

EXPLORING THE LINK BETWEEN ECONOMIC COMPLEXITY AND EMERGENT ECONOMIC ACTIVITIES

Alex Bishop and Juan Mateos-Garcia*

Recent studies have shown a strong link between the complexity of economies and their economic development. There remain gaps in our understanding of the mechanisms underpinning these links, in part because they are difficult to analyse with highly aggregated, official data sources that do not capture the emergence of new industrial activities, a potential benefit from complexity. We seek to address some of these gaps by calculating two indices of economic complexity for functional local economies (Travel to Work Areas) in Great Britain, and explore their link with these locations' economic performance. Seeking to gain a better understanding of the mechanism connecting economic complexity with economic performance, we create a measure of emergent technological activity in a location based on a combination of novel data sources including text from UK business websites and CrunchBase, a technology company directory, and study its link with economic complexity. Our results highlight the potential value of novel, unstructured data sources for the analysis of the links between economic complexity and regional economic development.

Keywords: complexity, economic geography, big data, emergent technologies.

JEL codes: R12; O31.

I. Introduction

In the past ten years, researchers have detected a robust link between the diversity and sophistication of productive capabilities in nations and their wealth in terms of GDP per capita and other important economic outcomes such as lower inequality (Hidalgo *et al.*, 2007; Hidalgo and Hausmann, 2009; Hartmann *et al.*, 2017).

If a location displays a comparative advantage in the supply of complex products, this suggests the presence of various hard to imitate capabilities enabling it to capture more market share (Balland and Rigby, 2017). Economic complexity also shapes a location's future specialisation trajectory through the Principle of Relatedness. According to this principle, nations tend to enter activities for which they already have relevant knowledge and capabilities (Hidalgo *et al.*, 2017; Frenken *et al.*, 2007). Economically complex locations have more 'options' to diversify their economy in response to new opportunities and shocks.

We would also expect complex economies to be able to generate 'emergent' phenomena in the form of new technologies and industries that recombine local capabilities, knowledge and skills in novel ways (Rotolo, Hicks and Martin, 2015). However, this link remains understudied. This is in part due to limitations in the data: it is not possible to identify novel phenomena with

structured data sources based on pre-existing product, industry and technological categories offering a lagging view of the economy. For example, new sectors such as the Internet of Things or Virtual Reality are not present in the Standard Industrial Classification (SIC) taxonomy used to structure economic data in the UK, dating back to 2007.

But how can we determine if more economically complex areas have a stronger propensity to nurture emergent activities?

The increasing availability of digital data, computational power, and advances in machine learning and natural language processing are opening up new opportunities to expand our understanding of structural change in the economy, including through the analysis of unstructured (text) data in patents and research papers, and the study of new data sources such as business websites (Bakhshi and Mateos-Garcia, 2016; Klinger, Mateos-Garcia and Stathoulopoulos, 2018; Einav and Levin, 2014; Gentzkow, Kelly and Taddy, 2017).

In this paper, we combine a novel dataset with information about 1.5 million business websites in the UK obtained from Glass, a big data startup, and CrunchBase, a global directory of technology companies ideal for the analysis

* Nesta. E-mail: alex.bishop@nesta.org.uk. An appendix to this article is available online.

of innovation and entrepreneurial ecosystems (Dalle, den Besten and Menon, 2017). We integrate these novel data sources with official economic data on employment, business activity and productivity in order to study the link between various measures of economic complexity, the emergence of new ideas and economic outcomes in local economies in the UK.

2. Definitions and data

2.1 Complexity

The seminal approach for estimating economic complexity, the Economic Complexity Index (ECI) is based on the method of reflections, a recursive algorithm that produces a sum of the number of products where a country has a revealed comparative advantage (or of industries where it specialises) ‘weighted’ by their ubiquity (their propensity to appear in many other countries) (Hidalgo *et al.*, 2007). We will transfer it to our empirical context of UK local economies at the Travel to Work Area (TTWA) level by using industrial specialisation profiles (instead of exports) to proxy local capabilities. A non-linear modification to ECI called Fitness is also commonly used to correct some shortcomings of ECI, like its tendency to over-estimate the complexity of regions with low complexity industries that are present in high complexity regions (see the online appendix for more information on the formal definitions of the indices we use, and their strengths and weaknesses) (Caldarelli *et al.*, 2012; Tachella *et al.*, 2012; Pugliese, Zaccaria and Pietronello, 2016; Albeaik *et al.*, 2017).

We will estimate our economic complexity indices using employment data from BRES (The Business Register and Employment Survey) and IDBR (Interdepartmental Business Register) accessed from NOMIS, an online portal with labour market data for the UK.

BRES is an annual survey of 80,000 businesses in Great Britain providing official employee and employment estimates by detailed geography and industry. The Interdepartmental Business Register is a register of 2.6 million VAT (Value-added tax) and PAYE (Pay as you earn) registered businesses in the UK derived from administrative sources. It is used as the sampling frame for official business surveys in the UK.

We study 218 Travel to Work Areas (TTWAs) capturing commuting areas in Great Britain. By focusing on functional economies instead of administrative regions (as in Mealy, Farmer and Teytelboym, 2017, and Mealy and Coyle, 2019) we hope to capture the span of industrial collaboration and knowledge spillovers that drives (and is enabled by) economic complexity. Unfortunately, we

are not able to include Northern Ireland in our analysis because BRES data is not available for this country.

We classify BRES and IDBR data into industries using 4-digit SIC codes. We have clustered these codes into industrial segments of related industries with the algorithm developed in Delgado, Porter and Stern (2015), which measures this relatedness based on employment and business co-location, occupational composition of the workforce and business to business trade based on input-output tables. This gives us a list of 72 unique industries that we use to calculate our measures of economic complexity.

2.2 Emergence

We define emerging economic activities as those which rely on emerging technologies characterised by radical novelty, fast growth, coherence, prominent impact, and uncertainty and ambiguity (Rotolo, Martin and Hicks, 2015). Researchers and policymakers are interested in these technologies because of their potential to transform production processes or business models and enhance productivity. In this paper, we seek to operationalise this definition by identifying businesses in the UK that develop or have adopted those emerging technologies. However, as we mentioned in Section 1, it is not possible to capture this information in official data sources organised around lagging taxonomies such as those we use to measure economic complexity. For this reason, we have to rely on other, less structured sources.

We identify emerging technologies using the evolution of businesses descriptions in CrunchBase, a directory of technology companies increasingly popular in the analysis of start-up ecosystems (Dalle, den Besten and Menon, 2017; Breschi, Lassébie and Menon, 2018; Nathan, Kemeny and Baden Almeer, 2017). More specifically, we construct a dataset of 238,629 companies in the CrunchBase directory consisting of their descriptions and the year when they were founded. We pre-process company descriptions and concatenate them based on the year when they were founded, and identify salient terms that year based on their Term-Frequency Inverse-Document Frequency (TFIDF) score, which normalises the occurrence of words in an observation (in this case a year) by their occurrence in the corpus.

As the results in table 1 show, this analysis captures different ‘eras’ in the digital technology landscape, starting with social networks and smartphones in 2008, moving into Artificial Intelligence since 2013, and then into Blockchain, crypto-currencies and virtual reality (VR) in the last three years. We believe that this

Table 1. Top five terms by TFIDF score and year, CrunchBase

Year	Top terms
2008	social network, online marketing, social networking, engine optimisation, app development
2009	social network, online marketing, social networking, twitter, engine optimisation
2010	social network, online marketing, platform allows, deal, social networks
2011	social network, smartphone, based platform, commerce platform, accelerator
2012	social network, application enables, smartphone, crowdfunding, accelerator
2013	accelerator, social network, crowdfunding, bitcoin, ai
2014	ai, peer, bitcoin, crowdfunding, accelerator
2015	ai, peer, vr, artificial intelligence, machine learning
2016	ai, blockchain, machine learning, vr, artificial intelligence
2017	blockchain, ai, cryptocurrency, artificial intelligence, decentralised
2018	blockchain, ai, crypto, cryptocurrency, decentralised

approach reflects key aspects of the above definition of technological emergence such as novelty, fast growth and prominent impact.

In principle, we could use CrunchBase to identify and map companies mentioning emergent terms, and in this way calculate indicators of emergence at the TTWA level; however, this could skew our measurement towards clusters of companies *developing* emerging technologies and neglect companies that *adopt and deploy* those technologies, an economically important behaviour which might be facilitated by a locations's economic complexity.

To avoid this bias, we will attempt to identify companies developing or adopting emerging technologies in a more comprehensive dataset capturing the universe of UK businesses with a website.

Glass is a UK big data startup that has developed technology that can "understand text at scale". The company has built a web crawler that reads and interprets public websites automatically. To identify UK businesses, the crawler is set to read websites representing public and private organisations that target a UK audience, or have adopted the .uk domain address. Websites are considered if they are written in English, mention a UK address in their page, and have some depth of representation for the organisation.

Starting with over 204 million web pages, 894,277 business websites with a qualified depth of data are found. Then, the content of each website is read and relevant text entities (e.g. business descriptions) are detected with state

of the art precision (>95%). Business descriptions are identified with a machine learning model that considers multiple features such as location on the web-page, use of specific keywords and phrases, sentence structure etc.

Based on descriptions and other attributes, each business is classified into one or more sectors and assigned a weight showing its proximity to the sector.¹ Specialised businesses tend to have a single sector with high weight, while those with a diversified activity have multiple sector predictions with lower weight values.

We geocode the business websites in the Glass data with their main postcode via NSPL (National Statistics Postcode Lookup), a lookup between postcodes and official geographies in the UK.² At the end of this process, we end up with ca. 400,000 unique geocoded business websites.

As mentioned, the Glass data are labelled with multiple sector tags through a predictive analysis of their website content. The LinkedIn taxonomy used for this contains 108 industries. We classify each website into its top sector based on the weight generated by Glass, and then into its official industrial segment through a look-up table.

One potential concern with the Glass data is that it could suffer from biases if, say, companies in some sectors or locations are more likely to have a website. We explore this potential issue by matching the Glass data with official data and comparing its sectoral distribution and geographical distribution. Our results suggest that the Glass data is broadly representative of industrial activity in the UK (see the online appendix for additional details).

In order to identify companies involved in emerging activities in the Glass data, we look for company descriptions with salient terms from the past three years in the CrunchBase data in table 1. In order to avoid low recall due to small differences in spelling or variation in terminology to refer to the same industries or technologies, we have expanded the initial vocabulary of emergent technology trends extracted from CrunchBase with similar terms based on a word2vec model trained on the Glass data.³

In general, the terms in the resulting list capture emerging technologies or related ideas and concepts (see list of results in the online appendix).

In addition to using the Glass dataset to detect emergence, we also analyse the structure of the language

in business websites and compare it between locations with different levels of economic complexity. This way, we seek to improve our understanding of the meaning of our economic complexity indices. In order to do this, we have extracted themes (clusters of terms) from business websites using a topic modelling algorithm. Essentially by considering our documents (business descriptions) as a mixture of topics, with each topic itself being a mixture of words, we can learn topics in an unsupervised manner from the likelihood of word co-occurrences within our documents. This approach allows us to analyse both the extent to which sectors and topics co-relate, and identify topics (and thus words) which correspond to high/low economic complexity.

Recent work (Gerlach, Peixoto and Altmann, 2018) has exploited a mathematical connection between topic models and finding community structure in networks, namely the mathematical equivalence between the Stochastic Block Model (SBM) and probabilistic Latent Semantic Indexing (pLSI), to develop an approach to topic modelling by deriving a non-parametric Bayesian parametrisation (topSBM) of pLSI adapted from a hierarchical SBM (Peixoto, 2017).

2.3 Additional Secondary Data

Previous studies have shown a robust relationship between economic complexity and other important economic variables (Hidalgo *et al.*, 2007; Hartmann *et al.*, 2017; Mealy and Coyle, 2019). Here, we use salary data, from ASHE (Annual Survey of Hours and Earnings) and local economy estimates of GVA per capita generated by the Office for National Statistics (ONS) as economic benchmarks for our economic complexity metrics.⁴ Because these benchmarking datasets are only available at the Local Authority District level, we use

the NSPL (National Statistics Postcode Lookup) to construct a weighted mapping between the two in order to convert median salaries and GVA per capita to the TTWA level (see the on-line appendix for a validation of this mapping.)

3. Findings

The data collection and processing we have undertaken results in three sets of variables: (1) Measures of economic complexity based on the ECI and Fitness definitions using employment (BRES) and establishment (IDBR) data, averaged over the period for which we have data (2015–17); (2) Indices of emergence based on the presence of emerging technology terms (from CrunchBase) in Glass' business website descriptions, and a model capturing topics of terms in business website descriptions, and economic outcomes related to productivity (median earnings and GVA per capita based on secondary data averaged between 2013 and 2018 (for earnings data) and between 2013 and 2015 (for GVA per capita data)).

Table 2 presents descriptive statistics for these variables, and figure 1 displays the bivariate (Spearman) correlations between them. The first thing worth noting is that our measures of complexity (ECI and Fitness) are strongly correlated with one another, suggesting that they are capturing similar features of local economies. All our economic complexity variables are also strongly associated with emergence, measured in terms of business counts and as a share of all businesses in a TTWA. This is consistent with the idea that locations that are more economically complex are also better able to develop and deploy emerging technologies. Complexity measures also display a positive correlation with economic outcomes, in line with recent findings

Table 2. Descriptive statistics for key variables used in the analysis^(a)

	n	max	min	q_25	q_50	q_75
BRES ECI	218	0.187	-0.147	-0.054	-0.013	0.040
IDBR ECI	214	0.108	-0.142	-0.058	0.009	0.052
BRES Fitness	218	1.367	0.901	1.020	1.069	1.124
IDBR Fitness	214	1.186	0.976	1.037	1.087	1.124
Emergent (share)	218	0.019	0.000	0.002	0.004	0.007
Emergent count (log)	218	8.066	-4.605	0.698	2.140	3.258
Median earnings (log)	218	10.612	10.012	10.134	10.188	10.238
GVA per capita (log)	218	10.811	9.565	9.782	9.882	10.020

Note: (a) economic complexity indices (BRES ECI, IDBR ECI, BRES Fitness, IDBR Fitness), number of businesses developing or adopting emerging technologies normalised by IDBR business count (Emergent (share)), log number of businesses developing or adopting emerging technologies (Emergent count (log)), median earnings from ASHE averaged across years (Median earnings (log)), and log GVA per capita averaged across years (GVA per capita (log)).

Figure 1. Pairwise correlations between economic complexity indices^(a)

	BRES ECI	IDBR ECI	BRES Fitness	IDBR Fitness	Emergent (share)	Emergent count (log)
BRES ECI	1.00	0.84	0.89	0.74	0.71	0.74
IDBR ECI	0.84	1.00	0.69	0.93	0.67	0.79
BRES Fitness	0.89	0.69	1.00	0.58	0.67	0.64
IDBR Fitness	0.74	0.93	0.58	1.00	0.55	0.68
Emergent (share)	0.71	0.67	0.67	0.55	1.00	0.78
Emergent count (log)	0.74	0.79	0.64	0.68	0.78	1.00
Median earnings (log)	0.49	0.45	0.51	0.36	0.46	0.32
GVA per capita (log)	0.52	0.40	0.53	0.30	0.52	0.30

Note: (a) economic complexity indices (BRES ECI, IDBR ECI, BRES Fitness, IDBR Fitness), number of businesses developing or adopting emerging technologies normalised by IDBR business count (Emergent (share)), log number of businesses developing or adopting emerging technologies (Emergent count (log)), median earnings from ASHE averaged across years (Median earnings (log)), and log GVA per capita averaged across years (GVA per capita (log)).

in Mealy and Coyle (2019) as well as studies using measures of economic complexity based on export data at the country level.

In the rest of the section we focus further on each of the set of variables that we have calculated and their relation with each other. First, we map the geography of complexity in Great Britain and seek to interpret complexity indices by looking at their composition, and also analysing salient topics in the website descriptions of businesses based on locations with different levels of complexity.

Having done this, we will focus on the geography of emergence in Britain, and its link with economic complexity. We will conclude by integrating our three data sources in a simple correlation analysis between economic complexity, emergence of new activities, and economic outcomes.

3.1 The geography of complexity in Great Britain

The maps in figure 2 present the geography of economic complexity in Great Britain based on the two algorithms we are focusing on (ECI in the first row, and Fitness in the second row) and data sources (employment in

the first column, and number of establishments in the second). Meanwhile, table 3 shows the ten top and bottom TTWAs according to each complexity index and measure of activity.

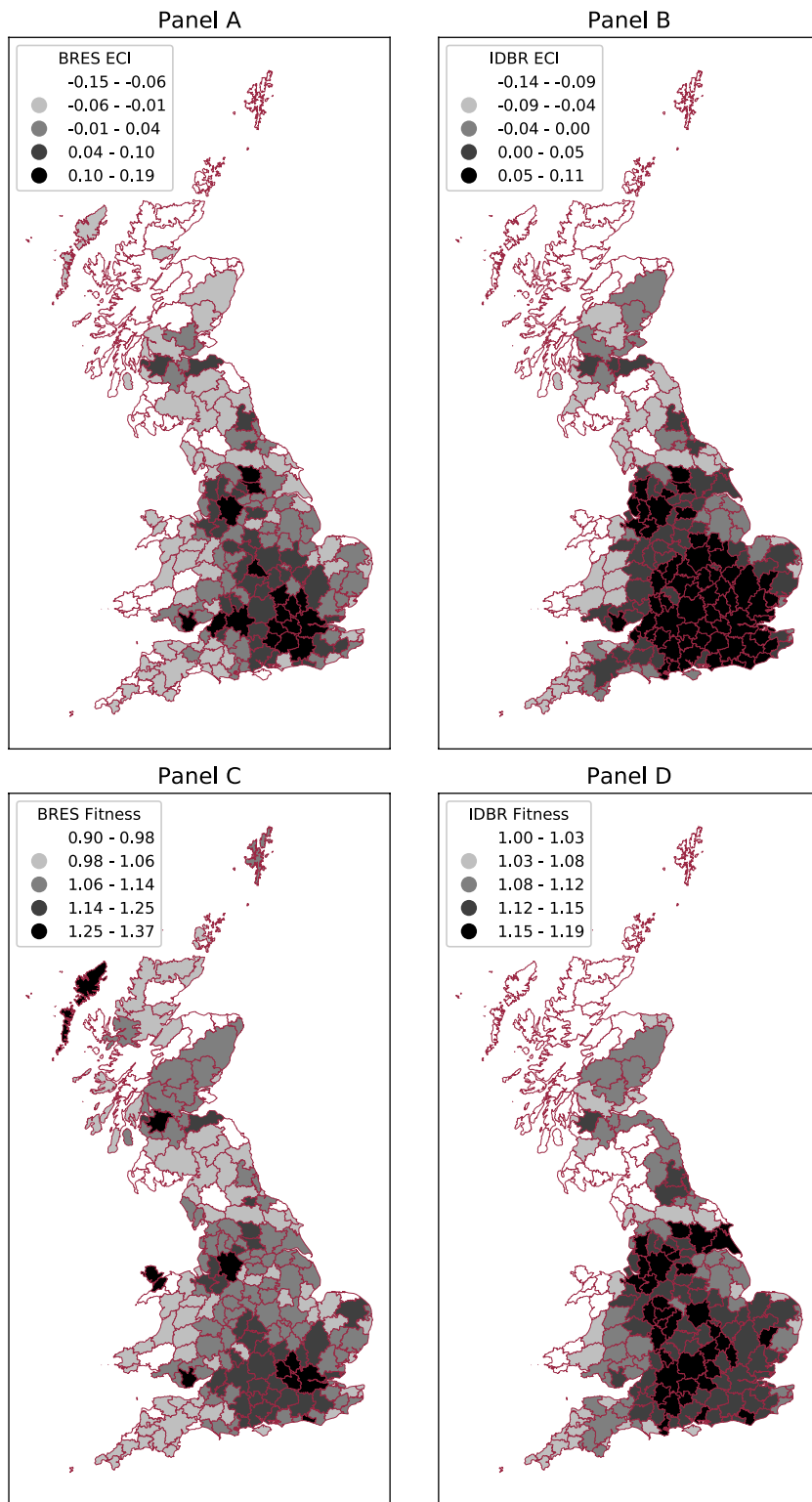
Starting with the maps, we see strong similarities between measures and data sources: all of them show ‘hubs’ of economic complexity around London and the South East of England stretching into the South West along the M4 corridor to reach Bristol and Cardiff, and in the North with Manchester, Leeds and Newcastle, and Edinburgh and Glasgow. The South West, North of Wales, the North West and the Scottish Highlands are zones of low complexity.

These similarities are also present in table 3, where there is a strong overlap in the top and bottom TTWAs according to our complexity measures. Perhaps unsurprisingly, measures of complexity based on employment data tend to favour larger cities such as London, Manchester, Leeds, Cardiff and Glasgow by comparison to those based on business establishments, which tend to be dominated by smaller cities in South England.

What are the sectoral drivers of economic complexity? Table 4 presents the sectors with the highest and lowest levels of complexity according to ECI and Fitness. High complexity sectors (those where high complexity TTWAs tend to specialise) mostly include digital, creative and knowledge intensive industries (here it is worth noting that the Fitness indices – and in particular those based on establishment data – also highlight a few manufacturing sectors such as armaments, casting and pesticides). Low complexity sectors include untraded and commoditised services industries and primary sectors which are more important in rural TTWAs, such as forestry.

The picture is quite similar when we study the language in business websites for locations with different levels of complexity. Table 5 presents the words included in topics that are more important in high complexity TTWAs. We see that this includes topics covering creative, digital, data and knowledge intensive activities. We also note with interest the presence of topics related to new and unique ideas and networking (topics 9 and 6 in the BRES ECI column) and internationalisation (topic 7 in the BRES Fitness column). By contrast, table 6 shows that the dominant topics in low complexity TTWAs refer to activities related to tourism, leisure and heritage. Terms such as ‘old’, ‘historical’ and ‘traditional’ recur in the business websites for these TTWAs.

Figure 2. Geography of complexity in Great Britain



Notes: Panels A and B display ECI based on BRES and IDBR data. Panels C and D display Fitness based on BRES and IDBR data.

Table 3. Top and bottom locations by complexity measure

	BRES ECI	IDBR ECI	BRES Fitness	IDBR Fitness
top 1	Leeds	Reading	Luton	High Wycombe and Aylesbury
top 2	Luton	Swindon	High Wycombe and Aylesbury	Swindon
top 3	Manchester	Leamington Spa	Slough and Heathrow	Andover
top 4	London	Guildford and Aldershot	London	Newbury
top 5	Reading	High Wycombe and Aylesbury	Manchester	Poole
top 6	Slough and Heathrow	Newbury	Brighton	Stevenage and Welwyn Garden City
top 7	Guildford and Aldershot	Luton	Cardiff	Bridlington
top 8	Bristol	Cheltenham	Glasgow	Bedford
top 9	High Wycombe and Aylesbury	Basingstoke	Western Isles	Salisbury
top 10	Basingstoke	Tunbridge Wells	Bangor and Holyhead	Trowbridge
last 10	Berwick	Elgin	Golspie and Brora	Elgin
last 9	Girvan	Oban	Elgin	Mull and Islay
last 8	Haverfordwest and Milford Haven	Mull and Islay	Pwllheli and Porthmadog	Campbeltown
last 7	Whitby	Dunoon and Rothesay	Cromer and Sheringham	Dunoon and Rothesay
last 6	Pwllheli and Porthmadog	Campbeltown	Aviemore and Grantown-on-Spey	Oban
last 5	Newton Stewart	Lochgilthead	Berwick	Lochgilthead
last 4	Orkney Islands	Dunfermline and Kirkcaldy	Newton Stewart	Dunfermline and Kirkcaldy
last 3	Cromer and Sheringham	Perth	Girvan	Perth
last 2	Fraserburgh	Pitlochry and Aberfeldy	Fraserburgh	Pitlochry and Aberfeldy
last 1	Lochgilthead	St Andrews and Cupar	Lochgilthead	St Andrews and Cupar

Table 4. Top and bottom industries by industry complexity measure

	BRES ECI	IDBR PCI	BRES Fitness product	IDBR Fitness product
top 1	services_computing	services_computing	services_content	manufacture_armaments
top 2	services_content	manufacture_armaments	services_computing	manufacture_pesticides
top 3	services_kibs	services_human_resources	services_kibs	services_computing
top 4	services_financial_legal	services_insurance_finance	services_financial_legal	manufacture_casting
top 5	services_insurance_finance	services_kibs	services_insurance_finance	services_creative
top 6	services_administrative	services_wholesale	services_publishing	services_kibs
top 7	services_facilities	services_creative	manufacture_armaments	manufacture_oils
top 8	services_publishing	services_content	services_facilities	manufacture_pharmaceuticals
top 9	services_wholesale	services_printing	services_air_transport	manufacture_steel_metal
top 10	services_telecommunications	services_administrative	services_administrative	services_human_resources
last 10	primary_forestry	services_passenger	manufacture_textiles	services_passenger
last 9	construction_construction	services_public_administration	services_agricultural	primary_forestry
last 8	primary_quarrying_concrete	services_hospitality	manufacture_furniture	services_consumer_retail
last 7	services_passenger	manufacture_ships	manufacture_food	manufacture_food
last 6	services_cultural	services_energy	construction_construction	services_agricultural
last 5	primary_oil_gas	services_cultural	services_cultural	services_education_primary
last 4	services_hospitality	services_sea_transport	primary_quarrying_concrete	services_public_administration
last 3	services_sea_transport	primary_forestry	primary_forestry	services_entertainment
last 2	manufacture_ships	primary_fishing	services_hospitality	services_cultural
last 1	primary_fishing	primary_oil_gas	primary_fishing	services_hospitality

Table 5. Top ten topics over-represented in high complexity (top 10%) TTWAs when compared to low complexity (bottom 50%) TTWAs for each complexity measure

	BRES ECI	BRES Fitness	IDBR ECI	IDBR Fitness
0	software, data, platform, communications, infrastructure, analysis, enterprise, implement	software, data, platform, communications, infrastructure, analysis, enterprise, implement	software, data, platform, communications, infrastructure, analysis, enterprise, implement	software, data, platform, communications, infrastructure, analysis, enterprise, implement
1	investment, growth, portfolio, finance, capital, invest, assets, investors	investment, growth, portfolio, finance, capital, invest, assets, investors	approach, understand, process, understanding relationships, effective, results, core	investment, growth, portfolio, finance, capital, invest, assets, investors
2	approach, understand, process, understanding, relationships, effective, results, core	approach, understand, process, understanding, relationships, effective, results, core	investment, growth, portfolio, finance, capital, invest, assets, investors	approach, understand process, understanding, relationships, effective, results, core
3	recruitment, talent, candidates, career, roles, search, executive, employers	creative, marketing, agency, media, content, agencies, social media, creativity	creative, marketing, agency, media, content, agencies, social media, creativity	creative, marketing, agency, media, content agencies, social media creativity
4	creative, marketing, agency, media, content, agencies, social media, creativity	focus, success, successful existing, manage, focussed, managed, challenges	recruitment, talent, candidates, career, roles, search, executive, employers	focus, success, successful, existing, manage, focussed, managed, challenges
5	focus, success, successful, existing, manage, focussed, managed, challenges	recruitment, talent, candidates, career, roles, search, executive, employers	focus, success, successful existing, manage, focussed, managed, challenges	recruitment, talent, candidates, career, roles search, executive, employers
6	businesses, network, strategy, strategic, strategies, insight, networks, connect	unique, create, bring, creating, passionate, passion, ideas, enhance	international, global, worldwide, countries, globe, globally, overseas, internationally	unique, create, bring, creating, passionate, passion, ideas, enhance
7	development, develop, partners, key, achieve, developing, partner, impact	international, global, worldwide, countries, globe, globally, overseas, internationally	unique, create, bring, creating, passionate, passion, ideas, enhance	international, global, worldwide, countries, globe, globally, overseas, internationally
8	deliver, leading, delivering, commitment, operate, operating, delivered, combined	development, develop, partners, key, achieve, development, partner, impact	development, develop, partners, key, achieve, developing, partner, impact	development, develop, partners, key, achieve, developing, partner, impact
9	unique, create, bring, creating, passionate, passion, ideas, enhance	value, professionals, long-term, provider, successfully, record, providers, objectives	expertise, enables, specific, ability, complex, highly, tailored, enabling	systems, innovative, solution, technical, industries, innovation, processes, operations

Consistent with previous observations, the dominant topics in TTWAs with high or low levels of complexity are largely independent of whether we use the ECI or Fitness index.

3.2 The geography of emergence in Great Britain

Having explored various indices and dimensions of economic complexity, we now turn to the analysis of emergence based on business website data.

We have identified 5,652 companies mentioning at least one of the emergent terms we highlighted in section 2.2. We map them in figure 3. Even after normalising by total business counts using IDBR data, this analysis shows that most of the TTWAs with a strong presence of emergent companies are in London and the South East of England. Creative and digital technology clusters such as Cambridge, Oxford, Guildford, Bristol and Brighton also rank highly in their 'emergence shares' (figure 3).

Table 6. Top ten topics under-represented in high complexity (top 10%) TTWAs when compared to low complexity (bottom 50%) TTWAs for each complexity measure

	BRES ECI	BRES Fitness	IDBR ECI	IDBR Fitness
0	road, room, park, village, town, hall, land, visitors	home, area, family, located England, serving, situated, south	road, room, park, village, town, hall, land, visitors	road, room, park, village town, hall, land, visitors
1	home, area, family, located, England, serving, situated, south	road, room, park, village, town, hall, land, visitors	home, area, family, located, England, serving, situated, south	home, area, family, located, England, serving, situated, south
2	local, year known, main, country, following, early, close	local, year, known, main, country, following, early, close	local, year, known, main, country, following, early, close	local, year, known, main, country, following, early, close
3	modern, traditional, Scotland, co, Manchester, owners, renowned, left	enjoy, hope, welcome, families, fun, friends, living, wonderful	modern, traditional, Scotland, co, Manchester, owners, renowned, left	enjoy, hope, welcome, families, fun, friends, living, wonderful
4	enjoy, hope, welcome, families, fun, friends, living, wonderful	day, happy, visit, feel, locationa, popular, away, stay	enjoy, hope, welcome, families, fun, friends, living, wonderful	day, happy, visit, feel location, popular, away, stay
5	day, happy, visit, feel, location, popular, away, stay	garden, accommodation, gardens, pet, animals, animal, dog, pets	day, happy, visit, feel, location, popular, away, stay	modern, traditional, Scotland, co, Manchester, owners, renowned, left
6	garden, accommodation, gardens, pet, animals, animal, dog, pets	original, old, began, plus, present, continued, added, moved	garden, accommodation, gardens, pet, animals, animal, dog, pets	garden, accommodation, gardens, pet, animals, animal, dog, pets
7	centre, open, history, city, opened, young, join, held	centre, open, history, city, opened, young, join, held	centre, open, history, city, opened, young, join, held	original, old, began, plus, present, continued, added, moved
8	original, old, began, plus, present, continued, added, moved	electrical, cleaning, air, waste, roof, clean, gas, renewable energy	original, old, began, plus, present, continued, added, moved	centre, open, history, city, opened, young, join, held
9	water, marine, sea, off-shore, aviation, boat, aircraft, crew	good, come, free, regular, little, regularly, week, days	electrical, cleaning, air, waste, roof, clean, gas, renewable energy	electrical, cleaning, air, waste, roof, clean, gas, renewable energy

3.3 Economic complexity and emergence

Figure 4 reveals an association between economic complexity and the levels of emergence in a location even after normalising our measure of emergence to control for differences in the economic composition of locations (that is, the fact that high complexity locations tend to specialise in digital and creative sectors that often involve emerging technologies). Our tentative conclusion is that the association between economic complexity and emergence is not fully explained by the industrial composition of high complexity TTWAs.

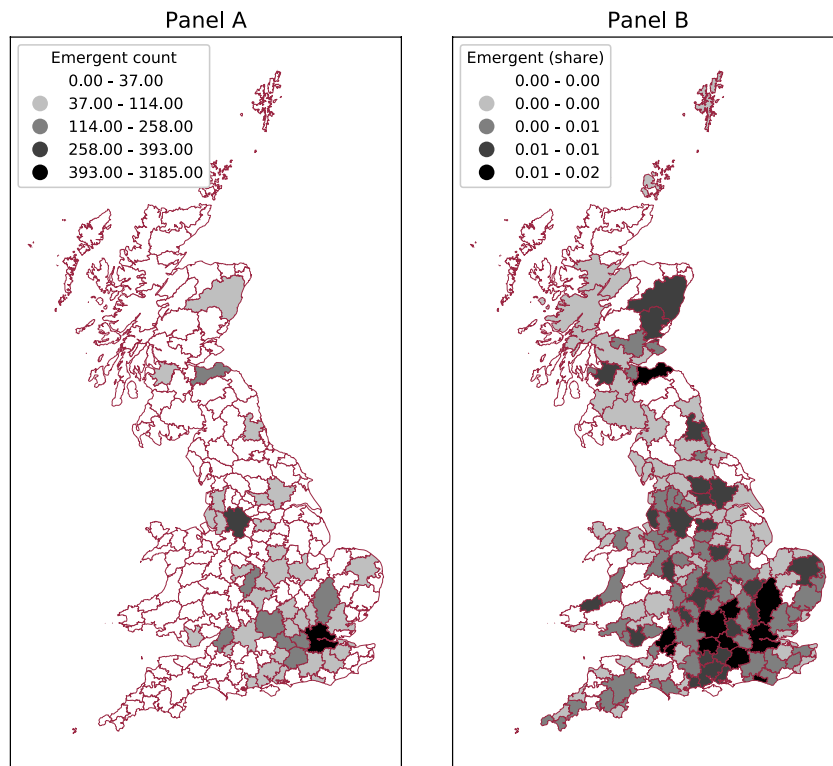
3.4 Synthesis

We conclude by considering our three key sets of variables (economic complexity, emergence and economic outcomes) together. Our goal is to explore the plausibility of emergence as a mechanism underpinning the relationship between economic complexity and economic outcomes. If this was the case, we would

expect to find a positive association between the three variables. For now, we simply calculate the Pearson correlation between the interaction of complexity and emergence, and economic outcomes. The results, which we present in figure 5, suggest the presence of this link: locations with high economic complexity and a strong presence of emergent companies also tend to have higher median earnings and GVA per capita.

However, and as noted before, this relationship is far from simple. In table 7 we show, for the top 50 British TTWAs by number of businesses in IDBR, whether they are above or below the median of all TTWAs in the interaction between economic complexity and emergence, and whether they are above or below the median in economic outcomes (in terms of median annual earnings). We see a significant number of TTWAs (particularly in the North) where high levels of economic complexity together with emergent economic activities

Figure 3. Geography of emergent activity in Great Britain



Note: Panel A displays the number of emergent companies identified in the Glass data and Panel B displays emergent companies as a share of all businesses (based on the IDBR data).

Figure 4. Correlation between complexity indices, specialisation in sectors with strong presence of emergent companies, and new normalised measure of emergence accounting for the presence of high-emergence activity sectors in a location

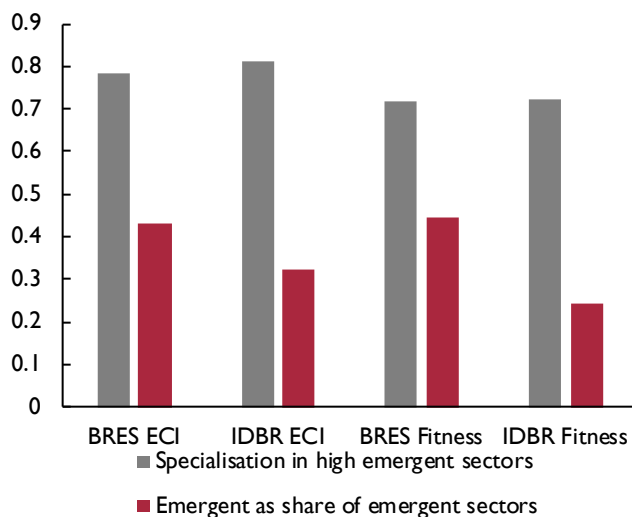


Figure 5. Correlation coefficient between interactions of economic complexity and emergence indices, and economic outcomes

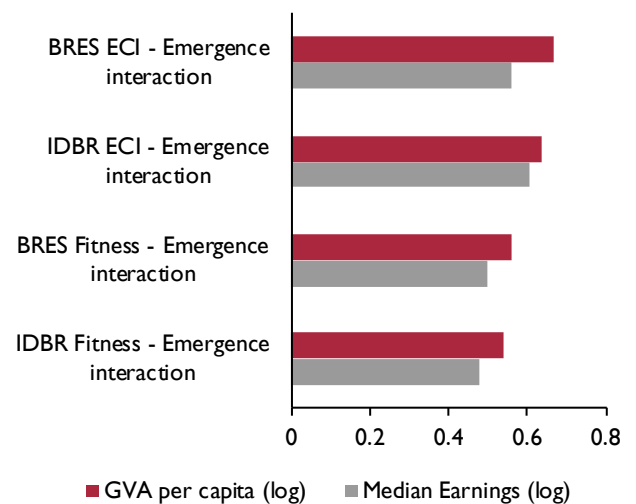


Table 7. TTWA position in the median earning and complexity*emergence distribution (top 50 TTWAs by number of businesses)

Complexity*Emergence below median	Complexity*Emergence above median
<i>Median earnings below median</i>	
Liskeard, Redruth and Truro, Falmouth, Wadebridge, St Austell and Newquay, Penzance, Launceston, Durham and Bishop Bude, Oswestry, Blyth and Ashington, Hull, Hexham, Hereford, Berwick.	Bradford, Huddersfield, Sheffield, Manchester, Liverpool, Wakefield and Castleford, Dudley, Cardiff, Doncaster, Auckland, Newcastle, Halifax, Sunderland.
<i>Median earnings above median</i>	
Shrewsbury, Bridlington, Fraserburgh, Peterhead, Turriff and Banff, Ludlow, Aberdeen, Fort William, Aviemore and Grantown-on-Spey, Ullapool, Alness and Invergordon, Thurso, Broadford and Kyle of Lochalsh, Wick, Portree, Golspie and Brora, Inverness, Birkenhead, Weston-super-Mare.	Leeds, Birmingham, Salisbury, Trowbridge, London, Crewe, Edinburgh, Bristol, Glasgow, Brighton, Milton Keynes, Swindon, Chester, Slough and Heathrow, Harrogate, Newbury, Bath, Motherwell and Airdrie, High Wycombe and Aylesbury, Bedford, Reading, Leamington Spa, Luton, Huntingdon, Andover, Northampton, York, Warrington and Wigan, Basingstoke.

are not (yet?) linked to improved economic outcomes. We also see some TTWAs where above median economic outcomes do not seem to be linked to high economic complexity/emergence. These include some rural areas, and perhaps notably, Aberdeen, a TTWA strongly reliant on economic activities linked to the oil industry.

4. Conclusion

We have mapped the economic complexity of UK TTWAs using two indices, ECI and Fitness, calculated using both employment data from BRES and establishment data from IDBR. Our results are broadly similar across complexity indices and activity measures: economic complexity appears in urban TTWAs, generally although not always in the South East of England, and involves knowledge-intensive, creative and digital sectors. High complexity areas show a stronger propensity to develop economic activities involving emergent technologies, a link that does not appear to be wholly driven by their tendency to specialise in 'complex' digital industries. We also show that areas with strong economic complexity and emergent activity also tend to experience better (higher productivity related) economic outcomes, consistent with (although of course not demonstrative of) the hypothesis that the generation of technological and industrial emergence is a mechanism through which economic complexity is transformed into local economic development.

We also identify some TTWAs outside of the South East of England where high levels of complexity and emergence do not correspond to superior economic outcomes. We hypothesise that these TTWAs could be missing other key resources required to generate impacts from their emergent activities, such as access to finance or markets – this is an area of high policy relevance deserving of further analysis.

Another preliminary policy implication from our analysis is that the development of economic complexity seems to rely on gaining specialisation in a set of knowledge intensive, digital 'vanguard' sectors that tend to cluster in a few locations as a consequence of agglomeration economies and other knowledge spillovers (Mateos-Garcia *et al.*, 2018). Policymakers seeking to develop their local economies should recognise that a policy of specialisation in such industries is risky and unlikely to generate benefits in the short term. It could also be detrimental for economic resilience against shocks concentrated on those sectors – economic complexity as measured here is, after all, not the same as diversity (Mealy *et al.*, 2018). The linking of novel data sources such as the Glass web data and official data make it possible to develop tools to aid policymakers in understanding the developing specialisation profiles of a local economy in a timely manner.

Our analysis is not without limitations. First, and most obviously, at this point we have simply described patterns of association between key variables. These need to be explored further within a multivariate modelling framework where we control for potential confounders such as population size, and workforce education & skills, also taking into account spatial autocorrelation between different locations. Second, there is the concern that our definition of emergence through CrunchBase data could skew our results towards creative and digital activities which are, as we mentioned, also over-represented in the Glass data. Generating measures of emergence which do not rely on ostensibly tech-oriented databases and sources, such as for example research publications and patents, is one potential avenue to address this limitation. Third, although our analysis suggests that technological emergence could be an important factor

underpinning the link between economic complexity and local economic development, the relation could go in the other direction – if wealthier areas are able to invest in novel technologies that make their economies more complex. Teasing causality within the network of interdependencies that drive local economic development will be another important and challenging avenue for further research. Unpicking the mechanisms that link economic complexity and emergence, such as knowledge spillovers, business networking and collaboration, labour flows and boundary-spanning entrepreneurship is one potential strategy to achieve this goal, and one where novel, highly detailed and timely, complex data sources such as those we have used in this paper have an important role to play.

NOTES

- 1 The sector classification has been trained using a sample of company classifications based on an industrial taxonomy created by LinkedIn.
- 2 We focus on business websites with less than five postcodes, where the main postcode tends to provide a valid match for geocoding.
- 3 Word2vec projects words in a corpus into a multidimensional space based on their co-location. The vectors for words which are semantically similar are located closer to each other in this space.
- 4 The ASHE data were also obtained from NOMIS. The GVA per capita data are available from <https://bit.ly/2krQBCs>.

REFERENCES

- Albeaik, S., Kaltenberg, M., Alsaleh, M. and Hidalgo, C.A. (2017), 'Improving the Economic Complexity Index', arXiv:1707.05826 [physics, q-n], July, <http://arxiv.org/abs/1707.05826>. arXiv: 1707.05826.
- Bakhshi, H. and Mateos-Garcia, J. (2016), 'New data for innovation policy', Working Paper, Nesta, London.
- Balland, P.-A. and Rigby, D. (2017), 'The geography of complex knowledge', *Economic Geography*, 93(1), January, <https://doi.org/10.1080/00130095.2016.1205947>.
- Breschi, S., Lassebie, J. and Menon, C. (2018), 'A portrait of innovative start-ups across countries'.
- Caldarelli, G., Cristelli, M., Gabrielli, A., Pietronero, L., Cala, A. and Tacchella, A. (2012), 'A network analysis of countries export flows: firm grounds for the building blocks of the economy', *PLOS ONE*, 7(10), pp.1–11, <https://doi.org/10.1371/journal.pone.0047278>.
- Dalle, J.-M., den Besten, M. and Menon, C. (2017), 'Using crunchbase for economic and managerial research'.
- Delgado, M., Porter, M.E. and Stern, S. (2015), 'Denying clusters of related industries', *Journal of Economic Geography*, 16(1), pp. 1–38.
- Einav, L. and Levin, J. (2014), 'Economics in the age of big data', *Science*, 346 (6210):1243089.
- Frenken, K., Van Oort, F. and Verburg, T. (2007), 'Related variety, unrelated variety and regional economic growth', *Regional Studies*, 41(5):685–97.
- Gentzkow, M., Kelly, B.T. and Taddy, M. (2017), 'Text as data', *Technical report*, National Bureau of Economic Research.
- Gerlach, M., Peixoto, T.P. and Altmann, E.G. (2018), 'A network approach to topic models', *Science Advances*, 4(7), <http://advances.sciencemag.org/content/4/7/eaq1360>.
- Hartmann, D., Guevara, M.R., Jara-Figueroa, C., Aristaran, M. and Hidalgo, C.A. (2017), 'Linking economic complexity, institutions, and income inequality', *World Development*, 93, pp. 75–93.
- Hidalgo, C.A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K. and Zhu, S. (2017), 'The principle of relatedness'.
- Hidalgo, C.A. and Hausmann, R. (2009), 'The building blocks of economic complexity', *Proceedings of the National Academy of Sciences*, 106(26), pp. 10570–5, June, <http://www.pnas.org/content/106/26/10570>.
- Hidalgo, C.A., Klinger, B., Barabasi, A.-L. and Hausmann, R. (2007), 'The product space conditions the development of nations', *Science*, 317(5837), pp. 482–7, <http://www.sciencemag.org/content/317/5837/482.short>.
- Klinger, J., Mateos-Garcia, J. and Stathoulopoulos, K. (2018), 'Deep learning, deep change? Mapping the development of the Artificial Intelligence General Purpose Technology', arXiv preprint arXiv:1808.06355.
- Mateos-Garcia, J., Klinger, J. and Stathoulopoulos, K. (2018), *Creative Nation: How the Creative Industries are Powering the UK's Nations and Regions*, London: Nesta.
- Mealy, P. and Coyle, D. (2019), 'Greater Manchester economic prosperity review: economic complexity analysis', Bennett Institute for Public Policy, University of Cambridge.
- Mealy, P., Farmer, J.D. and Teytelboym, A. (2017), *A New Interpretation of the Economic Complexity Index*, arXiv:1711.08245 [q-n], <http://arxiv.org/abs/1711.08245>. arXiv: 1711.08245.
- Nathan, M., Kemeny, T. and Almeer, B. (2017), 'Using crunchbase to explore innovative ecosystems in the US and UK'.
- Peixoto, T.P. (2017), 'Bayesian stochastic blockmodeling', <https://arxiv.org/abs/1705.10225>.
- Pugliese, E., Zaccaria, A. and Pietronero, L. (2016), 'On the convergence of the fitness-complexity algorithm', *The European Physical Journal Special Topics*, 225(10), pp. 1893–911, October, <https://doi.org/10.1140/epjst/e2015-50118-1>.
- Rotolo, D., Hicks, D. and Martin, B.R. (2015), 'What is an emerging technology?', *Research Policy*, 44(10), pp. 1827–43, 2015.
- Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A. and Pietronero, L. (2012), 'A new metrics for countries' fitness and products' complexity', *Scientific Reports*, 2:723.