**
ln *Cluster Spec*$_{Patent}$	**.060 (.010)**	**.046 (.010)**
ln *Related Clusters Spec*$_{Employ}$		**.083 (.013)**
ln *Related Clusters Spec*$_{Patent}$		−**.002 (.018)**
ln *Cluster Spec in Neighbors*$_{Employ}$		**.096 (.013)**
ln *Cluster Spec in Neighbors*$_{Patent}$		.030 (.016)
EA fixed effects	Yes	Yes
Industry fixed effects	Yes	Yes
R-Squared	.381	.383

Note: Bold numbers refer to coefficients significant at 1% level. See notes in Table 2.

modest positive correlation (0.24), which suggests that they emphasize different types of agglomeration. Some regional clusters have high employment strength and low patenting strength as of 1990 (e.g., the chemical cluster in Memphis, TN-MS-AR), and others have high patenting strength and low employment strength (e.g., the chemical cluster in New York-Newark-Bridgeport, NY-NJ-CT-PA).[30]

The employment strength of a cluster captures most types of supply-side and demand-side externalities (input–output, skills, technology, etc.). The patenting strength of a cluster emphasizes the potential for knowledge externalities, but could also capture input–output and other externalities. Controlling for employment-based cluster strength, we assess whether patenting-based cluster strength relates positively to region-industry employment growth. This would suggest that innovation in the cluster is important for the employment growth of the industries in the cluster, and that knowledge linkages may be a relevant mechanism at work.

We begin in model 4-1 by estimating region-industry employment growth using a modified version of Eq. (1), with the initial cluster strength based on patenting and employment. We find that the patenting strength and the employment strength of the cluster both seem to contribute to industry employment growth.[31] The positive estimated effect of patenting cluster strength suggests that knowledge externalities matter for employment creation. The positive estimated coefficient of employment cluster strength suggests that a broader set of mechanisms, such as access to specialized inputs and sophisticated demand, are also at work. However, we cannot assess the relative importance of each type of mechanism.

Model 4-2 shows the most comprehensive specification with all the cluster environment variables (cluster, related clusters, and neighboring clusters) based on employment and patenting. It confirms the findings in model 4-1. The estimated effects of the employment and patenting strength in the cluster are each positive and significant. In contrast, only the employment strength in related clusters and in neighboring clusters seems to matter for employment growth. This suggests that the complementarities between employment and innovation may be more relevant within the focal cluster than with related or neighboring clusters.

6.1. Clusters and patenting growth

The analysis thus far has focused on the role of regional clusters in employment growth. To better understand the interplay between innovation and employment performance in clusters, we also examine the role of regional clusters in region-industry innovation growth. Is a region-industry patenting growth facilitated by the pre-existing patenting in the cluster, by the pre-existing employment in the cluster, or by both?

We examine these questions in Table 5. We begin with model 5-1, which relates EA-industry patenting growth to the initial level of patenting-based industry specialization and cluster specialization, and a comprehensive set of industry and EA fixed effects. We find that a higher initial level of patenting in an EA-industry is associated with a reduction in the growth rate in the number of granted patents. The estimated annual rate of convergence is above 11%.[32] However, an initially stronger cluster based on patenting is associated with an increase in the growth rate of patenting. The strength of related clusters and neighboring clusters also contributes to patenting growth (model 5-2). A one standard-deviation increase above the mean in each aspect of the patenting-based cluster environment is associated with a 1.2 percentage point increase in the expected *annual* patenting growth (0.3%, 0.4%, and 0.5% for *Cluster Spec*$_{Patent}$, *Related Clusters Spec*$_{Patent}$, and *Cluster Spec in Neighbors*$_{Patent}$, respectively). This positive relationship between clusters and patenting growth is consistent with studies that find a firm's innovative activity relates to the presence of knowledge externalities in a location (Audretsch, 1998; Baptista and Swann, 1998; Feldman and Audretsch, 1999).[33]

Next, we explore whether there are complementarities between employment and patenting within clusters that improve the patenting of the industries in the clusters. We examine this in model 5-3 by estimating a modified version of Eq. (2), with the initial cluster strength measured based both on employment and patenting. The patenting and employment strength of a cluster each have a positive and significant relationship with patenting growth, with no statistical difference between the variables. The positive estimated coefficient of *Cluster Spec*$_{Employ}$ suggests that the employment strength of the cluster facilitates opportunities for innovation. The most comprehensive model (5-4) – which includes all the cluster environment variables based on both employment and patenting – shows that all these variables have a positive estimated coefficient. These findings are consistent with a broad set of externalities arising in clusters, related clusters, and neighboring clusters.

Overall, our findings suggest that there are complementarities between employment and innovation performance in clusters. Region-industry growth in employment and patenting are each positively associated with the pre-existing cluster strength in terms of both employment and innovation. Importantly, these results hold for each type of industry (service, low-tech manufacturing, and high-tech manufacturing). We interpret this as suggestive evidence

[30] The chemical cluster in Memphis, TN-MS-AR contained 16 industries; for each industry, the employment specialization of the region in the other 15 industries was high (*Cluster Spec*$_{Employ}$ above 2 – top quartile), while the patenting specialization was low (*Cluster Spec*$_{Patent}$ around 0.5 – below median). In contrast, the chemical cluster in New York-Newark-Bridgeport, NY-NJ-CT-PA had 20 industries; for each industry, the employment specialization of the region in the other 19 industries was low (below 1), while the patenting specialization was high (above 1.5).

[31] The coefficient of the employment-based cluster strength is significantly higher than that of the patenting-based cluster strength. A one standard-deviation increase above the mean for *Cluster Spec*$_{Employ}$ and *Cluster Spec*$_{Patent}$ is associated with a 1.2% and 0.3% increase in the annual employment growth of an EA-industry, respectively.

[32] The large convergence rate for region-industry patenting is consistent with large convergence effects in region-industry wages and productivity in other studies (Henderson et al., 1995; Cingano and Schivardi, 2004).

[33] Similarly to the analysis in Table 3, we examine the impact of region size, cluster size, and industry type on industry patenting growth. Agglomeration at the cluster level holds for large and small regions and US clusters. The initial patenting strength of the cluster matters for the patenting growth of all types of industries, but the estimated coefficient is significantly larger for high-tech industries than for service industries.

Table 5
EA-industry patenting growth – the duality of employment and patenting in clusters.

	Patenting Growth$_{1990–05}$ (N = 55,083)			
	5-1	5-2	5-3	5-4
ln *Industry Spec*$_{Patent}$	**−.807 (.010)**	**−.812 (.010)**	**−.814 (.010)**	**−.818 (.010)**
ln *Cluster Spec*$_{Employ}$			**.101 (.009)**	**.069 (.010)**
ln *Cluster Spec*$_{Patent}$	**.102 (.014)**	**.052 (.015)**	**.070 (.014)**	***.031 (.015)***
ln *Related Clusters Spec*$_{Employ}$				**.058 (.017)**
ln *Related Clusters Spec*$_{Patent}$		**.138 (.025)**		**.101 (.026)**
ln *Cluster Spec in Neighbors*$_{Employ}$				***.030 (.014)***
ln *Cluster Spec in Neighbors*$_{Patent}$		**.162 (.021)**		**.118 (.022)**
EA fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes
R-Squared	.468	.471	.471	.474

Notes: See notes in Table 2. All models include EA-industry with positive employment in 1990.

Table 6
Emergence of new regional industries (probit; marginal effects).

	Emergence$_{2005}$ (N = 48,430)			
	6-1	6-2	6-3	6-4
ln *Cluster Spec*$_{Employ}$	**.033 (.002)**	**.032 (.002)**	**.026 (.002)**	**.025 (.002)**
ln *Cluster Spec*$_{Patent}$		**.009 (.003)**		***.006 (.003)***
ln *Related Clusters Spec*$_{Employ}$			**.015 (.004)**	**.015 (.004)**
ln *Related Clusters Spec*$_{Patent}$				−.005 (.006)
ln *Cluster Spec in Neighbors*$_{Employ}$			**.021 (.003)**	**.018 (.003)**
ln *Cluster Spec in Neighbors*$_{Patent}$				**.015 (.006)**
EA fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes
(Pseudo) R-Squared	.304	.306	.306	.306

Notes: Bold and Bold-Italic numbers refer to coefficients significant at 1% and 5% levels, respectively. The reported coefficients are the marginal effects of the estimated probit model. Standard errors are clustered by EA-Cluster. Model 1 includes a dummy equal to 1 if zero employment in 1990 in the cluster (outside the industry); similarly all models include dummies equal to 1 if the explanatory variable has been corrected because of zero employment or zero patents. Our findings only change trivially when dropping these dummies.

that multiple types of externalities arise across related industries within clusters.

7. Clusters and emergence of new regional industries

Our analysis shows a positive relationship between cluster strength and the growth of existing regional industries. We next examine whether the externalities that arise in clusters also relate to the *emergence* of industries new to the region. Building on studies of industry dynamics (Swann, 1998; Porter, 1998; Klepper, 2007, 2010; Neffke et al., 2011; Frenken et al., 2011), we use the cluster framework to explore the role of the initial cluster strength in the subsequent emergence of related industries in that cluster. Rather than focusing on particular types of linkages between new and existing industries, our measures of initial strength in a cluster capture various types of linkages that could be relevant for the emergence of industries.

The results are presented in Table 6. A higher level of cluster specialization measured by employment (model 6-1) is positively associated with an increase in the probability of the emergence of new regional industries in the cluster as of 2005 (the estimated marginal effect is 0.033).[34] Model 6-2 includes the regional cluster strength based both in employment and patenting. Both variables are positively associated with the emergence of new regional industries, but the estimated marginal effect of employment-based cluster strength is significantly higher (0.032 versus 0.009). This suggests that multiple types of externalities within clusters contribute to the emergence of new industries. Importantly, we find that this positive cluster effect is not restricted to manufacturing industries. Cluster strength facilitates the entry of new service industries as well.

We then specify in (6-3) a comprehensive model of the cluster environment using the employment-based variables (*Cluster Spec*$_{Employ}$, *Related Cluster Spec*$_{Employ}$, and *Cluster Spec in Neighbors*$_{Employ}$), and find that each is associated with a significantly higher probability of the emergence of new regional industries.[35] The estimated probability that a new industry emerges in a region is 0.22 at the mean of the explanatory variables, and increases to 0.32 (45% increase) if the cluster environment variables are valued at the mean plus one standard deviation. The findings are robust in model 6-4 to including both the employment and patenting cluster environment variables. Most prior studies have neglected the existence of inter-regional spillovers that influence the entry of an industry to a particular region, but we find that the strength of neighboring clusters contributes to the emergence of industries new to the region. These findings call for further research on the flow of knowledge, labor, and other inputs across nearby regions.

Our analysis suggests that the initial cluster strength (based on employment and patenting) facilitates the emergence of industries in a regional cluster. This, in turn, could be based on different types of firms: startups or incumbent firms entering through existing or new facilities (Dunne et al., 1988; Feldman et al., 2005; Klepper, 2007; Frenken and Boschma, 2007; Delgado et al., 2010; Alcacer and Delgado, 2013). Analysis of the types of firms that drive the emergence and growth of regional industries within clusters is an important area for future research.

[34] This finding is robust to using alternative industry groupings to define cluster strength: two-digit SIC and the BCD.

[35] This result holds with the exclusion of region and industry fixed effects.

8. Conclusions

In this paper, we investigate the role of cluster agglomeration in the growth of individual industries located within a cluster. We find the co-existence of convergence at the region-industry with agglomeration economies across related industries within clusters. The rate of employment growth in a region-industry is declining in the initial level of employment at the region-industry. At the same time, industries located in a cluster with higher initial employment strength in related industries are associated with higher employment growth, a finding that is robust across different regions, clusters, and industry types. Industry employment growth is also increasing with the strength of related clusters in the region and with the strength of similar clusters in geographically adjacent regions.

We also find complementarities between the employment and innovation performance in clusters. The initial employment and patenting strength of a regional cluster are each positively associated with region-industry growth in employment and patenting. These findings inform the debate on the importance of co-locating production and innovation for regional performance (Dertouzos et al., 1989; Porter and Rivkin, 2012; Berger, 2013). Furthermore, this analysis offers suggestive evidence that a broad set of externalities – such as knowledge spillovers, access to specialized inputs, and sophisticated demand – arise in clusters.

We also use the cluster framework to examine the role of a strong cluster in the emergence of new industries within the cluster. Our analysis shows that new regional industries in services and manufacturing are born out of strong regional clusters. In addition, the strength of the cluster in nearby regions seems to matter for the emergence of regional industries. These findings suggest that cluster agglomeration plays a role in the path of regional diversification (Porter, 1990, 1998; Swann, 1992, 1998; Neffke et al., 2011).

Taken together, these findings have a number of implications. First, the traditional distinction between industry specialization and regional diversity is misplaced. This dichotomy overlooks the role played by complementary economic activity in shaping economic growth, and the role of clusters as the manifestation of complementarity. Narrow regional specialization in an industry is likely to result in diminishing returns. However, the presence of complementary activity via clusters relates positively to growth. Clusters may encourage various agglomeration mechanisms, including a firm's ready access to key inputs, better interactions with customers, and better experimentation and innovation opportunities.

Second, prior studies have focused on individual channels through which complementarities might operate. Building on Feldman and Audretsch (1999), for example, numerous studies have emphasized the role of the local scientific knowledge base and the potential for knowledge spillovers in shaping opportunities for innovation and entrepreneurship. While our results are consistent with such findings, the impact of related economic activity on economic performance is far broader. The presence of clusters, which foster multiple types of complementarities, seems to be a key driver of the emergence and growth of industries for different industry types.

Our findings also carry several policy implications. First, regional policies that encourage complementarities across related economic activity are likely to be more effective than those that prioritize particular industries or clusters where there is little pre-existing strength within the region. Hence, policymakers should pursue policies that leverage a region's cluster strengths (Porter, 1990, 1998, 2003; Cortright, 2006; Rodríguez-Clare, 2007; Ketels and Memedovic, 2008). Moreover the notion of cluster strength operationalizes the view that policy action should focus on building upon pre-existing comparative advantage (Hausmann and Rodrik, 2003).[36]

Second, the benefits arising from clusters often span multiple jurisdictions (and even neighboring states). Our results suggest that the growth and emergence of regional industries relate to the cluster composition in nearby regions. Policies that enhance complementarities across jurisdictions, such as supporting infrastructure and institutions that facilitate access to demand, skills, or suppliers in neighboring clusters, may be important tools for regional development. More research is needed to assess the role of specific state and federal-level policies in catalyzing cluster growth.

Third, regional performance depends on the co-location of employment and innovation in a region. There are positive complementarities between the employment and innovation activities across industries within a regional cluster that contribute to the growth of regional industries. This calls for policies that foster complementarities between production and R&D within clusters – for example, the creation of public or semi-public goods that are cluster-specific, including institutions for collaboration, training programs, and policies to facilitate the collaboration between producers and local R&D institutions (Porter and Emmons, 2003; Porter and Rivkin, 2012; Berger, 2013).

Our analysis also raises several directions for future research. While the current analysis takes the initial cluster strength as a given, and holds cluster definitions constant, cluster structure can evolve over time. For example, whereas electronic computers may have been the central industry within the information technology cluster on an historical basis, software may be the core industry within that cluster going forward. However, few studies have examined how the co-location patterns of industries change over time, or how the historical composition of industries in a region shape the emergence of new industries (Klepper, 2007, 2010). Understanding the drivers of the evolution of clusters, particularly the types of constituent firms and their networks, is a crucial direction for future research.

Finally, qualitative studies of clusters emphasize the central role of specialized local institutions – from training facilities to infrastructure investments – in allowing potential complementarities to be realized (Porter and Emmons, 2003; Sölvell et al., 2006). While the precise design, role, and operation of such institutions vary widely by circumstance, little theoretical or empirical research has examined the impact of these localized institutions on regional economic performance.

Acknowledgements

We are grateful to the editor and reviewers for great comments. This paper has benefited greatly from the sustained research by Richard Bryden and Weifeng Weng in assembling the US Cluster Mapping Project dataset. Ben Jones, Jim Davis, Juan Alcacer, Bill Kerr, Matthew Kahn, Jim Dana, Wayne Gray, Megan MacGarvie, Anita McGahan, Alex Bryson, Iain Cockburn, Ram Mudambi, Frank Neffke, Christian Ketels, and Sam Zyontz contributed helpful suggestions. We are particularly grateful for many insightful comments by participants in the NBER Productivity Lunch Series, the Research Seminar at IESE Business School, the NBER Summer Institute (IPE group), and the North American Regional Science Association International meetings. Mercedes Delgado gratefully acknowledges support from the Kauffman Foundation and the Innovation Policy group at the NBER. Any opinions and conclusions

[36] For example, Hausmann and Rodrik (2003) argue that policy action should focus on innovative projects that can generate new knowledge about the relatedness among existing and potential new products. We think that these types of knowledge externalities may be facilitated in clusters of related industries.

expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.respol.2014.05.007.

References

Alcacer, J., 2006. Location choices across the value chain: how activity and capability influence collocation. Management Science 52 (10), 1457–1471.

Alcacer, J., Delgado, M., 2013. Spatial Organization of Firms and Location Choices through the Value Chain, Harvard Business School Working Paper No. 13-025.

Aharonson, B., Baum, J., Plunket, A., 2008. Inventive and uninventive clusters: the case of Canadian biotechnology. Research Policy 37, 1108–1131.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dorddrecht.

Audretsch, D.B., 1998. Agglomeration and the location of innovative activity. Oxford Review of Economic Policy 14 (2), 18–29.

Audretsch, D.B., Feldman, M.P., 1996. Knowledge spillovers and the geography of innovation and production. American Economic Review 86 (3), 630–640.

Baptista, R., Preto, M.T., 2010. Long term effects of new firm formation by type of start-up. International Journal of Entrepreneurship and Small Business 11 (4), 382–402.

Baptista, R., Swann, P., 1998. Do firms in clusters innovate more? Research Policy 27, 525–540.

Baptista, R., Swann, P., 1999. A comparison of clustering dynamics in the US and UK computer industries. Journal of Evolutionary Economics 9, 373–399.

Barro, R.J., Sala-i-Martin, X., 1991. Convergence across states and regions. Brookings Papers on Economic Activity 1, 107–182.

Beaudry, C., Schiffauerova, A., 2009. Who's right, Marshall or Jacobs? The localization versus urbanization debate. Research Policy 38 (2), 318–337.

Berger, S. (Ed.), 2013. Making in America: From Innovation to Market. MIT Press, Cambridge, MA.

Bloom, N., Schankerman, M., Van Reenen, J., 2012. Identifying technology spillovers and product market rivalry. Econometrica 81 (4), 1347–1393.

Bönte, W., 2004. Innovation and employment growth in industrial clusters: evidence from aeronautical firms in Germany. International Journal of the Economics of Business 11 (3), 259–278.

Boschma, R., 2005. Proximity and innovation: a critical assessment. Regional Studies 39, 61–74.

Bresnahan, T., Gambardella, A. (Eds.), 2004. Building High-Tech Clusters. Silicon Valley and Beyond. Cambridge University Press, New York.

Buenstorf, G., Klepper, S., 2009. Heritage and agglomeration: the akron tire cluster revisited. The Economic Journal 119, 705–733.

Ciccone, A., Hall, R., 1996. Productivity and the density of economic activity. American Economic Review 86 (1), 54–70.

Cingano, F., Schivardi, F., 2004. Identifying the sources of local productivity growth. Journal of the European Economic Association 2 (4), 720–742.

Combes, P.P., 2000a. Economic structure and local growth: France, 1984–1993. Journal of Urban Economics 47, 329–355.

Combes, P.P., 2000b. Marshall-Arrow-Romer Externalities and City Growth, CERAS Working Paper No. 99-06.

Cortright, J., 2006. Making Sense of Clusters: Regional Competitiveness and Economic Development. The Brookings Institution Metropolitan Policy Program http://www.brookings.edu/reports/2006/03cities_cortright.aspx

Delgado, M., Porter, M.E., Stern, S., 2010. Clusters and entrepreneurship. Journal of Economic Geography 10 (4), 495–518.

Delgado, M., Porter, M.E., Stern, S., Harvard Business School Working Paper 2013. Defining Clusters of Related Industries. Mimeo, Boston.

Dertouzos, M.L., Lester, R.K., Solow, R.M., MIT Productivity Commission, 1989. Made in America: Regaining the Productive Edge. MIT Press, Cambridge, MA.

Dumais, G., Ellison, G., Glaeser, E.L., 2002. Geographic concentration as a dynamic process. Review of Economics and Statistics 84 (2), 193–204.

Dunne, T., Roberts, M., Samuelson, L., 1988. Patterns of firm entry and exit in U.S. manufacturing industries. RAND Journal of Economics 19, 495–515.

Duranton, G., 2007. Urban evolutions: the fast, the slow, and the still. American Economic Review 97 (1), 197–221.

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

Duranton, G., Puga, D., 2001. Nursery cities: urban diversity, process innovation, and the life cycle of products. American Economic Review 91 (5), 1454–1477.

Ellison, G., Glaeser, E., 1997. Geographic concentration in U.S. manufacturing industries: a dartboard approach. Journal of Political Economy 105, 889–927.

Ellison, G., Glaeser, E., Kerr, W., 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. The American Economic Review 100, 1195–1213.

Everitt, B.S., Landau, S., Leese, M., Stahl, D., 2011. Cluster Analysis, 5th ed. John Wiley & Sons, Ltd., Chichester, UK.

Feldman, M.P., Audretsch, D., 1999. Innovation in cities: science-based diversity, specialization, and localized competition. European Economic Review 43, 409–429.

Feldman, M.P., Francis, J., Bercovitz, J., 2005. Creating a cluster while building a firm: entrepreneurs and the formation of industrial clusters. Regional Studies 39 (1), 129–141.

Feser, E.J., 2005. Benchmark Value Chain Industry Clusters for Applied Regional Research, Regional Economics Applications Laboratory. University of Illinois, Urbana-Champaign.

Feser, E.J., Bergman, E.M., 2000. National industry cluster templates: a framework for applied regional cluster analysis. Regional Studies 34 (1), 1–19.

Feser, E., Renski, H., Goldstein, H., 2008. Clusters and economic development outcomes. Economic Development Quarterly 22 (4), 324–344.

Forni, M., Paba, S., 2002. Spillovers and the growth of local industries. Journal of Industrial Economics 50, 151–171.

Frenken, K., Boschma, R., 2007. A theoretical framework for evolutionary economic geography: industrial dynamics and urban growth as a branching process. Journal of Economic Geography 7 (5), 635–649.

Frenken, K., Cefis, E., Stam, E., 2011. Industrial Dynamics and Economic Geography: A Survey, Working Paper No. 11. 07. Eindhoven Centre for Innovation Studies (ECIS), Eindhoven University of Technology.

Frenken, K., van Oort, F.G., Verburg, T., 2007. Related variety, unrelated variety, and regional economic growth. Regional Studies 41 (5), 685–697.

Fuchs, E., Kirchain, R., 2010. Design for location? The impact of manufacturing offshore on technology competitiveness in the optoelectronics industry. Management Science 56 (12), 2323–2349.

Fujita, M., Krugman, P., Venables, A.J., 1999. The Spatial Economy. MIT Press, Cambridge, MA.

Fujita, M., Thisse, J.F., 2002. Economics of Agglomeration: Cities, Industrial Location, and Regional Growth. Cambridge University Press, Cambridge, UK.

Glaeser, E.L., Kallal, H.D., Scheinkman, J.A., Shleifer, A., 1992. Growth in cities. Journal of Political Economy 100 (6), 1126–1152.

Glaeser, E.L., Kerr, W.R., 2009. Local industrial conditions and entrepreneurship: how much of the spatial distribution can we explain? Journal of Economics and Management Strategy 18 (3), 623–663.

Gompers, P., Lerner, J., Scharfstein, D., 2005. Entrepreneurial spawning: public corporations and the formation of new ventures, 1986–1999. Journal of Finance 60 (2), 577–614.

Haltiwanger, J., Jarmin, R.S., Miranda, J., 2009. "Who Creates Jobs? Small vs. Large vs. Young," NBER Working Paper No. 16300.

Hausmann, R., Rodrik, D., 2003. Economic development as self-discovery. Journal of Development Economics 72, 603–633.

Helper, S., Krueger, T., Wial, H., 2012. Locating American Manufacturing: Trends in the Geography of Production. Brookings Institution, Washington, DC.

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

Henderson, J.V., 2003. Marshall's scale economies. Journal of Urban Economics 53, 1–28.

Henderson, J.V., Kuncoro, A., Turner, M., 1995. Industrial development in cities. Journal of Political Economy 103, 1067–1090.

Higgins, M.J., Levy, D., Young, A.T., 2006. Growth and convergence across the United States: evidence from county-level data. Review of Economics and Statistics 88 (4), 671–681.

Jaffe, A., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. Quarterly Journal of Economics 108, 577–598.

Johnson, K.P., Kort, J.R., 2004. Redefinition of the BEA Economic Areas. http://www.bea.gov/scb/pdf/2004/11November/1104Econ-Areas.pdf

Ketels, C., Memedovic, O., 2008. From clusters to cluster-based economic development. International Journal of Technological Learning, Innovation, and Development 1, 375–392.

Klepper, S., 2007. Disagreements, spinoffs, and the evolution of detroit as the capital of the U.S. automobile industry. Management Science 53, 616–631.

Klepper, S., 2010. The origin and growth of industry clusters: the making of Silicon Valley and Detroit. Journal of Urban Economics 67 (1), 15–32.

Krugman, P., 1991. Increasing returns and economic geography. Journal of Political Economy 99 (3), 483–499.

LaFountain, C., 2005. Where do firms locate? Testing competing models of agglomeration. Journal of Urban Economics 58 (2), 338–366.

LeSage, J.P., 1999. The theory and practice of spatial econometrics. Econometrics Toolbox Functions for Matlab http://www.spatial-econometrics.com

Lorenzen, M., Mudambi, R., 2013. Clusters, connectivity, and catch-up: bollywood and Bangalore in the global economy. Journal of Economic Geography 13, 501–534.

Markusen, A., 1996. Sticky places in slippery space: a typology of industrial districts. Economic Geography 72 (3), 293–313.

Marshall, A., 1920. Principles of Economics. Macmillan, London.

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.

Pisano, G.P., 1997. The Development Factory: Unlocking the Potential of Process Innovation. Harvard Business School Press, Boston.

Porter, M.E., 1990. The Competitive Advantage of Nations. Free Press, New York.

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

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

Porter, M.E., Emmons, W., 2003. Institutions for Collaboration: Overview, Harvard Business School Note 703-436.

Porter, M.E., Rivkin, J.W., 2012. Choosing the United States. Harvard Business Review 90 (3), 80–91.

Rodríguez-Clare, A., 2007. Clusters and comparative advantage: implications for industrial policy. Journal of Development Economics 82, 43–57.

Rosenthal, S.S., Strange, W.C., 2009. Small establishments/big effects: agglomeration, industrial organization, and entrepreneurship. In: Glaeser, E. (Ed.), The Economics of Agglomeration. National Bureau of Economic Research, University of Chicago Press, Chicago, pp. 277–302.

Rosenthal, S.S., Strange, W.C., 2004. Evidence on the nature and sources of agglomeration economies. In: Henderson, J.V., Thisse, J.F. (Eds.), Handbook of Regional and Urban Economics, vol. 4. Elsevier North-Holland, Amsterdam.

Saxenian, A., 1994. Regional Advantage: Culture and Competition in Silicon Valley and Route 128. Harvard University Press, Cambridge, MA.

Scherer, F.M., 1982. Inter-industry technology flows and productivity growth. The Review of Economics and Statistics 64 (4), 627–634.

Silverman, B.S., 1999. Technological resources and the direction of corporate diversification. Management Science 45 (8), 1109–1124.

Sölvell, Ö., Lindqvist, G., Ketels, C., 2006. Cluster Initiatives in Developing and Transition Economies, Report, Ivory Tower AB, Stockholm. http://www.cluster-research.org/dldevtra.htm

Sorenson, O., Audia, P.G., 2000. The social structure of entrepreneurial activity: geographic concentration of footwear production in the United States, 1940–1989. American Journal of Sociology 106 (2), 424–462.

Swann, P., 1992. The Dynamics of Industrial Clusters. Oxford University Press, Oxford.

Swann, P., 1998. Clusters in the US computing industry. In: Swann, P., Prevezer, M., Stout, D. (Eds.), The Dynamics of Industrial Clustering: International Comparisons in Computing and Biotechnology. Oxford University Press, Oxford, pp. 76–105.

Venables, A.J., 1996. Equilibrium locations of vertically linked industries. International Economic Review 37 (2), 341–360.

Vernon, R., 1966. International investment and international trade in the product cycle. Quarterly Journal of Economics 80 (2), 190–207.