REGIONAL STUDIES
2021, VOL. 55, NO. 3, 479–494
https://doi.org/10.1080/00343404.2020.1802418

Routledge Taylor & Francis Group
RSA Regional Studies Association

OPEN ACCESS Check for updates

Relatedness, complexity and local growth

Benjamin Davies[a] and David C. Maré[b]

ABSTRACT
We derive a measure of the relatedness between economic activities based on weighted correlations of local employment shares. Our approach recognizes variation in the extent of local specialization and adjusts for differences in data quality between cities. We use our measure to estimate activity and city complexity, and examine the contribution of relatedness and complexity to urban employment growth in New Zealand. Relatedness and complexity are complementary in promoting employment growth in New Zealand's largest cities, but do not contribute to employment growth in its smaller cities.

KEYWORDS
relatedness; complexity; smart specialization

JEL R11, R12
HISTORY Received 29 April 2019; in revised form 8 July 2020

INTRODUCTION

The spatial concentration of economic activities in cities generates agglomeration economies arising from labour market pooling, input sharing and knowledge spillovers (Marshall, 1920). The principle of relatedness (Hidalgo et al., 2018; Vicente et al., 2018) suggests that the advantages of proximity may also accrue to interacting activities that are similar in ways other than spatially. Such interactions support the growth of complex activities that rely on specialized combinations of complementary knowledge and skills (Balland et al., 2020; Hidalgo & Hausmann, 2009).

In this paper we derive a measure of the relatedness between economic activities based on weighted correlations of local employment shares. Our approach extends measures used in previous studies (Balland et al., 2019; Boschma et al., 2015; Farinha Fernandes et al., 2019; Hidalgo et al., 2007; Rigby et al., 2019) by recognizing the extent of local specialization and adjusting for differences in employment data quality between geographical areas. These attributes make our measure suitable for studying small areas, where measurement errors and random fluctuations in employment are proportionally large. We use our measure to estimate activity and city complexity based on an eigenvector approximation of the Method of Reflections (Caldarelli et al., 2012; Hidalgo & Hausmann, 2009).

Balland et al. (2019, p. 1252) propose that relatedness and complexity capture 'the risks and rewards of competing diversification strategies', and that the ideal local growth strategy involves expanding into complex activities related to existing local competencies. Rigby et al. (2019) test this proposition empirically by analysing whether European city-regions whose historical growth paths aligned more closely with Balland et al.'s ideal experienced faster employment and gross domestic product (GDP) growth. However, Balland et al.'s and Rigby et al.'s analyses both use data on European city-regions, and leave open the question of whether relatedness and complexity provide information about growth prospects in smaller, non-European contexts.

We address this question by examining the contribution of relatedness and complexity to urban employment growth in New Zealand, a small and geographically isolated country with limited access to agglomeration economies (McCann, 2009). Our data cover a range of urban areas that are smaller than, but contain similar activities to, previously studied regions. These data allow us to investigate whether the mechanisms through which relatedness and complexity promote employment growth operate only in sufficiently large cities. Our results suggest that

CONTACT
[a] **(Corresponding author)** ben.davies@motu.org.nz
Motu Economic and Public Policy Research, Wellington, New Zealand.
[b] dave.mare@motu.org.nz
Motu Economic and Public Policy Research, Wellington, New Zealand.
Supplemental data for this article can be accessed at https://doi.org/10.1080/00343404.2020.1802418

relatedness and complexity are complementary in promoting employment growth in New Zealand's largest cities but do not contribute to employment growth in its smaller cities. Our analysis supports the characterization of cities as networks of interacting activities; in our data, the benefits of such interaction are more apparent in larger cities, where workers engaged in related activities interact more frequently.

The main contributions of this paper are twofold. First, we provide relatedness and complexity measures that overcome some of the drawbacks of discrete measures used in previous studies. Second, we provide new empirical evidence on the (lack of) relevance of these measures for analysing growth dynamics in small geographical areas. We also demonstrate how the apparent relevance of relatedness and complexity depends on the measures used. Ultimately, we find that the conclusions drawn using previous measures are not robust to using our measures, which we believe are more appropriate in our context.

The remainder of this paper is structured as follows. Next, we discuss the principle of relatedness, the concept of complexity, and the potential relevance of relatedness and complexity to employment growth. We then explain our approach to measuring relatedness and complexity, and compare our measures with those used in previous studies. We describe our data and document our empirical findings. Finally, we provide a brief conclusion and discuss some remaining research questions.

LITERATURE REVIEW AND RELEVANT MECHANISMS

Activities are related if they require similar knowledge or inputs (Hidalgo et al., 2018). Such similarities can be inferred from, for example, worker flows (Jara-Figueroa et al., 2018; Neffke & Henning, 2013), input–output linkages and shared labour pools (Delgado et al., 2016), and patent applications (Balland et al., 2019; Boschma et al., 2015).

The principle of relatedness describes the empirical relationship between the probability that a location specializes in a new activity and the presence of related activities in that location (Hidalgo et al., 2018). This relationship motivates studies of related and unrelated variety (Content & Frenken, 2016; Frenken et al., 2007; Fritsch & Kublina, 2018), and related diversification (Boschma, 2017; Neffke et al., 2011; Rigby, 2015). Such studies form a subset of the literature on urban and regional growth and innovation. A dominant focus within this literature is on the relevance of relatedness for processes of innovation (Boschma, 2005; Feldman & Audretsch, 1999), entrepreneurship (Neffke et al., 2018) and industrial diversification (Neffke & Henning, 2013). This focus reflects the microfoundations of the relatedness literature, which emphasize local complementarities and the consequent knowledge creation occurring between related knowledge bases (Asheim & Gertler, 2005).

The principle of relatedness influences European regional policy by underpinning spatially differentiated policy approaches. Local policies are context specific, in light of the relatedness patterns among local economic activities as well as the local institutional context (Barca et al., 2012; Boschma, 2014). The policy emphasis, as with the relatedness literature, is on innovation, entrepreneurial, and research and development processes, and the support of innovation-led growth. Such processes are the focus of smart specialization policies (Foray et al., 2009, 2011), which encourage regions to upgrade their economic structure 'by building on their existing capabilities' (Balland et al., 2019, p. 1253) and which are a core component of the reformed Cohesion Policy (Barca, 2009; McCann & Ortega-Argilés, 2015).

Balland et al. (2019) propose a framework for analysing smart specialization that connects the principle of relatedness to the concept of complexity (Hidalgo & Hausmann, 2009). The complexity literature characterizes economic activities as embodiments of tacit knowledge and skills (Hidalgo et al., 2007; Hidalgo & Hausmann, 2009), and emphasizes the role of social and economic networks in knowledge and skill accumulation (Hidalgo, 2015; Sorenson, 2005; Sorenson et al., 2006). Such networks facilitate sharing and learning among individuals and firms, supporting the growth of complex activities that rely on specialized combinations of complementary knowledge and skills.

The literature on evolutionary economic geography (Boschma & Frenken, 2006) similarly emphasizes the tacit knowledge embedded in firms, and the ability of firms to absorb and recombine such knowledge (Hidalgo, 2015). This emphasis draws upon the characterization of growth and innovation as recombinant processes (Schumpeter, 1934; Weitzman, 1998). Because tacit knowledge, by definition, is difficult to transfer, its absorption and recombination may rely on indirect transfer mechanisms such as the migration of workers among firms (Breschi & Lissoni, 2005; Jara-Figueroa et al., 2018). These mechanisms constitute local interactions that facilitate collective learning and the production of complex knowledge. The localization of these interactions contributes to geographical variation in the concentration of complex knowledge (Balland & Rigby, 2017) and activities that rely on such knowledge. In particular, more complex activities tend to concentrate in larger cities (Balland et al., 2020) where collaborative networks are larger and denser.

Complex activities are economically valuable because they generate high rents (Rigby et al., 2019) and embody tacit knowledge, which provides a source of competitive advantage (Maskell & Malmberg, 1999). This argument, coupled with the principle of relatedness, underlies Balland et al.'s (2019, p. 1252) framework for analysing smart specialization. They characterize smart specialization as a way to 'leverage existing strengths' and 'generate novel platforms on which regions can build competitive advantage in high value-added activities'. Balland et al. operationalize these objectives by suggesting that regions expand into complex activities that are related to existing local specializations. Complex activities deliver economic rents and competitive advantage, while activities related to existing specializations can leverage the knowledge and skills possessed by local employees and firms.

Balland et al.'s framework highlights the complementarity between relatedness and complexity in the context of urban and regional development. Expanding into activities related to, but not of greater complexity than, existing specializations may lead to 'lock-in' that prevents growth and innovation because local capabilities do not expand. Likewise, attempts to expand into complex activities that are not related to existing specializations may fail because local workers and firms lack the requisite knowledge and skills.

Balland et al. use their framework to analyse technological growth in European regions. They capture technologies by patent classes and measure growth in patent claims. Such claims reflect the creation of new knowledge. They estimate the relatedness and complexity of different patent classes and find that these indices correlate with growth within each class. Rigby et al. (2019) use a similar empirical set-up to show that relatedness and complexity correlate positively with employment and GDP growth.

There are at least two reasons why relatedness and complexity may contribute to employment growth. First, clusters of related activities promote innovation (Delgado et al., 2014) by bringing together complementary ideas (Feldman & Audretsch, 1999; Jacobs, 1969). To the extent that such innovation produces long-term economic growth, competitive forces drive employment growth in local clusters of related activities in order to capitalize on their potential to facilitate knowledge creation and spillovers (Asheim & Gertler, 2005). Similarly, to the extent that complex activities deliver economic rents and competitive advantage, markets reallocate employees towards complex activities in order to extract these benefits.

Second, cities dense with related and complex activities are more robust to labour market shocks. In such cities, if demand for employees in one activity falls then displaced workers can quickly regain employment in another activity requiring similar knowledge and skills (Morkutė et al., 2017; Neffke & Henning, 2013). Moreover, because complex activities require *combinations* of knowledge and skills, workers engaged in such activities can regain employment in the alternative activities that use some of this knowledge or these skills.

MEASURING RELATEDNESS AND COMPLEXITY

Activity relatedness

We infer activities' relatedness from employee co-location patterns. Such patterns reveal activities' mutual reliance on spatially heterogeneous inputs, such as the tacit knowledge and skills embedded in firms. Our relatedness measure relies on variation in activities' relative sizes in different cities. This variation may arise due to firm birth or death, or differential growth rates. We do not attempt to separate these sources of variation.

The mathematical foundations of our relatedness measure are as follows. Consider an economy comprising a set C of cities and a set $\mathcal{A}$ of activities. Let E_c^a denote the number of employees in city $c \in \mathcal{C}$ and activity $a \in \mathcal{A}$. Total city c employment is given by

$$E_c = \sum_{a \in \mathcal{A}} E_c^a,$$

while national activity a employment is equal to

$$E^a = \sum_{c \in \mathcal{C}} E_c^a.$$

Summing over all cities and activities yields national employment:

$$E = \sum_{c \in \mathcal{C}} \sum_{a \in \mathcal{A}} E_c^a.$$

Comparing the local share

$$LS_c^a = \frac{E_c^a}{E_c}$$

of activity a in city c with its share E^a/E of national employment reveals whether the activity is relatively over-represented in city c. Such over-representation indicates local specialization in activity a relative to the national economy.

We measure the relatedness of activities a_i and a_j using the correlation between the corresponding vectors $(LS_1^{a_i}, LS_2^{a_i}, \ldots, LS_{|\mathcal{C}|}^{a_i})$ and $(LS_1^{a_j}, LS_2^{a_j}, \ldots, LS_{|\mathcal{C}|}^{a_j})$ of local employment shares. This correlation is high when a_i and a_j are relatively over-represented in similar cities, revealing firms' tendency to co-locate in pursuit of agglomeration economies. First, we compute the weighted covariance

$$\Omega_{a_i a_j} = \sum_{c \in \mathcal{C}} \frac{E_c}{E} \left(LS_c^{a_i} - \sum_{c \in \mathcal{C}} \frac{E_c}{E} LS_c^{a_i} \right) \left(LS_c^{a_j} - \sum_{c \in \mathcal{C}} \frac{E_c}{E} LS_c^{a_j} \right)$$

$$= \sum_{c \in \mathcal{C}} \frac{E_c}{E} \left(\frac{E_c^{a_i}}{E_c} - \frac{E^{a_i}}{E} \right) \left(\frac{E_c^{a_j}}{E_c} - \frac{E^{a_j}}{E} \right) \quad (1)$$

between the local share vectors for activities a_i and a_j, where the weighting factor E_c/E is equal to city c's share of national employment. Second, we divide $\Omega_{a_i a_j}$ by the city share-weighted standard deviations of $(LS_1^{a_i}, LS_2^{a_i}, \ldots, LS_{|\mathcal{C}|}^{a_i})$ and $(LS_1^{a_j}, LS_2^{a_j}, \ldots, LS_{|\mathcal{C}|}^{a_j})$, yielding the weighted correlation of local employment shares for activities a_i and a_j. Finally, we map this correlation to the closed unit interval $[0, 1]$ using the linear transformation $x \mapsto (x+1)/2$. Hence, our measure of the relatedness between activities a_i and a_j is given by

$$R_{a_i a_j} = \frac{1}{2} \left(\frac{\Omega_{a_i a_j}}{\sqrt{\Omega_{a_i a_i} \Omega_{a_j a_j}}} + 1 \right). \quad (2)$$

$R_{a_i a_j}$ has a range of [0, 1], is largest when activities a_i and a_j have equal local shares in each city $c \in \mathcal{C}$, and is smallest when the percentage point difference between activity a_i's local and national shares has equal magnitude but opposite sign to that difference for a_j in all cities. We assume that $\Omega_{aa} > 0$ for each $a \in \mathcal{A}$ so that (2) is well defined.

Comparison with previously used relatedness measures

The bracketed terms in the summand of (1) are equal to the percentage point difference between activities' local and national employment shares, and thus measure the extent to which activities are locally over-represented. An alternative measure of local over-representation is the location quotient

$$LQ_c^a = \frac{E_c^a/E_c}{E^a/E}, \tag{3}$$

which exceeds unity if and only if activity a comprises a larger share of city c employment than of national employment. Hidalgo et al. (2007), following Balassa (1965), use a similar metric to identify the commodities in which countries exhibit revealed comparative advantage (RCA) and infer commodities' similarity from RCA co-occurrences. Boschma et al. (2015) and Balland et al. (2019) use this approach to estimate the similarity between different technologies using patent data from US cities and European regions.

Inferring activity relatedness from RCA co-occurrence patterns is problematic for at least three reasons. First, near-zero denominators of (3), caused by activities contributing negligible shares of national employment, exacerbate measurement errors in the numerator of (3). Our measure (2) reduces the impact of such errors by comparing percentage point differences in local and national shares rather than ratios of such shares.

Second, RCA co-occurrence patterns ignore variation in the extent of local specialization and are sensitive to small perturbations in location quotients near unity. To see why, consider the indicator variable

$$\text{RCA}_c^a = \begin{cases} 1 & \text{if } LQ_c^a \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

for the event in which city c has RCA in activity a. This variable is constant with respect to LQ_c^a on the intervals $[0, 1)$ and $(1, \infty)$, and is discontinuous at $LQ_c^a = 1$ and, therefore, in E_c^a. In contrast, our measure (2) recognizes variation in the extent of local specialization and varies continuously with local activity employment.

Third, the RCA approach is sensitive to external influence and noise in employment counts within small cities. If an activity exits a small city, then all other activities in that city are likely to become over-represented relative to their national shares because the increase in their local shares will be proportionally larger than any change in national shares. Thus, the identification of activities in which small cities appear to be specialized is sensitive to internal migration and to fluctuations in local employment in other activities. Activity specializations in larger cities are less noisy because local shares are less sensitive to absolute fluctuations in local employment. The RCA approach does not recognize differences in signal quality between cities of different size. In contrast, our measure (2) is more robust to noise induced by small cities because it gives such cities less weight than large cities.

Mean local relatedness and relatedness density

Equation (2) implies that $R_{aa} = 1$ for each activity $a \in \mathcal{A}$. Therefore, the local share-weighted mean relatedness of activity a with the activities in city c can be written as

$$\sum_{a' \in \mathcal{A}} \frac{E_c^{a'}}{E_c} R_{aa'} = LS_c^a + RD_c^a,$$

where we define

$$RD_c^a = \sum_{a' \in \mathcal{A} \setminus \{a\}} \frac{E_c^{a'}}{E_c} R_{aa'}$$

as the 'relatedness density' of activity a in city c. Boschma et al. (2015) and Balland et al. (2019) suggest an alternative measure

$$\frac{\sum_{a' \in \mathcal{A} \setminus \{a\}} \text{RCA}_c^{a'} R_{aa'}}{\sum_{a' \in \mathcal{A} \setminus \{a\}} R_{aa'}}$$

of relatedness density. However, this measure has the same limitations as inferring relatedness from RCA co-occurrences: amplified measurement errors, disregard for the extent of local specialization, discontinuity at unit location quotients, and fragility in small cities.

Activity complexity

Complexity captures the knowledge intensity of activities (Balland et al., 2020) by encoding the extent to which they rely on specialized combinations of knowledge and skills. We define activity complexity using the second eigenvector of the row-standardized activity relatedness matrix. Our approach extends Caldarelli et al.'s (2012) eigenvector approximation of Hidalgo and Hausmann's (2009) Method of Reflections.

For ease of exposition, we first summarize the RCA-based approach to estimating complexity before introducing the relatedness-based approach used in our empirical analysis. Let

$$\phi_c^{(0)} = \sum_{a \in \mathcal{A}} \text{RCA}_c^a$$

denote the number of activities in which city $c \in \mathcal{C}$ has RCA and let

$$\psi_a^{(0)} = \sum_{c \in \mathcal{C}} \text{RCA}_c^a$$

denote the number of cities that have RCA in activity $a \in \mathcal{A}$. Consider the sequences $(\phi_c^{(k)})_{k \geq 0}$ and $(\psi_a^{(k)})_{k \geq 0}$ defined by the system:

$$\phi_c^{(k)} = \frac{1}{\phi_c^{(0)}} \sum_{a \in \mathcal{A}} \text{RCA}_c^a \psi_a^{(k-1)} \tag{4}$$

$$\psi_a^{(k)} = \frac{1}{\psi_a^{(0)}} \sum_{c \in \mathcal{C}} \text{RCA}_c^a \phi_c^{(k-1)}. \tag{5}$$

Hidalgo and Hausmann (2009) argue that the limit point of $(\psi_a^{(k)})_{k \geq 0}$ measures the complexity of activity a because it captures 'the complexity that emerges from the interactions between the increasing number of individual activities that conform an economy.' Defining the vector $\psi^{(k)} = (\psi_1^{(k)}, \psi_2^{(k)}, \ldots, \psi_{|\mathcal{A}|}^{(k)})$, substituting (4) into (5) and letting $k \to \infty$ yields

$$P\psi^{(\infty)} = \psi^{(\infty)},$$

where we define

$$\psi^{(\infty)} = \lim_{k\to\infty} \psi^{(k)}$$

and where $P = (p_{a_i a_j})$ is the matrix with

$$p_{a_i a_j} = \sum_{c\in C} \frac{\mathrm{RCA}_c^{a_i}}{\psi_{a_i}^{(0)}} \frac{\mathrm{RCA}_c^{a_j}}{\phi_c^{(0)}} \tag{6}$$

as the entry in row a_i and column a_j. P is then the transition matrix for a Markov chain on $\mathcal{A}$, which we interpret as follows. Suppose that a specialist in activity a_i relocates to another city specializing in a_i and that, on arrival, they change jobs to one of the local specializations in the new city. If all feasible outcomes of the relocation and job-change decision are equally likely, then $p_{a_i a_j}$ is the probability that the specialist shifts to activity a_j.

We estimate activity complexity using the spectral properties of P as follows. Consider the standardized vector

$$\hat{\psi}^{(k)} = \frac{\psi^{(k)} - \overline{\psi^{(k)}}1}{\mathrm{sd}(\psi^{(k)})}, \tag{7}$$

where $\overline{\psi^{(k)}}$ and $\mathrm{sd}(\psi^{(k)})$ denote the mean and standard deviation of the components of $\psi^{(k)}$, and where 1 denotes the $|\mathcal{A}| \times 1$ vector of ones. According to Hidalgo and Hausmann (2009), the vector of activity complexities is given by the limit

$$\hat{\psi}^{(\infty)} = \lim_{k\to\infty} \hat{\psi}^{(k)}.$$

Caldarelli et al. (2012) show (as also shown in Appendix A in the supplemental data online), that if P has eigenvectors $e_1, e_2, \ldots, e_n$ and corresponding eigenvalues of decreasing absolute value then (7) implies

$$\hat{\psi}^{(\infty)} = \frac{e_2 - \overline{e_2}1}{\mathrm{sd}(e_2)}. \tag{8}$$

The matrix P is derived from RCA co-occurrence patterns that can produce unreliable relatedness estimates for the reasons identified above. Consequently, the spectral properties of P are not robust to, for example, small values of E_c^a or perturbations in location quotients near unity. We overcome this weakness by replacing P with a row-standardized copy of the activity relatedness matrix $R = (R_{a_i a_j})$, where $R_{a_i a_j}$ is the relatedness between activities a_i and a_j defined in (2). In the resulting Markov chain on the activity set $\mathcal{A}$, transitions from node a_i to node a_j occur with probability

$$\frac{R_{a_i a_j}}{\sum_{a\in\mathcal{A}} R_{a_i a}}. \tag{9}$$

Using our relatedness measure, rather than RCA co-occurrences, to define the stochastic structure of the inter-activity Markov chain retains potentially important information about activities' spatial distribution. Our approach thus improves upon the Method of Reflections, which discards such information by discretising the extent of local specialization.

We define the complexity C^a of activity a as the a^{th} component of the standardized second eigenvector of the row-standardized relatedness matrix R. Thus, our procedure is consistent with the eigenvector approximation suggested by Caldarelli et al. (2012) except that we replace the transition probability $p_{a_i a_j}$ with (9). The resulting vector $(C^1, C^2, \ldots, C^{|\mathcal{A}|})$ partitions $\mathcal{A}$ into two subsets of similar size according to the sign of each component (Mealy et al., 2019; Shi & Malik, 2000). We set the sign of C^1 such that C^a is positively correlated with the weighted mean size

$$\sum_{c\in C} \frac{E_c^a}{E^a} E_c$$

of cities that contain activity a. This choice recognizes that large cities facilitate a deeper division of labour than do small areas (Jacobs, 1969) and that such division is needed for complex knowledge to develop (Balland et al., 2020; Hidalgo & Hausmann, 2009).

City complexity

We estimate city complexity symmetrically to activity complexity. For each pair $c_i, c_j \in \mathcal{C}$, we compute the activity size-weighted covariance:

$$\sum_{a\in\mathcal{A}} \frac{E^a}{E}\left(\frac{E_{c_i}^a}{E^a} - \sum_{a\in\mathcal{A}} \frac{E^a}{E}\frac{E_{c_i}^a}{E^a}\right)\left(\frac{E_{c_j}^a}{E^a} - \sum_{a\in\mathcal{A}} \frac{E^a}{E}\frac{E_{c_j}^a}{E^a}\right)$$
$$= \sum_{a\in\mathcal{A}} \frac{E^a}{E}\left(\frac{E_{c_i}^a}{E^a} - \frac{E_{c_i}}{E}\right)\left(\frac{E_{c_j}^a}{E^a} - \frac{E_{c_j}}{E}\right)$$

in city shares of activity employment, from which we derive the relatedness between city c_i and c_j by converting to a weighted correlation and linearly mapping the result to $[0, 1]$. Hence, our city relatedness index measures the extent to which cities have more similar local activity portfolios than would be expected if employees were assigned to activities randomly. We define the complexity C_c of city c as the c^{th} component of the standardized second eigenvector of the row-standardized city relatedness matrix, consistent with our definition of activity complexity. We choose the sign of C_1 such that C_c is positively correlated with the local share-weighted mean complexity:

$$\sum_{a\in\mathcal{A}} \frac{E_c^a}{E^a} C^a$$

of activities in city c. Thus, by construction, complex activities tend to concentrate in complex cities.

DATA

We apply our relatedness and complexity measures to historical New Zealand Census data aligned to current industry, occupation and urban area codes. These data provide usual resident employment counts in each census from 1981 to 2013.[1] We capture cities by 2013 urban area code and identify activities using industry–occupation pairs. We capture industries by a manual grouping of New Zealand Standard Industry Output Category codes[2] and occupations by one-digit 1999 New Zealand Standard Classification of Occupations code.

We restrict our analysis to persistently large urban areas and activities in order to mitigate the impact of two confidentiality requirements imposed by Statistics New Zealand, the agency that provides our data. First, the employment count in each cell – that is, each combination of urban area, industry, occupation and census year – is randomly rounded up or down to a multiple of three. Second, cells with unrounded counts below six are suppressed. We identify 50 urban areas with at least 1400 employed usual residents and 199 industry–occupation pairs with at least 800 employees in each census year between 1981 and 2013. We pool remaining pairs into a single residual activity that represents about 18% of national employment across the years in our data.

Some urban areas delineate zones within a city. For example, the urban area classification separates New Zealand's largest city, Auckland, into its north, west, south and central zones. We merge such zones in order to increase the consistency between our city classification and functional economic areas. Our classification delivers 41 distinct cities.

We use our relatedness measure (2) to estimate local shares, relatedness densities, and activity and city complexities for each census year. We structure our estimates as panel data in which observations correspond to city–activity pairs in a given census year. We exclude all observations associated with the residual activity, and all observations for census years 1986, 1996 and 2006. Therefore, our data comprise a panel of 41 cities and 199 non-residual activities in census years 1981, 1991, 2001 and 2013.[3]

The cities in our data have year-specific employed usual resident populations ranging from 1434 (Queenstown in 1981) to 573,150 (Auckland in 2013). These populations have mean 29,947 and median 6952. About 36% of the employees captured in our data were usual residents of Auckland at the date of the corresponding census.

The activities in our data span 61 industries and nine occupations.[4] The non-residual activities in our data have year-specific employee populations ranging from 648 (mining plant and machine operators and assemblers in 2001) to 74,565 (education professionals in 2013). These populations have mean 6139 and median 2798.

Defining activities as industry–occupation pairs allows for variation in activity relatedness and complexity within industries and occupations. This variation may capture important differences in workers' knowledge and skills. For example, primary school teachers may require different knowledge and skills than university professors, even though both groups are employed within the education sector. Such differences may lead to different co-location patterns with other activities, and, consequently, different relatedness and complexity estimates. Likewise, managing a consulting firm may require different knowledge and skills than managing a trawling company, even though both tasks share an occupation classification of 'manager'.

Our activity classification is coarser than the classifications used in previous studies.[5] This coarseness is necessary because the employment counts in our data are small. Using a finer classification would increase the knowledge and skill homogeneity among workers within each activity, but decrease the information density of our data. We believe that our choice of 199 persistently large activities balances the trade-off between classification detail and information density, and leaves sufficient variation in activities' prevalence across cities to obtain defensible relatedness and complexity estimates. Our activity classification also delivers a relatively dense city–activity employment matrix: every activity appears in at least 16 of the 41 cities in our data.[6]

EMPIRICAL ANALYSIS

Activity space

We first define an 'activity space' that captures the network structure of activities based on our relatedness estimates. Our construction echoes the 'product space' of commodities defined by Hidalgo et al. (2007). We describe activity space by a weighted network $N = (\mathcal{A}, \mathcal{E})$, where $\mathcal{A}$ is the set of 199 non-residual activities in our data and where each edge $\{a_i, a_j\} \in \mathcal{E}$ has weight equal to the pairwise relatedness $R_{a_i a_j}$ between activities a_i and a_j.

Figure 1 presents a network map of activity space based on 2013 Census employment data. Nodes represent activities (industry–occupation pairs) and have radii proportional to activities' sizes. Darker nodes represent more complex activities. We use Fruchterman and Reingold's (1991) algorithm to position nodes using the edge weights of N. This algorithm places more related activities closer together. In order to reveal the strongest inter-activity connections, we display the subnetwork N' of N induced by the 500 edges of largest weight. We omit from our visualization all components of N' containing three nodes or fewer. These restrictions yield a network with four components, 84 nodes and 476 edges. To guide our readers, we label clusters of nodes in our visualization of N' with indicative sector labels that describe the majority of activities in each cluster.

Our map of activity space contains a densely connected cluster of activities associated with the distributive services sector. These activities include high- and low-skill occupations in the wholesale trade industry, and medium-skill occupations in the real estate and road transport industries. Below this cluster is a group of high-skill occupations in the professional services sector. Such activities tend to concentrate in large cities and, consequently, share strong co-location patterns.

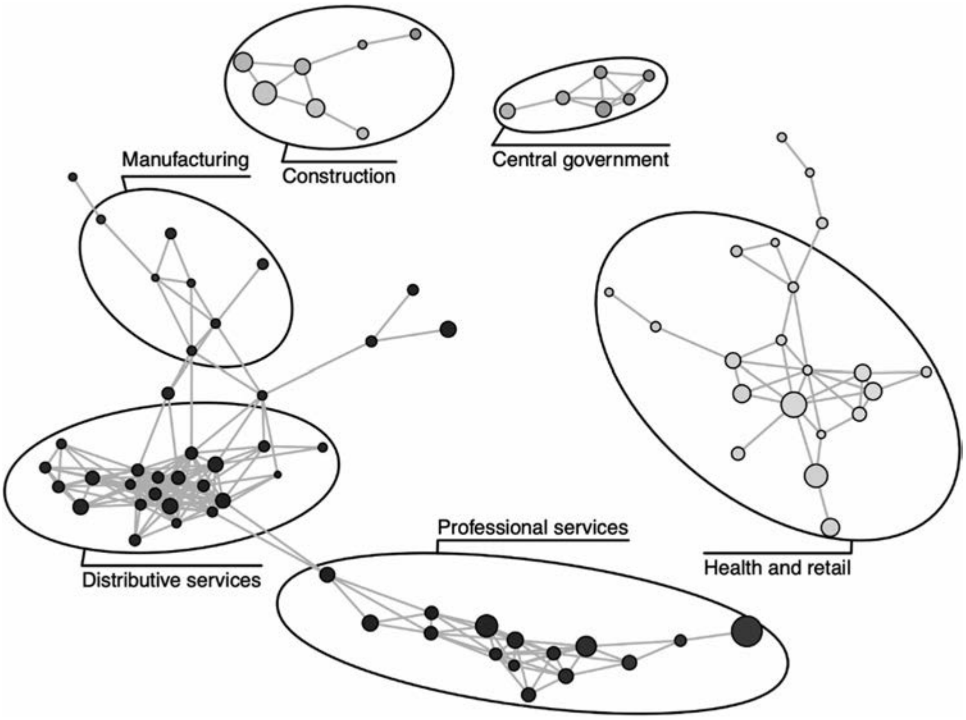

Figure 1. Network map of activity space based on relatedness estimates for 2013.
Note: Nodes represent activities (industry–occupation pairs). Larger, darker nodes represent larger, more complex activities. Labels describe the sector containing the majority of activities in each cluster (indicative only).

The distributive and professional services sectors contain the most complex activities based on our estimates using data from the 2013 Census. These sectors require specialized combinations of knowledge and skills that have low spatial transferability. In contrast, the least complex activities are associated with medium- to low-skill occupations in the health and retail industries. These activities have low complexity because they do not depend on other activities being present locally. Activities associated with the health and retail industries tend to be locally over-represented in cities with few specializations, which our measure (2) captures as high relatedness.

The smaller two components in our visualization of activity space represent groups of activities of medium complexity. These groups correspond to activities associated with central government, which concentrate in New Zealand's capital city, and with the construction sector, which concentrates in cities with high residential growth.

Smart specialization opportunities

We embed our relatedness and complexity estimates within Balland et al.'s (2019) framework for analysing smart specialization. Their framework identifies low-risk, high-return development opportunities as locally under-represented activities – that is, activities a with location quotient $LQ_c^a < 1$ in city c – that have high mean local relatedness and high complexity.

Figure 2 plots mean local relatedness against complexity for locally under-represented activities in three cities: Auckland, New Zealand's most populous city with 573,150 employed usual residents at the 2013 Census; Wellington, the nation's capital with 185,844 employed usual residents at the 2013 Census; and Huntly, a small coal mining town with 1611 employed usual residents at the 2013 Census.

Auckland's relatively large and diverse labour market facilitates specialization in the most complex activities. Figure 2 shows that there is limited scope for Auckland to expand into new low-risk, high-return activities because it is already specialized in such activities. In contrast, Wellington appears poised for employment growth in complex activities because such activities are locally under-represented but also highly related to existing specializations.

Huntly's small size makes it appear relatively specialized in all but the nationally largest activities because few local employees are needed to obtain location quotients above unity. For example, only 876 of the 1,503,018 usual residents employed nationally at the 2013 Census were employed as trades workers in the specialized food retailing industry. Thus, a single employee in that activity delivers a location quotient in Huntly of about 1.07. This example highlights the instability of RCA-based relatedness measures for small cities and activities.

Complexity measure comparison

We discussed above why our complexity measure is more robust to noisy employment data than measures based on RCA co-occurrences. We demonstrate our measure's relative robustness using the following bootstrap procedure. First, we randomize the activity portfolio mix within each city by randomly sampling employees with replacement from the observed city-specific activity employment

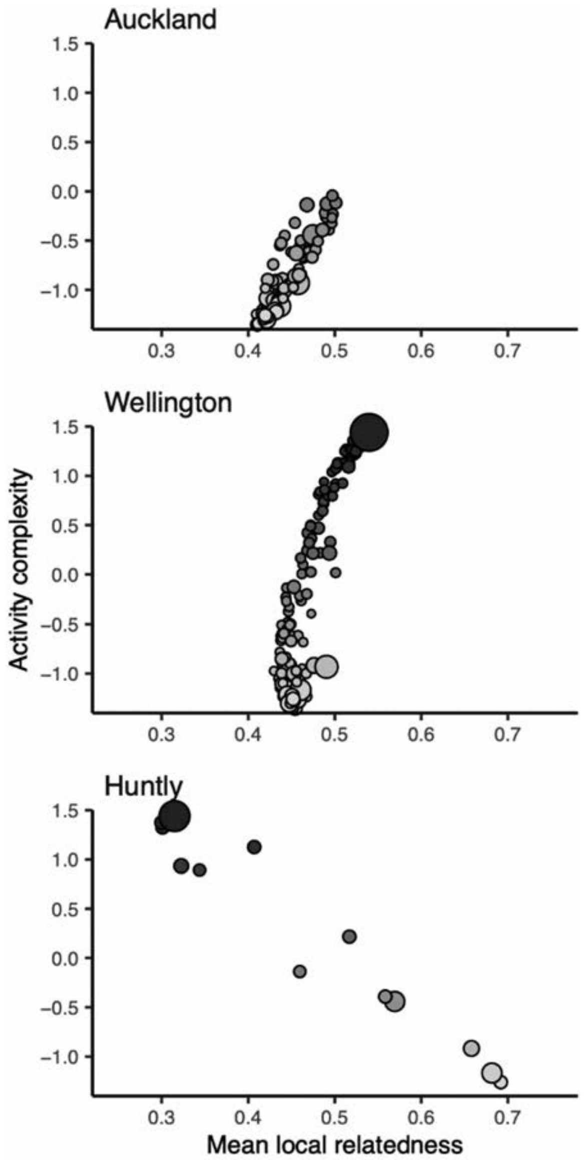

Figure 2. Smart specialization opportunities in Auckland, Wellington and Huntly based on 2013 Census data.
Notes: Points represent activities and are scaled by local share and coloured by activity complexity.

distributions. Next, we use the resulting randomized national employment counts to compute activity complexities using our measure, and the RCA-based measure used in previous studies. This RCA-based measure is implemented in the R package *EconGeo* (Balland, 2017). We repeat this procedure $n = 50$ times, and record the complexity C^{ams} of activity $a \in \mathcal{A}$ using measure m in each sample $s \in \{1, 2, \ldots, n\}$.[7] We then compute the bias

$$b^{am} = \frac{1}{n}\sum_{s=1}^{n}(C^{ams} - C^{am})$$

and Bessel-corrected standard deviation

$$\sigma^{am} = \sqrt{\frac{1}{n-1}\sum_{s=1}^{n}\left(C^{ams} - \frac{1}{n}\sum_{t=1}^{n}C^{amt}\right)^2}$$

in the estimated complexity of activity a using method m across bootstrap samples. Here C^{am} is the complexity of activity a estimated using measure m before bootstrapping. Finally, we compute the means and standard errors of b^{am} and σ^{am} across the 199 activities in our data for each measure m. We summarize these means and standard errors in Table 1, which reports population-weighted and unweighted values for each census year in our data.

The mean biases in our bootstrap activity complexity estimates are not significantly different than zero using our measure or the measure implemented in *EconGeo*. However, the year-specific standard deviations of the estimates using our measure are significantly smaller than those standard deviations using *EconGeo*, suggesting that our measure is more robust to random fluctuations in employment counts. Comparing weighted and unweighted values reveals that the bootstrap estimates using our complexity measure tend to be more precise for larger activities. In contrast, the estimates using *EconGeo* do not exhibit consistent relationships between precision and activity size across years.

Do relatedness and complexity predict employment growth?

Last, we evaluate the contribution of relatedness and complexity to local employment growth. We define the growth rate in city c, activity a employment as the annualised percentage point change

$$G_c^a = 100\left(\left(\frac{E_c^a}{L.E_c^a}\right)^{1/n} - 1\right),$$

where n is the number of years between consecutive observations; and $L.$ is the lag operator.[8] We regress G_c^a on lagged values of local share, relatedness density, activity complexity and city complexity. We use the 16,591 observations for which G_c^a is computable[9] and weight observations by the corresponding lagged share $L.E_c^a/L.E$ of national employment.

We transform our local share estimates by subtracting their weighted mean and multiplying by 100 to obtain demeaned percentage point shares. We also standardize relatedness density, and activity and city complexity, to have zero weighted mean and unit weighted variance. These transformations calibrate our interaction terms to have zero values at covariates' weighted means, easing the interpretation of the coefficients on non-interaction terms. Table 2 presents descriptive statistics for our transformed data before and after weighting by lagged shares of national employment. Comparing the weighted and unweighted means reveals that, on average, observations with larger city–activity employment counts are associated with slower local growth rates, greater local shares, lower relatedness densities and higher city complexities.

Table 3 presents our regression results. Columns (1) and (2) show that relatedness dense activities grew slower during our period of study. On average, and holding activity complexity constant at its weighted mean, a 1 SD (standard deviation) rise in relatedness density corresponds

Table 1. Bias and standard deviation in bootstrapped activity complexity estimates, aggregated across activities.

	Davies and Maré		EconGeo	
Year	Bias	SD	Bias	SD
Population-weighted means (standard errors)				
1981	0.000 (0.003)	0.080 (0.006)	0.025 (0.010)	0.177 (0.010)
1991	−0.002 (0.002)	0.082 (0.006)	−0.005 (0.012)	0.143 (0.010)
2001	−0.001 (0.002)	0.070 (0.006)	−0.005 (0.008)	0.127 (0.010)
2013	0.003 (0.001)	0.059 (0.005)	0.052 (0.011)	0.197 (0.013)
Unweighted means (standard errors)				
1981	0.000 (0.005)	0.124 (0.008)	0.000 (0.012)	0.174 (0.011)
1991	0.000 (0.003)	0.114 (0.007)	0.000 (0.014)	0.173 (0.011)
2001	0.000 (0.003)	0.107 (0.007)	0.000 (0.009)	0.148 (0.011)
2013	0.000 (0.002)	0.093 (0.006)	0.000 (0.010)	0.134 (0.011)

Table 2. Descriptive statistics for data used in the main regressions.

	Unweighted		Weighted			
Variable	Mean	SD	Mean	SD	Minimum	Maximum
Local growth rate (G_c^a)	0.613	5.010	0.242	4.275	−32.935	34.472
Local share ($L.LS_c^a$)	−0.736	0.976	0.000	1.639	−1.337	22.293
Davies and Maré						
Relatedness density ($L.RD_c^a$)	0.316	1.285	0.000	1.000	−5.112	4.756
Activity complexity ($L.C^a$)	0.073	0.996	0.000	1.000	−1.244	1.756
City complexity ($L.C_c$)	−0.756	0.840	0.000	1.000	−1.991	1.380
EconGeo						
Relatedness density ($L.RD_c^a$)	−0.486	0.857	0.000	1.000	−3.316	4.061
Activity complexity ($L.C^a$)	−0.075	0.885	0.000	1.000	−0.972	3.842
City complexity ($L.C_c$)	−0.908	0.651	0.000	1.000	−1.604	1.823

Note: There is a total of 16,714 observations for each variable. Observations are city–activity–year tuples. The minimums and maximums of G_c^a correspond to city–activity employment counts that fell from 489 to 9 and rose from 9 to 174 between consecutive censuses.

to a 0.42 percentage point decline in local employment growth per year. This effect does not change significantly when we control for city complexity.

The coefficient estimates in columns (1) and (2) of Table 3 may be biased by unobservable, time-varying activity and city factors that are correlated with our chosen covariates. We control for these factors in column (3) by including activity–year and city–year fixed effects. This allows us to identify the effects of cross-sectional variation in local growth rates, controlling for the period-specific growth experienced by each activity across New Zealand and for the period-specific growth experienced by each city as a whole. However, we lose the ability to separately identify coefficients on activity and city complexity because these covariates are perfectly collinear with our fixed effects. The negative coefficient on local share implies that employment growth was faster for activities that initially represented smaller shares of local employment. Thus, on average, cities diversified their local activity portfolios during our period of study. The negative coefficients on the interaction of local share with activity and city complex imply that this diversification was faster into more complex activities and within more complex cities.

Balland et al.'s (2019) framework characterizes high relatedness density as an indicator of low-risk local investment options and activity complexity as an indicator of high reward options. Their framework thus suggests that complex activities with high local relatedness offer the strongest prospects for future growth. If this were true in our setting, then we would expect a strong positive coefficient on the interaction of relatedness density and activity complexity. However, our estimates in column (3) of Table 3 show only a weak and insignificant interaction (−0.059) after controlling for time-varying activity and city factors.

In contrast with our results, Rigby et al. (2019) estimate a positive, statistically significant relationship between smart specialization prescriptions and employment growth. They use the relatedness and complexity measures provided by the R package *EconGeo* (Balland, 2017). To determine whether our conclusions differ from Rigby et al.'s due to differences in methodology or because of differences in context, we compare our regression estimates in columns (1)–(3) of Table 3 with the estimates obtained using *EconGeo*'s relatedness and complexity measures. We transform these measures using the procedure described above and

Table 3. Main regression estimates.

	Dependent variable: Local growth rate (G_c^a)				
	Davies and Maré			EconGeo	
	(1)	(2)	(3)	(4)	(5)
Local share ($L.LS_c^a$)	−0.118	−0.115	−0.243***	0.177**	−0.159*
	(0.120)	(0.122)	(0.059)	(0.073)	(0.082)
Relatedness density ($L.RD_c^a$)	−0.419***	−0.414***	−0.085**	0.022	−0.147*
	(0.083)	(0.096)	(0.040)	(0.103)	(0.081)
Activity complexity ($L.C^a$)	−0.104	−0.067		0.521***	
	(0.138)	(0.144)		(0.103)	
Activity complexity × local share	−0.156	−0.173	−0.233***	0.097	−0.310***
	(0.140)	(0.141)	(0.076)	(0.062)	(0.112)
Activity complexity × relatedness density	0.211***	0.165*	−0.059	0.021	0.127***
	(0.076)	(0.087)	(0.055)	(0.053)	(0.032)
City complexity ($L.C_c$)		0.117		0.086	
		(0.101)		(0.115)	
City complexity × local share		0.010	−0.061**	0.180**	0.094***
		(0.064)	(0.026)	(0.071)	(0.032)
City complexity × relatedness density		−0.095	−0.106***	−0.214**	0.133**
		(0.099)	(0.035)	(0.088)	(0.065)
Activity–year and city–year fixed effects			Yes		Yes
Observations	16,591	16,591	16,591	16,591	16,591
R^2	0.015	0.016	0.755	0.021	0.754

Note: Values are ordinary least squares (OLS) estimates with analytic weights equal to lagged shares of total employment. Heteroskedasticity-robust standard errors are shown in parentheses: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. Observations are city–activity–year tuples. Constant terms are included in the models but suppressed in the table.

Table 4. Weighted covariate means and standard deviations, by regression subsample.

	Local growth rate	Local share	Relatedness density	Activity complexity	City complexity
All data	0.242 (4.275)	0.000 (1.639)	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)
Local over-representation subsamples					
$L.\text{RCA}_c^a = 1$	−0.194 (4.283)	0.172 (1.902)	0.537 (0.776)	0.179 (1.040)	−0.119 (1.008)
$L.\text{RCA}_c^a = 0$	0.784 (4.203)	−0.214 (1.203)	−0.667 (0.831)	−0.222 (0.900)	0.148 (0.969)
Urban area type subsamples					
AHWC	0.539 (3.993)	−0.186 (1.190)	−0.174 (0.915)	0.123 (1.023)	0.399 (0.861)
Main excluding AHWC	−0.240 (4.684)	0.129 (1.634)	0.178 (0.876)	−0.196 (0.929)	−0.649 (0.816)
Secondary	−0.417 (4.861)	0.680 (2.668)	0.624 (1.281)	−0.294 (0.876)	−0.988 (0.656)
Minor	−0.958 (4.422)	2.173 (4.793)	1.254 (1.748)	−0.451 (0.783)	−1.214 (0.523)

Notes: Weighted standard deviations are shown in parentheses.
AHWC, Auckland, Hamilton, Wellington and Christchurch.

present descriptive statistics for the resulting transformed covariates in Table 2.

Columns (4) and (5) of Table 3 report coefficient estimates using *EconGeo* before and after introducing activity–year and city–year fixed effects. We use the same analytic weights in these columns as in columns (1)–(3). The estimates using our relatedness and complexity measures suggest different patterns than the estimates using *EconGeo*. For example, our measures imply a weak negative relationship between activity complexity and local employment growth, whereas the *EconGeo* measures imply a strong positive relationship. Likewise, introducing fixed effects weakens the interaction between activity complexity and relatedness density using our measures but strengthens it using *EconGeo*. These differences in patterns may be context-specific, and we therefore encourage further research on how and why such differences arise using data from different geographical contexts.

Are the effects of relatedness and complexity context dependent?

Estimating regression coefficients using all observations in our data may mask effects that are relevant only for particular activities or local contexts. We analyse subsamples of our data in order to investigate the variation in attributes of the activities and cities to which the growth benefits of relatedness and complexity accrue. Table 4 reports weighted means and standard deviations for each subsample. We transform our subsample data so that local share has zero within-subsample weighted mean, and so that relatedness density, activity complexity and city complexity have zero weighted mean and unit weighted variance. We use the model specification in column (3) of Table 3 as our preferred specification throughout our subsample analyses.

Table 5 summarizes our subsample analyses. Its second and third columns of coefficient estimates show the contribution of relatedness and complexity to local employment growth in existing and potential specializations. We identify these contributions by re-estimating our preferred model specification among observations with $L.\text{RCA}_c^a = 1$ and observations with $L.\text{RCA}_c^a = 0$. The interaction of activity complexity and relatedness density was negative and insignificant in our subsample of locally under-represented activities. Thus, Balland et al.'s (2019) characterization of smart specialization does not explain employment dynamics in our data.

The mechanisms underlying relatedness and complexity may operate only at certain city scales. For example, knowledge spillovers may be relevant only in cities with sufficiently thick labour markets. We analyse the scale dependence of relatedness and complexity by partitioning our data into four subsamples according to cities' urban area types. 'Main' urban areas represent the largest urbanized areas in New Zealand; 'secondary' and 'minor' urban areas tend to be smaller and less densely populated.[10] We separate the four largest cities – Auckland, Hamilton, Wellington and Christchurch (AHWC) – from other cities classified as main urban areas. Table 4 shows that, on average, larger cities grew faster during our period of study, had more diverse local labour markets and contained more complex activities.

The last four columns of Table 5 summarize our subsample analyses by urban area type. Holding activity and city complexity constant at their weighted means, we find strong negative relationships between local employment growth and relatedness density in main urban areas but weak positive relationships in smaller cities. We find a positive relationship between local employment growth and the interaction of activity complexity and relatedness density in all four subsamples, but this relationship is significant in our subsample of the four largest cities only. This scale-dependence echoes McCann and Ortega-Argilés (2015) claim that peripheral regions have limited capacities to pursue development paths in line with smart specialization prescriptions. Overall, our results support the characterization of cities as dense networks of interacting activities: the benefits of such interaction are more apparent in larger cities where workers engaged in related activities interact more frequently.

Table 5. Subsample regression estimates.

	Dependent variable: Local growth rate (G_c^a)						
		Local over-representation		Urban area type			
	All data	$\boldsymbol{L}.\text{RCA}_c^a = \mathbf{1}$	$\boldsymbol{L}.\text{RCA}_c^a = \mathbf{0}$	**AHWC**	**Main excluding AHWC**	**Secondary**	**Minor**
Local share ($L.LS_c^a$)	−0.243***	−0.266***	−0.908	0.199	−0.880***	−0.383***	−0.459***
	(0.059)	(0.080)	(0.570)	(0.310)	(0.134)	(0.119)	(0.125)
Relatedness density ($L.RD_c^a$)	−0.085**	0.029	0.032	−0.338***	−0.294**	0.030	0.436
	(0.040)	(0.070)	(0.081)	(0.076)	(0.122)	(0.203)	(0.354)
Activity complexity × local share	−0.233***	−0.103	−1.378***	−0.305	−0.750***	−0.308***	−0.060
	(0.076)	(0.091)	(0.435)	(0.317)	(0.159)	(0.119)	(0.203)
Activity complexity × relatedness density	−0.059	−0.058	−0.011	0.208*	0.094	0.217	0.297
	(0.055)	(0.067)	(0.059)	(0.111)	(0.144)	(0.231)	(0.376)
City complexity × local share	−0.061**	−0.139***	−0.032	−0.044	−0.031	0.058	0.039
	(0.026)	(0.035)	(0.042)	(0.049)	(0.033)	(0.070)	(0.047)
City complexity × relatedness density	−0.106***	−0.135**	−0.096**	0.048	0.044	−0.038	−0.093
	(0.035)	(0.060)	(0.040)	(0.057)	(0.030)	(0.069)	(0.093)
Observations	16,591	8,750	7,841	2,384	7,144	5,128	1,935
R^2	0.755	0.795	0.810	0.887	0.756	0.595	0.617

Notes: Values are ordinary least squares (OLS) estimates with analytic weights equal to lagged within-subsample employment shares. Heteroskedasticity-robust standard errors are shown in parentheses: $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Observations are city–activity–year tuples. Constant terms are included in the models but are suppressed from the table. All models include activity–year and city–year fixed effects.
AHWC, Auckland, Hamilton, Wellington and Christchurch.

Table 6. Robustness regression estimates.

	Dependent variable: Local growth rate (G_c^a)			
	All cities	**All zones**	**AHWC cities**	**AHWC zones**
Local share ($L.LS_c^a$)	−0.243***	−0.266***	0.199	−0.539***
	(0.059)	(0.044)	(0.310)	(0.157)
Relatedness density ($L.RD_c^a$)	−0.085**	0.059*	−0.338***	0.017
	(0.040)	(0.032)	(0.076)	(0.052)
Activity complexity × local share	−0.233***	−0.138**	−0.305	0.276**
	(0.076)	(0.060)	(0.317)	(0.135)
Activity complexity × relatedness density	−0.059	0.030	0.208*	−0.069
	(0.055)	(0.040)	(0.111)	(0.066)
City complexity × local share	−0.061**	−0.084***	−0.044	−0.051*
	(0.026)	(0.023)	(0.049)	(0.030)
City complexity × relatedness density	−0.106***	−0.142***	0.048	0.057*
	(0.035)	(0.026)	(0.057)	(0.034)
Observations	16,591	21,352	2384	6589
R^2	0.755	0.719	0.887	0.820

Notes: Values are ordinary least squares (OLS) estimates with analytic weights equal to lagged within-subsample employment shares. Heteroskedasticity-robust standard errors are shown in parentheses: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. Observations are city–activity–year tuples. Constant terms are included in the models but are suppressed from the table. All models include activity–year and city–year fixed effects.
AHWC, Auckland, Hamilton, Wellington and Christchurch.

Do different spatial scales suggest different relationships?

Merging zones within cities precludes us from using intra-city variation to generate our regression estimates. This variation may be relevant if relatedness and complexity capture mechanisms that operate at smaller spatial scales than can by observed by comparing cities. We explore this possibility in Table 6, which reports coefficient estimates obtained using our relatedness and complexity measures before and after merging urban zones.[11] We generate pre-merge estimates by recomputing activity and city relatedness and complexities using all 50 urban areas in our data, rather than the 41 cities obtained after merging zones, and re-estimating the model specification in column (3) of Table 3.

Separating zones changes what we measure to be related and complex. For example, our activity relatedness and complexity estimates using 2013 Census data before merging zones share Spearman correlations of 0.757 and 0.545 with their values after merging zones. These differences in rankings reflect differences in activity co-location patterns observed at different spatial scales.

The first two columns of Table 6 report coefficient estimates using city- and zone-level variation across all cities in our data. Both columns provide strong evidence of activity portfolio diversification during our period of study, particularly among complex activities and within complex cities. The interaction between activity complexity and relatedness density remains insignificant when we separate zones within cities.

The last two columns of Table 6 report coefficient estimates using city- and zone-level variation among New Zealand's four largest cities. Among these cities, we find no evidence of diversification at the city level but strong evidence at the zone level. Zones in New Zealand's largest four cities appear to have diversified towards complex activities, but these activities were not related to existing specializations. We attribute these patterns to spatial sorting within cities: complex activities became more spatially concentrated at the zone level, with limited observable changes at the city level. Such sorting occurs in sufficiently large cities only, which may explain why we fail to identify differences in growth patterns between cities and zones when we pool across all cities in our data. Accounting for intra-city spatial sorting lies beyond the scope of this paper but remains a potential avenue for future research.

CONCLUSIONS

This paper makes two substantive contributions to the existing literature. First, we describe new measures of relatedness and complexity that are more appropriate than existing measures for analysing the interaction among economic activities in small cities. Second, we use these measures to obtain new evidence on the contribution of relatedness and complexity to urban employment growth in New Zealand, where local labour markets are smaller than those in geographical areas studied previously.

We find some evidence that relatedness and complexity are complementary in promoting employment growth within the larger cities in our data, but no evidence that relatedness and complexity contribute to employment growth within the smaller cities in our data. Our results differ from those obtained using the relatedness and complexity measures applied in previous studies. This difference highlights the importance of using contextually appropriate measures when evaluating the potential efficacy of urban and regional growth and innovation policies. We encourage further research using data from different contexts to compare our measures to those used in previous studies.

Overall, we do not identify strong effects of relatedness and complexity on growth in local activity employment in New Zealand. It is an open question whether this absence means that these effects do not operate or that New Zealand cities lack the scale for such operation. Our results may reflect the limited capacity for knowledge specialization within New Zealand's local labour markets. Alternatively, our failure to identify strong effects may reflect how, within New Zealand and during our period of study, policies were not explicitly designed to encourage or support relatedness and complexity. The absence of such targeted policies may have prevented any potential employment growth benefits of smart specialization policies from being realized.

Likewise, it is an open question whether there is a limited spatial horizon for the growth and innovation benefits of related activities' interaction. Our comparison of city- and zone-level growth dynamics in Table 6 shows that the apparent benefits of such interaction depend upon the spatial scale at which those benefits are measured. This dependence may, in turn, depend on the presence of geographically distributed inter-firm networks, such as integrated supply chains and 'virtual enterprises' that facilitate more distant interactions among related activities. We leave these dependencies to future studies.

Finally, whereas our analysis focuses on the contribution of relatedness and complexity to urban employment growth, future research could examine the contribution of these measures to other outcomes such as innovation, entrepreneurship, and firm entry and exit.

ACKNOWLEDGEMENTS

Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The results presented in the study are the work of the authors, not Statistics New Zealand. The authors thank Andrew Coleman, Edward Glaeser, Dean Hyslop, Guanyu Zheng, seminar participants at Motu and the University of Canterbury, and three anonymous referees for their comments and suggestions.

DISCLOSURE STATEMENT

No conflict of interest was reported by the authors.

FUNDING

This work was supported by Te Pūnaha Matatini and Building Better Homes, Towns and Cities National Science Challenge.

NOTES

1. We capture employee locations using residential addresses rather than workplace addresses due to the substantial incompleteness of available workplace data.
2. Our manual grouping identifies 65 distinct industries, which represent aggregations of the 103 NZSIOC codes used by Statistics New Zealand to generate input–output tables. These NZSIOC codes, in turn, represent aggregations of 214 three-digit 2006 ANZSIC codes.
3. We analyse decade-separated censuses, rather than consecutive censuses, in order to balance the anticipated trade-off between predicting local employment growth using out-of-date relatedness and complexity estimates, and allowing too little time for local conditions to affect employment dynamics. The interval between 2001 and 2013 is longer than a decade because the 2011 Census was postponed due to the 2011 Christchurch earthquake.
4. Davies and Maré (2019, tabs 11–13) provide year-specific industry and occupation employment counts, and indicate the industry–occupation pairs included in our data.
5. Balland and Rigby (2017) study 438 technology classes, Balland et al. (2019) study 617 technology classes, and Rigby et al. (2019) study 652 technology classes.
6. The minimum is obtained by service and sales workers in the air and space transport industry.
7. We obtain similar results using larger values of n.
8. Thus $n = 10$ for city–activity observations in 1991 and 2001, while $n = 12$ for observations in 2013.
9. We lose observations for two suppression-induced reasons. First, there are 8538 observations for which the unrounded value of E_c^a is below the suppression threshold, preventing us from computing the local share LS_c^a, relatedness density RD_c^a and growth rate G_c^a. Second, there are another 7507 observations for which the unrounded value of $L.E_c^a$ falls below the suppression threshold, preventing us from computing G_c^a. We account for suppressed values of LS_c^a by computing relatedness indices based on pairwise complete observations.
10. Statistics New Zealand imposes minimum population thresholds in its classification of urban areas. These thresholds are 30,000 for main urban areas, 10,000 for secondary urban areas and 1000 for minor urban areas.
11. These merges affect Auckland (four zones), Hamilton (three zones), Wellington (four zones), and Napier and Hastings (two zones). These four cities contribute 58.3% of the employees in our data across all years.

ORCID

Benjamin Davies http://orcid.org/0000-0002-1749-4567
David C. Maré http://orcid.org/0000-0003-4716-2755

REFERENCES

Asheim, B. T., & Gertler, M. S. (2005). The geography of innovation: Regional innovation systems. In J. Fagerberg, D. C.

Mowery, & R. R. Nelson (Eds.), *The Oxford handbook of innovation* (pp. 291–317). Oxford University Press.
Balassa, B. (1965). Trade liberalisation and 'revealed' comparative advantage. *Manchester School*, *33*(2), 99–123. https://doi.org/10.1111/j.1467-9957.1965.tb00050.x
Balland, P.-A. (2017). *Economic geography in R: Introduction to the EconGeo package*. Papers in Evolutionary Economic Geography 17.09. Utrecht University.
Balland, P.-A., Boschma, R., Crespo, J., & Rigby, D. L. (2019). Smart Specialisation policy in the European Union: Relatedness, knowledge complexity and regional diversification. *Regional Studies*, *53*(9), 1252–1268. https://doi.org/10.1080/00343404.2018.1437900
Balland, P.-A., Jara-Figueroa, C., Petralia, S., Steijn, M., Rigby, D., & Hidalgo, C. A. (2020). Complex economic activities concentrate in large cities. *Nature Human Behaviour*, *4*(3), 248–254. https://doi.org/10.1038/s41562-019-0803-3
Balland, P.-A., & Rigby, D. L. (2017). The geography of complex knowledge. *Economic Geography*, *93*(1), 1–23. https://doi.org/10.1080/00130095.2016.1205947
Barca, F. (2009). *An agenda for a reformed Cohesion Policy: A place-based approach to meeting European Union challenges and expectations* (Independent report prepared at the request of Danuta Hübner, Commissioner for Regional Policy). European Commission.
Barca, F., McCann, P., & Rodríguez-Pose, A. (2012). The case for regional development intervention: Place-based versus place-neutral approaches. *Journal of Regional Science*, *52*(1), 134–152. https://doi.org/10.1111/j.1467-9787.2011.00756.x
Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, *39*(1), 61–74. https://doi.org/10.1080/0034340052000320887
Boschma, R. (2014). Constructing regional advantage and smart specialisation: Comparison of two European policy concepts. *Italian Journal of Regional Science*, *13*(1), 51–68. https://doi.org/10.3280/SCRE2014-001004
Boschma, R. (2017). Relatedness as driver of regional diversification: A research agenda. *Regional Studies*, *51*(3), 351–364. https://doi.org/10.1080/00343404.2016.1254767
Boschma, R., Balland, P.-A., & Kogler, D. F. (2015). Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. *Industrial and Corporate Change*, *24*(1), 223–250. https://doi.org/10.1093/icc/dtu012
Boschma, R. A., & Frenken, K. (2006). Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *Journal of Economic Geography*, *6*(3), 273–302. https://doi.org/10.1093/jeg/lbi022
Breschi, S., & Lissoni, F. (2005). 'Cross-firm' inventors and social networks: Localized knowledge spillovers revisited. *Annales d'Économie et de Statistique*, 79/80, 189–209. https://doi.org/10.2307/20777575
Caldarelli, G., Cristelli, M., Gabrielli, A., Pietronero, L., Scala, A., & Tacchella, A. (2012). A network analysis of countries' export flows: Firm grounds for the building blocks of the economy. *PLoS ONE*, *7*(10). https://doi.org/10.1371/journal.pone.0047278
Content, J., & Frenken, K. (2016). Related variety and economic development: A literature review. *European Planning Studies*, *24*(12), 2097–2112. https://doi.org/10.1080/09654313.2016.1246517
Davies, B., & Maré, D. C. (2019). *Relatedness, complexity and local growth*. Motu Working Paper 19-01. Motu Economic and Public Policy Research.
Delgado, M., Porter, M. E., & Stern, S. (2014). Clusters, convergence and economic performance. *Research Policy*, *43*(10), 1785–1799. https://doi.org/10.1016/j.respol.2014.05.007
Delgado, M., Porter, M. E., & Stern, S. (2016). Defining clusters of related industries. *Journal of Economic Geography*, *16*(1), 1–38. https://doi.org/10.1093/jeg/lbv017
Farinha Fernandes, T., Balland, P.-A., Morrison, A., & Boschma, R. (2019). What drives the geography of jobs in the US? Unpacking relatedness. *Industry and Innovation*, *26*(9), 988–1022. https://doi.org/10.1080/13662716.2019.1591940
Feldman, M. P., & Audretsch, D. B. (1999). Innovation in cities: Science-based diversity, specialisation and localized competition. *European Economic Review*, *43*(2), 409–429. https://doi.org/10.1016/S0014-2921(98)00047-6
Foray, D., David, P. A., & Hall, B. (2009). *Smart Specialisation – The concept*. Knowledge Economists Policy Brief No. 9. European Commission.
Foray, D., David, P. A., & Hall, B. (2011). *Smart Specialisation: From academic idea to political instrument, the surprising career of a concept and the difficulties involved in its implementation*. MTEI Working Paper. École Polytechnique Fédérale de Lausanne.
Frenken, K., van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, *41*(5), 685–697. https://doi.org/10.1080/00343400601120296
Fritsch, M., & Kublina, S. (2018). Related variety, unrelated variety and regional growth: The role of absorptive capacity and entrepreneurship. *Regional Studies*, *52*(10), 1360–1371. https://doi.org/10.1080/00343404.2017.1388914
Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, *21*(1), 1129–1164. https://doi.org/10.1002/spe.4380211102
Hidalgo, C. (2015). *Why information grows: The evolution of order, from atoms to economies*. Basic Books.
Hidalgo, C. A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Glaeser, E., He, C., Kogler, D. F., Morrison, A., Neffke, F., Rigby, D., Stern, S., Zheng, S., & Zhu, S. (2018). The principle of relatedness. In A. J. Morales, C. Gershenson, D. Braha, A. A. Manai, & Y. Bar-Yam (Eds.), *Unifying themes in complex systems IX* (pp. 451–457). Springer.
Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences, USA*, *106*(26), 10570–10575. https://doi.org/10.1073/pnas.0900943106
Hidalgo, C. A., Klinger, B., Barabasi, A.-L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, *317*(5837), 482–487. https://doi.org/10.1126/science.1144581
Jacobs, J. (1969). *The economy of cities*. Vintage.
Jara-Figueroa, C., Jun, B., Glaeser, E. L., & Hidalgo, C. A. (2018). The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. *Proceedings of the National Academy of Sciences, USA*, *115*(50), 12646–12653. https://doi.org/10.1073/pnas.1800475115
Marshall, A. (1920). *Principles of economics* (8th ed.). Macmillan.
Maskell, P., & Malmberg, A. (1999). Localised learning and industrial competitiveness. *Cambridge Journal of Economics*, *23*(2), 167–185. https://doi.org/10.1093/cje/23.2.167
McCann, P. (2009). Economic geography, globalisation and New Zealand's productivity paradox. *New Zealand Economic Papers*, *43*(3), 279–314. https://doi.org/10.1080/00779950903308794
McCann, P., & Ortega-Argilés, R. (2015). Smart Specialisation, regional growth and applications to European Union Cohesion Policy. *Regional Studies*, *49*(8), 1291–1302. https://doi.org/10.1080/00343404.2013.799769
Mealy, P., Doyne Farmer, J., & Teytelboym, A. (2019). Interpreting economic complexity. *Science Advances*, *5*(1). https://doi.org/10.1126/sciadv.aau1705
Morkutė, G., Koster, S., & Van Dijk, J. (2017). Employment growth and inter-industry job reallocation: Spatial patterns and relatedness. *Regional Studies*, *51*(6), 958–971. https://doi.org/10.1080/00343404.2016.1153800

Neffke, F., Hartog, M., Boschma, R., & Henning, M. (2018). Agents of structural change: The role of firms and entrepreneurs in regional diversification. *Economic Geography*, *94*(1), 23–48. https://doi.org/10.1080/00130095.2017.1391691

Neffke, F., & Henning, M. (2013). Skill relatedness and firm diversification. *Strategic Management Journal*, *34*(3), 297–316. https://doi.org/10.1002/smj.2014

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. https://doi.org/10.1111/j.1944-8287.2011.01121.x

Rigby, D. L. (2015). Technological relatedness and knowledge space: Entry and exit of US cities from patent classes. *Regional Studies*, *49*(11), 1922–1937. https://doi.org/10.1080/00343404.2013.854878

Rigby, D. L., Roesler, C., Kogler, D., Boschma, R., & Balland, P.-A. (2019). *Do EU regions benefit from smart specialization?* Papers in Evolutionary Economic Geography 19.31. Utrecht University.

Schumpeter, J. (1934). *The theory of economic development*. Harvard University Press.

Shi, J., & Malik, J. (2000). Normalised cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(8), 888–905. https://doi.org/10.1109/34.868688

Sorenson, O. (2005). Social networks, informational complexity and industrial geography. In D. Fornahl, C. Zellner, & D. B. Audretsch (Eds.), *The role of labour mobility and informal networks for knowledge transfer* (Vol. 6, pp. 79–96). Springer.

Sorenson, O., Rivkin, J. W., & Fleming, L. (2006). Complexity, networks and knowledge flow. *Research Policy*, *35*(7), 994–1017. https://doi.org/10.1016/j.respol.2006.05.002

Vicente, J., Balland, P.-A., & Crespo, J. (2018). Les fondements micro du changement structurel régional: Que nous enseignent 25 ans de proximités? *Revue d'Économie Régionale & Urbaine*, *5*(6), 1013–1041. https://doi.org/10.3917/reru.185.1013

Weitzman, M. L. (1998). Recombinant growth. *Quarterly Journal of Economics*, *113*(2), 331–360. https://doi.org/10.1162/003355398555595