Crowdfunding and industrial diversification

Nicola Cortinovis

Papers in Evolutionary Economic Geography

26.07

Utrecht University Human Geography and Planning

Crowdfunding and industrial diversification

Abstract

Crowdfunding (CF) has emerged as a novel source of entrepreneurial finance, yet its role in shaping regional industrial dynamics remains poorly understood. Adopting an evolutionary economic geography perspective and exploiting a newly developed database, this paper examines the relationship between crowdfunding activity and the emergence of new local industrial specializations. The analysis shows that industries receiving funds through CF are more likely to become part of local specialization patterns, especially when they are related to the existing industrial structure. Moreover, these associations are stronger in counties characterized by higher levels of credit insecurity.

Keywords: Crowdfunding, industrial diversification, relatedness, credit insecurity, US
JEL Codes: O14, O31

1. Introduction

The study of industrial diversification—i.e., the emergence of new industries in a location—has attracted significant attention in the economic geography literature in recent years. Scholars in evolutionary economic geography (EEG) have offered a number of contributions highlighting how the current industrial structure of a location plays a crucial role in shaping the emergence of new – typically related – industries (Hidalgo et al. 2018; Boschma 2017). The main mechanism behind the diversification process relies on entrepreneurship and cognitive proximity: local entrepreneurs and spinoffs leverage existing capabilities in combination with newly acquired ones to develop new products and services (Hidalgo et al. 2007; Neffke et al. 2018; Frenken and Boschma 2015). The success of these ventures leads to the emergence of new industries, in a process of regional branching (Boschma et al. 2017; Boschma and Frenken 2010).

This strand of literature has stimulated discussions in economic geography and related fields, which have offered additional contributions, stressing the importance of factors other than cognitive proximity. From the role of regional institutions (Cortinovis et al. 2017; Boschma and Capone 2015), local leadership and human agency (Beer et al. 2019; Grillitsch and Sotarauta 2020) and the presence of specific ecosystems (Content et al. 2022; Jacobs et al. 2014; Feldman 2023) to external linkages (Qiao et al. 2024; Cornejo Costas et al. 2025; Yeung 2021; Kogler et al. 2023), economic geographers have highlighted a number of factors shaping and affecting diversification dynamics. In this burgeoning literature, however, the role of (entrepreneurial) finance has been largely overlooked.

This limited knowledge of the relationship between access to capital and diversification is problematic for multiple reasons. First, to the extent that entrepreneurial efforts are essential for diversification (Hidalgo et al. 2007; Neffke et al. 2018), challenges and barriers, such as access to funding, faced by start-ups and entrepreneurs may hinder the emergence of new industries. Understanding how access to capital may affect diversification can clarify some of the microlevel mechanisms at play and help identify bottlenecks and opportunities (Feldman 2023). Second, the literature on the uneven geography of finance shows how access to capital remains unequal across space (Ughetto et al. 2019; Lee and Brown 2017; Brown et al. 2019; Hamdani et al. 2019). Structurally weak and peripheral regions are especially affected by limited financial resources (Kärnä and Stephan 2022; Fanjul et al. 2023), which may further lower their chances of diversification. Finally, in recent decades, significant innovations have taken place in the financial sector in general (Alessandrini et al. 2009), some of which have opened new opportunities in the provision of entrepreneurial finance (Bruton et al. 2015; Block et al. 2018). An important development in these respects has been the emergence of crowdfunding (CF) platforms as possible sources of capital for early-stage and niche entrepreneurs (Sorenson et al. 2016; Mollick and Robb 2016; Kim and Hann 2019).

Against this backdrop, this article aims to conceptually and empirically explore crowdfunding as an enabling factor in industrial diversification. While crowdfunding remains modest in terms of scale relative to other sources of entrepreneurial finance, its distinctive characteristics offer a particularly relevant perspective to examine how finance shapes regional diversification. Taking an evolutionary perspective, I connect the broader literature on diversification (Neffke et al. 2018; Boschma and Frenken 2010; Hidalgo et al. 2018; Neffke et al. 2011) with contributions in entrepreneurship and crowdfunding (Mollick and Robb 2016; Sorenson et al. 2016; Breznitz and Noonan 2020) to highlight how CF platforms may not only help address the financial constraints of firms (Cumming et al. 2021; Kim and Hann 2019; Eldridge et al. 2021) but also convey relevant insights, knowledge and connections crucial for entrepreneurs (Johan and Taylor 2019; Martínez-Climent et al. 2021; Acs et al. 2013). Drawing on these literatures, I conceptualize CF as an enabling factor – rather than a direct cause – behind industrial diversification and propose three main mechanisms. Empirically, I investigate the relationship between the provision of CF and the emergence of new industries in US counties leveraging a new database (KIUS 2.0) linking Kickstarter projects to a specific 3-digit NAICS industry, US county and year.

Three main contributions arise from this conceptual and empirical work. First, this paper contributes to the EEG by explicitly investigating the role of one type of entrepreneurial finance for local diversification (Clayton et al. 2024; Uhlbach et al. 2022). I find that the provision of resources through CF is associated with a higher probability for the targeted industry to become part of the portfolio of specializations of that location in the following three years. This suggest that access to entrepreneurial finance even through a source of funding typically smaller in size than banking or venture capital like CF can support local diversification processes (Feldman 2023; Binz et al. 2016). Second, my analysis reveals that the impact of the CF is strongest in industries related to the current industrial structure of the county, which supports the idea of related diversification (Boschma 2017; Boschma and Frenken 2011). Third, the impact of CF is especially strong in areas that are most affected by credit insecurity, in line with the literature on the uneven geography of finance (Ughetto et al. 2019; Lee and Brown 2017; Cowling et al. 2021). CF may thus be an important factor for unlocking diversification especially in areas facing lower access to capital.

The paper is structured as follows. The second section reviews the literature on EEG, links it to the extant contributions on crowdfunding and derives three testable hypotheses. The third section discusses the methodology and data development, whereas the fourth section presents the results of the baseline analysis and robustness checks. The last section draws conclusions and suggests avenues for further research. The reader interested in the details of the matching exercise and the data development and validation efforts is directed to the appendices.

2. Review of the literature

From an EEG perspective, the main mechanism behind the emergence of new industries in a given location relies on processes of discovery and innovation carried out by individuals and firms (Hidalgo et al. 2018; Boschma et al. 2017; Neffke et al. 2011; Esposito and Van Der Wouden 2022). The recombination dynamics of new and old knowledge foster the creation of new products and services and the emergence of new industries (Dam and Frenken 2022; Hidalgo et al. 2007; Hausmann et al. 2007; Frenken and Boschma 2015). The process of diversification can thus be thought of as a regional branching process in which entrepreneurs and firms innovate while remaining close to the industries already present in the local area to leverage preexisting knowledge and capabilities (Boschma and Frenken 2010).

While start-ups, spin-offs and firms are fundamental in the theory of regional branching (Neffke et al. 2018; Frenken and Boschma 2015), the role of financial resources in fuelling diversification has received only scant attention in the EEG literature. This is to some extent in contrast with related streams of research, which more explicitly recognize the role of capital resources in the emergence of products and technologies (Markard and Truffer 2008; Binz et al. 2016). One of the few contributions in EEG directly considering the role of finance in diversification is that of Uhlbach et al. (2022), which shows that public R&D spending contributes to the emergence of new technological specializations in European regions. Similar results have been obtained in relation to the presence of knowledge-intensive business services (KIBS), including banking and finance (Content et al. 2022). More recently, Clayton et al. (2024) studied the impacts of different sources of funding (VC, federal and state spending) on the survival of firms across the different development stages of the biotech cluster in the North Carolina system (Clayton et al. 2024). Interestingly, these sources of funding appear to have different impacts over time, but none of them affect firm survival in the nascent stage of the cluster. From this point of view, CF represents an interesting alternative, given its relevance for early-stage ventures and high-risk and niche projects (Mollick and Robb 2016; Bruton et al. 2015; Sorenson et al. 2016).

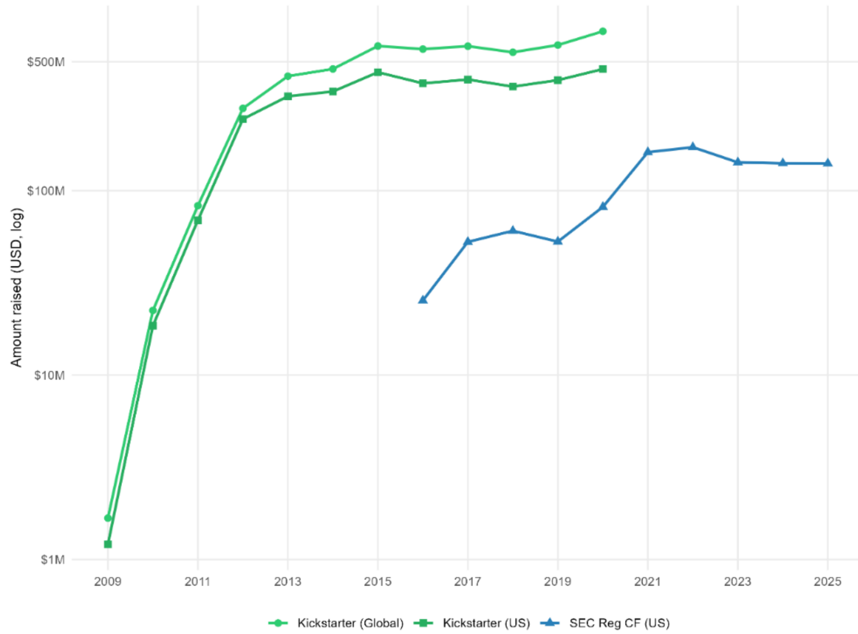
Crowdfunding: a short introduction

CF refers to a form of venture financing that uses internet platforms to gather financial contributions for investment purposes. This new type of financial instrument first emerged in the early 2000s and has gained significant popularity since then (Agrawal et al. 2014). As shown in Figure 1, the economic scale of crowdfunding has grown substantially over the past fifteen years. Kickstarter alone, one the main CF platforms and the main source of this analysis raised over $500 million globally in 2020, with US-based campaigns accounting for the large majority of projects on this platform[1]. Equity-based crowdfunding, based on information by the US Securities and Exchange Commission (SEC) (only available from 2016[2]), has surpassed $150 million annually by 2022, representing a rapidly growing alternative financing channel for early-stage ventures that remain underserved by traditional funding channels.

[1] The data on Kickstarter at the global level is provided and made publicly available (for the period 2009-2020) by Leland (2022). The data on Kickstarter in the US comes from the KIUS database (version 2.0).

[2] The data on equity-based crowdfunding offerings in the US is collected and made available by SEC on their website (https://www.sec.gov/data-research/sec-markets-data/crowdfunding-offerings-data-sets).

Figure 1: Trends in CF in the US and globally based on Kickstarter and SEC CF Offerings



Note: The y-axis is displayed on a logarithmic scale to accommodate the large range of values across series and years; labels are expressed in USD millions for readability.

The literature identifies four types of CF, with equity-based CF (in which investors receive some equity and future profits from the supported organization) and reward-based CF (in which backers receive an in-kind reward) being the most popular and more widely studied. Donation-based CF (where money is donated) and lending-based CF (in which investors simply expect to have their principal repaid) have instead received less attention (Agrawal et al. 2014; Kuppuswamy and Bayus 2015).

As mentioned above, CF presents three main characteristics that set it apart from traditional financing channels and make it particularly relevant for early-stage ventures and innovative yet niche ideas (Kuppuswamy and Bayus 2015; Moritz and Block 2016).

First, from a financial perspective, CF is a more cost-effective solution than banking or venture capital (VC) for small and young ventures due to its online accessibility, the use of nonfinancial rewards and other community-based benefits (Agrawal et al. 2014; Belleflamme et al. 2014). As funding decisions in CF are often based on the appeal of the product or additional nonfinancial rewards, the background, social or financial situation of the entrepreneur are less important. In this way, CF effectively creates funding opportunities for entrepreneurs who are typically disregarded by traditional financial channels (Cumming et al. 2021; Kim and Hann 2019; Mollick and Robb

2016; Macht and Weatherston 2014). Moreover, the ability to obtain funds through online platforms increases the chances of obtaining additional funds through other channels (Colombo and Shafi 2021; Hornuf et al. 2018; Macht and Weatherston 2014). Some studies show that successful CF campaigns help venture capitalists (Sorenson et al. 2016) and business angels (Yu et al. 2017) identifying business opportunities that would otherwise be unnoticed.

Second, CF provides not only capital but also other critical resources for entrepreneurs, particularly knowledge and information (Agrawal et al. 2014; Bargoni et al. 2022; Hervé and Schwienbacher 2018). Through CF platforms, entrepreneurs can actually test whether there is demand for their product or service and even contributing to generate such demand (Mollick and Kuppuswamy 2016; Mollick 2014). This is extremely helpful for early-stage entrepreneurs to either obtain market validation or quickly realize that there is not enough interest for their creation (Mollick 2014; Belleflamme et al. 2014; Macht and Weatherston 2014). Even more importantly, CF platforms open a communication channel between entrepreneurs and potential backers, allowing them to share ideas, provide suggestions and recombine knowledge (Hervé and Schwienbacher 2018; Martínez-Climent et al. 2021). For example, some functions and design improvements of the very first smart watch—Pebble, one of the most successful campaigns on Kickstarter—were put forward by backers and taken on by the creating team (Agrawal et al. 2014). As knowledge is often inherently embedded in geography and spatial context (Acs et al. 2013; Breschi 2001; Boschma 2017), start-ups or firms in emerging industries may lack the necessary knowledge or capabilities. Leveraging the knowledge and expertise of the crowd can thus provide critical insights and complement local knowledge assets.

Finally, CF platforms can help build and leverage network relationships. Scholars on CF often stress how these platforms are used by entrepreneurs both to activate connections within the CF communities (so-called "internal social capital") and to build and mobilize local relations ("external social capital") outside those of family and friends (Colombo et al. 2015; Giudici et al. 2018; Cai et al. 2021). These network connections are of paramount importance for entrepreneurs to identify new opportunities and acquire critical resources, not only in terms of finance and knowledge but also in terms of legitimacy, reputation and personal connections (De Luca et al. 2019; Wald et al. 2019). Leveraging these network relationships also allows entrepreneurs to build a customer base, identify suppliers and partners, increase public exposure and awareness and mobilize local stakeholders (Macht and Weatherston 2014; Cuvero et al. 2023; Sotarauta et al. 2025).

These theoretical arguments and observations have motivated some empirical studies attempting to measure and quantify the economic impact of CF. Mollick & Kuppuswamy (2016), one of the few studies examining the effect of reward-based CF campaigns, suggest that CF enables the creation of new business ventures and that most newly funded and preexisting companies survive at least the first 1-4 years after the campaign, with approximately one-third of them reporting revenues above 100,000 USD. In terms of employment, leading a successful campaign on average results in approximately 2.2 additional employees (Mollick & Kuppuswamy, 2016), with estimates connecting successful projects to an increase in approximately 5000 permanent employees and approximately 160,000 temporary ones (Mollick 2016). In the context of equity CF, Walthoff-Borm et al. (2018) focus on survival and innovation performance, finding that funded projects are associated with higher failure rates but also more patent applications. Interestingly, the conclusions of Eldridge et al. (2021) are somewhat different, indicating that equity-based CF does not influence innovation significantly but is positively associated with growth opportunities. As mentioned above, some scholars highlight how successful CF campaigns lead to an increase in the funding provided by business angels and venture capitalists in the same location (Sorenson et al. 2016; Yu et al. 2017).

Can crowdfunding foster industrial diversification?

While relevant for firms, it is not conceptually clear how CF may facilitate the process through which a new industry emerges locally. Diversification is a complex process driven and shaped by a multiplicity of forces, from the industrial-technological composition of the region (Boschma 2017; Dam and Frenken 2022; Antonietti and Montresor 2021) to institutions (Grillitsch and Sotarauta 2020; Cortinovis et al. 2017) and network-mediated external factors (Balland and Boschma 2021; Kogler et al. 2023). Against this backdrop, while CF is unlikely to represent an immediate and direct cause of diversification, some of its characteristics closely connect to factors highlighted in the literature as facilitating diversification dynamics.

First, the provision of entrepreneurial finance strongly relates to industrial diversification because of its relevance for innovative entrepreneurship. Small firms play a crucial role in innovation and industry emergence (Audretsch et al. 2020; Audretsch 2002; Jara-Figueroa et al.

2018; Neffke et al. 2018) but are often affected by limited financing opportunities (McKenzie and Woodruff 2008). As CF expands the pool of available financial resources, especially for start-up and small businesses (Cumming et al. 2021; Dunkley 2016; Mollick and Robb 2016), CF provides vital resources for new firms and supports the emergence and growth of new industries. The impact of CF resources may even be strengthened if resources from venture capitalists and business angels become available at a later stage (Sorenson et al. 2016; Yu et al. 2017; Butticè et al. 2017; Sewaid et al. 2026). These indications suggest the use of CF contributes to easing the investment constraints of entrepreneurs. As access to finance is one of the key resources for the emergence of a new industrial path (Binz et al. 2016; Clayton et al. 2024; Feldman 2023), these direct and indirect (through follow-up VC investments for instance) financial impacts of CF may provide important support for the emergence of new industries.

Second, CF helps mobilize and access knowledge and insights crucial for diversification. By interacting on online platforms, entrepreneurs can access new knowledge, discuss with users and fine-tune their ideas (Agrawal et al. 2014; Martínez-Climent et al. 2021). As these insights and knowledge may not be directly available in the local area, CF platforms may function akin to knowledge exchanges through interactions with experts (Bathelt and Gibson 2015; Bai et al. 2024; Chai and Freeman 2019; Painter et al. 2020). In addition, CF may mobilize additional cognitive resources and capabilities by stimulating and broadening the base of active entrepreneurs. By facilitating access to capital resources for entrepreneurs excluded by traditional financial channels (Mollick 2016; Walthoff-Borm et al. 2018), CF increases the pool of entrepreneurs and activates innovative ideas and cognitive capabilities available in the local area. Overall, CF can help foster diversification dynamics by facilitating the inflow of knowledge from elsewhere as well as mobilizing latent cognitive resources of "accidental entrepreneurs" who would typically struggle with funding their ventures (Mollick and Robb 2016; Cumming et al. 2021; Bao 2025).

Finally, the network connections generated through crowdfunding platforms can also facilitate diversification. By fostering the relational infrastructures that connect entrepreneurs, investors, and consumers, CF can promote the exchange of ideas, practices, and market signals. A combination of successful projects may help reveal, legitimize and exploit the available capabilities in local areas (De Luca et al. 2019; Wald et al. 2019), pushing other local entrepreneurs to further leverage them (Hausmann and Rodrik 2003). Moreover, network relations and the increased legitimacy of a successful CF campaign can attract the attention of relevant stakeholders—such as large companies, public investors and policy makers—mobilizing regional support systems that coordinate diversification efforts across different domains (Grillitsch and Sotarauta 2020; Beer et al. 2019; Binz et al. 2016; Sotarauta et al. 2025). From this perspective, the network relationships emerging from CF help connect, mobilize and better leverage local capabilities and institutions, increasing the capacity for structural change and diversification.

These different arguments support a possible role of CF in stimulating the emergence of new industries locally and clarify that CF should be better understood as an enabling factor rather than a direct cause of diversification. From this perspective, CF facilitates the processes of entrepreneurial discovery, fostering entrepreneurial dynamics (Hausmann and Rodrik 2003; Neffke et al. 2018) through the provision of capital, the mobilization of knowledge and the creation of network relations (Agrawal et al. 2014; Martínez-Climent et al. 2021; Macht and Weatherston 2014; De Luca et al. 2019). Through these mechanisms, CF helps remove bottlenecks and tap into capabilities and opportunities and thus facilitate the emergence of new activities locally. Because of the specific characteristics of CF (relatively smaller projects, geographically more dispersed), these theoretical mechanisms may be particularly relevant for areas characterized by a lower availability of capital, a limited presence of traditional financial institutions and providers and, more generally, poorer and more peripheral regions. Some recent evidence in this respect suggests that CF indeed stimulates entrepreneurial dynamics in rural areas (Cortinovis 2025).

Hypotheses

Building on the discussion thus far, I propose three testable hypotheses on the role of CF as a possible enabling factor in processes of industrial diversification. More specifically, the first hypothesis connects CF to industrial diversification through the three channels highlighted above, namely, the provision of entrepreneurial finance (Binz et al. 2016; Clayton et al. 2024; Feldman 2023), access to knowledge (Agrawal et al. 2014; Martínez-Climent et al. 2021; Balland and Boschma 2021) and the creation of new network relations (Grillitsch and Sotarauta 2020; Wald et al. 2019; De Luca et al. 2019). Based on these arguments, I hypothesize the following:

Hypothesis 1: The probability that a region specializes in a new industry is positively related to the industry receiving resources through crowdfunding.

Second, the EEG literature theorizes that regional development occurs through a process of regional branching (Boschma and Frenken 2010; Hidalgo et al. 2007), in which entrepreneurs innovate and recombine local capabilities, typically staying close to the industries in which the region is already specialized (Hidalgo et al. 2018; Frenken and Boschma 2015; Boschma et al. 2017). Given my characterization of CF as an enabling factor, I expect CF to facilitate diversification most in the case of industries related to the current local industrial structure (Boschma 2017; Boschma and Frenken 2011) and theorize the following:

Hypothesis 2: Relatedness density positively moderates the effect of receiving resources through crowdfunding on the probability that a region specializes in a new industry.

Finally, different contributions in the extant literature highlight how access to entrepreneurial finance is strongly affected by geography (Lee and Brown 2017; Ughetto et al. 2019; Kim and Hann 2019). As noted above, easing the budget constraints of local entrepreneurs via CF may enable the pursuit of projects that would otherwise remain unfunded through traditional channels. This effect is likely to be most pronounced in credit-insecure areas, where CF can partially substitute for missing or underdeveloped financial infrastructure and thus stimulate entrepreneurial activity and broaden the local economic base. To better characterize the impact of CF on diversification, I exploit the heterogeneity across space in access to capital resources to test whether the relationship between CF and diversification is stronger for areas that face greater credit insecurity than for regions for which access to credit is easier. Therefore, I hypothesize the following:

Hypothesis 3: The relationship between new industry specialization and successful crowdfunding campaigns is stronger for more credit insecure regions.

3. <u>Data and methodology</u>

<u>Data development</u>

One of the main limitations in the study of the impacts of CF, either at the project level or at the aggregate level, is the availability of data. To analyse the role of CF in industrial diversification, I developed a new database with exactly those specifications. The data (KIUS 2.0) has been developed using webscraped data from Kickstarter (as also done by Bao (2025)), which contains information on the time and location of approximately 170,000 CF campaigns. To complement the existing data with industry information, I rely on the large language model behind chatGPT 3.5. For every project, I provide the chatGPT API with the title, blurb (short introductory description of the project under the project's title) and the relevant category and subcategory in which the project is listed[3] and ask the algorithm to classify the project in one of the official codes of the 2017 NAICS classification. Overall, the data development effort provided information on the most relevant industry for approximately 170,000 individual projects. Appendix 2 provides more information on the data development and validation.

Additional data sources

The developed database linking projects to industries and including location (city) and year was aggregated to the level of 3-digit NAICS industry, county and year. In aggregating the data, I used information on the amount of money raised (pledge), the amount of money asked by creators (goal) and whether the goal has been reached to define two key variables: the number of successful large and very large campaigns (respectively defined as having raised either at least 5,000 USD or at least 50,000 USD) and the amount of money pledged to the local industry. I complemented the CF data with the County Business Patterns (CBP) database (Eckert et al. 2020), providing information about the number of employees, plants and annual payroll at the correct level of disaggregation. Additionally, I included data on venture projects by industry, county and year from the EIKON database. The final dataset includes industry, county, and year information for US counties for the period 2009-2016[4].

Construction of the variables

[3] Kickstarter groups projects in 15 categories (e.g. Comics, Crafts, Fashion, Technology) and more than 100 subcategories (e.g. within Technology: 3D printing, Apps, Fabrication tools, etc.). The creators of the projects are responsible for choosing the category and subcategory.

[4] While the database on Kickstarter projects covers the period from April 2009 (when Kickstarter was launched) to October 2023, the coverage of CBP is most complete only until 2016 (Eckert et al. 2020).

To study diversification, I use the standard approach in the EEG literature and estimate an entry model (Cornejo Costas et al. 2025; Steijn 2021; Boschma and Capone 2015), which I enrich, including main explanatory variable information on successful campaigns on Kickstarter.

Dependent variable

The starting point for building the dependent variable is to calculate in which industry each county specializes in a given year. I do this by computing a location quotient using the number of employees active in a certain industry and location. The location quotient is calculated as follows:

$$LQ_{i,c,t} = \frac{E_{i,c,t}/E_{*,c,t}}{E_{i,*,t}/E_{*,*,t}}$$

where $E_{i,c,t}$ represents the number of employees in industry i, county c, and year t, $E_{*,c,t}$ the total number of employees in county c, and year t, $E_{i,*,t}$ the number of employees in industry i in year t, in the whole county, and $E_{*,*,t}$ the total number of employees in the whole county in year t. Following (O'Donoghue and Gleave 2004; Tian 2013), I opt to use a standardized (log) location quotient and a bootstrapping approach to compute industry-specific thresholds to define specializations (Cortinovis et al. 2017). This is particularly important because using a threshold of 1 for the location quotient is particularly problematic with fine-grained data at the geographic and industry levels (Pominova et al. 2022). I follow the bootstrapping procedure presented by Tian (2013), taking the log of $LQ_{i,c,t}$ (to account for its skewed distribution) and standardizing it (subtracting its mean and dividing it by its standard deviation) to define $SLQ_{i,c,t}$. I then sample with replacement and generate 1000 samples for each industry to calculate the 90th percentile for each of the 1000 samples. Finally, I compute the threshold ($threshold_{i,t}$) above which specialization occurs by taking the average of the 90th percentiles across these 1000 samples. On the basis of this information, I define whether county c specializes in industry i in year t when $SLQ_{i,c,t}$ is greater than the industry-time specific threshold:

$$spec_{i,c,t} = \begin{cases} 1 \; if \; SLQ_{i,c,t} > threshold_{i,t} \\ 0, otherwise \end{cases}$$

To identify industry entries into counties, I exploit information on sectoral specializations, $spec_{i,c,t}$, defined as a binary variable equal to 1 if county c is specialized in industry i at time t and 0 otherwise. I detect entry events by comparing a county's specialization vector over a three-year window. The idea is that an industry is considered to have entered a county once that county becomes specialized in that industry, provided it was not specialized in it three years earlier. Formally, if $spec_{i,c,t} = 0$ and $spec_{i,c,t+3} = 1$, then industry i is coded as having entered county c at time t, and the entry variable $entry_{i,c,t}$ is set to 1. To illustrate, consider the specialization vectors of county c in 2010 and 2013:

$$spec_{i,c,2010} = \begin{matrix} 1 \\ 0, \\ 0 \end{matrix} \qquad spec_{i,c,2013} = \begin{matrix} 1 \\ 1 \\ 0 \end{matrix}$$

Comparing the two vectors, I define the variable entry for the year 2013 as follows:

$$entry_{i,c,2013} = \begin{matrix} NA \\ 1 \\ 0 \end{matrix}$$

The first industry is coded as *NA* because county *c* was already specialized in it in 2010, meaning that the entry already occurred prior to the observation window. The second industry receives a value of 1, indicating that the county became specialized in it between 2010 and 2013. The third industry receives a 0, as specialization was absent in both periods, and while entry was possible it did not occur. Importantly, once an industry has entered a county—that is, once specialization first appears—I assign *NA* to all subsequent years for that industry–county group to avoid counting the same entry multiple times if specialization persists.

Independent variables

The main explanatory variables I am interested in pertain, on the one hand, the provision of crowdfunding to a specific industry in a specific location and year and, on the other hand, a measure of relatedness density to assess how closely related the industry targeted by crowdfunding is with respect to the existing specializations of the location.

To study the exposure to CF for a specific industry-location-year group, I focus on the extensive margin and build a dummy variable taking value 1 when the industry has received support via CF.

To define my CF variable I follow Mollick (2014) and consider only successful projects which are relatively large (above 5000 USD for a single project). Based on my theoretical framework, the measure of CF should thus not be conceptualized as a simple cash injection of (at least) 5,000 USD[5], which by itself enables the transformation of an industry. Rather, this variable is a proxy for the presence of entrepreneurial and innovative projects of a minimal operational scale attempting to mobilize financial, cognitive and network resources. In this respect, choosing a relatively low threshold for identifying industry-county which have received CF allows me to fully exploit the granularity of my database. In formal terms

$$CF\ (L)_{i,c,t} = \begin{cases} 1\ if\ there\ is\ at\ least\ 1\ large\ successful\ KS\ campaign\ in\ i,c,t \\ 0,otherwise \end{cases}$$

Nonetheless, to avoid running the risk of setting a threshold too low, I have also built a similar variable ($CF\ (VL)$), considering projects that have raised at least 50,000 USD individually. Since for both these variables, it is not clear after how long the impact of a successful CF campaign may start materializing in the local industry and economy, I consider both $CF\ (L)$ and $CF\ (VL)$ as permanent (i.e. once an industry-location has received funding through Kickstarter, it will remain so for all the following years[6]). Finally, I also include additional measures related to the intensive margin of CF treatment, specifically the number of successful large campaign ($CF\ Projects\ (L)$) and the amount raised ($Amount\ raised\ (USD)$, from both large and small campaigns).

To capture the industry composition, I rely on the method developed by Hidalgo et al. (2007) and commonly applied in the evolutionary economic geography literature (Boschma and Capone 2015; Content et al. 2022; Cortinovis et al. 2017; Hidalgo et al. 2018). Using my measure of specializations developed above $spec_{i,c,t}$, I calculate the proximity index between each pair of industries on the basis of the co-occurrence of specializations in the same location (Balland 2017;

[5] It is important to stress that this crowdfunding variable is not intended to measure the magnitude of capital inflow, but rather the occurrence of entrepreneurial activity of at least a minimal operational scale. A threshold that is too high would risk excluding many genuinely entrepreneurial initiatives. By contrast, using a relatively low cutoff may introduce some “irrelevant” cases, but this would statistically weaken (rather than inflate) the estimated relationship.

[6] One of the robustness checks shows that removing this assumption and letting $CF\ (L)$ vary by year does not affect the results.

Steijn 2021; van Eck and Waltman 2009). The measure of relatedness is calculated as the ratio between the observed cooccurrences and a random benchmark:

$$\varphi_{i,j,t} = \frac{c_{i,j,t}}{\left[\left(\frac{S_{i,t}}{T_t}\right) * \left(\frac{S_{j,t}}{T_t - S_{i,t}}\right) + \left(\frac{S_{j,t}}{T_t}\right) * \left(\frac{S_{i,t}}{T_t - S_{j,t}}\right)\right] * \left(\frac{T_t}{2}\right)}$$

In the equation above, $c_{i,j,t}$ is the co-occurrence count of specializations in sectors i and j in year t, $S_{i,t}$ and $S_{j,t}$ are the total number of occurrences of each of the two industries, respectively, and T_t is the total number of occurrences of any sector. Essentially, $\varphi_{i,j,t}$ is the ratio between the number of counties in which the specializations in i and j simultaneously occur (nominator) and the number of counties in which I would expect the specializations to cooccur at random (denominator). Intuitively, the measure of relatedness captures the systematic co-occurrence of pairs of specializations. The values of $\varphi_{i,j}$ take a value of 1 when the co-occurrence found in the data is as frequent as at random, whereas it carries a higher (lower) weight when co-occurrence occurs more (less) frequently. I further modify the matrix Φ by setting the values of $\varphi_{i,j,t}$, which are equal to or less than 1, to 0 since their relatedness is coincidental. The matrix is then used to construct a density indicator (Hidalgo et al. 2007), capturing for industry *i*, in county *c*, at time *t*, how many of the theoretically possible industries related to *i* is county *c* already specialized in:

$$density_{i,c,t} = \frac{\sum_{k \neq i} \varphi_{i,k,t} * spec_{k,c,t}}{\sum_{k \neq i} \varphi_{i,k,t}}$$

<u>*Model and methodology*</u>

To study the relationship between the acquisition of a new industrial specialization and CF, I use a linear probability model:

$$entry_{i,c,t+3} = \beta * CF_{i,c,t} + \gamma * \boldsymbol{C}^1_{i,c,t} + \theta * \boldsymbol{C}^2_{c,t} + \xi_i + \pi_c + \tau_t + \varepsilon_c$$

where $entry_{i,c,t+3}$ is a dummy variable taking the value of 1 if the county becomes specialized in the targeted industry three years since it received resources through Kickstarter. The variable $CF_{i,c,t}$

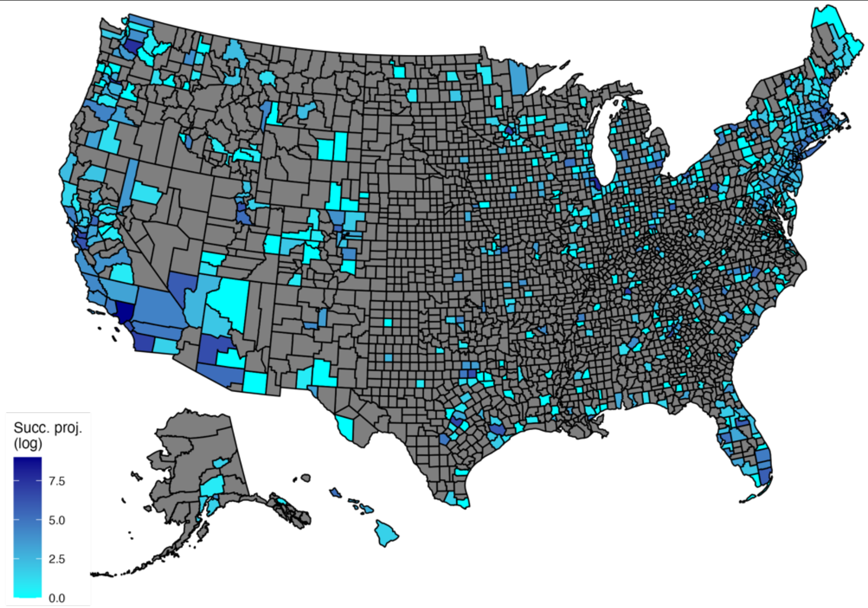
captures the receipt of resources through CF in the focal industry-county at time *t*. The two matrices $\boldsymbol{C}^1_{i,c,t}$ and $\boldsymbol{C}^2_{c,t}$ represent two sets of control variables. The first one, $\boldsymbol{C}^1_{i,c,t}$, varies by industry, county and year and includes a dummy for venture capital projects and a measure of relatedness density. With *VC (dummy)* variable I try to account for alternative measures of entrepreneurial financing, specifically venture capital. This is important since VC may follow CF and also facilitate the entry of a new industry. *Relatedness density* accounts for proximity to the current industrial structure of the county which the EEG literature stresses as a crucial driver of diversification (Boschma and Frenken 2010; Cortinovis et al. 2017). The second one, $\boldsymbol{C}^2_{c,t}$, includes variables varying by county and time only, such as GDP per capita, the total number of employees and the share of GDP accruing to financial activities in the county. The first two variables (*GDP (log)*, *Tot. empl (log)*) aim at controlling for the general economic performance of the area. *Fin. Share GDP*, computed as the share of GDP accruing to financial activities, controls for the local importance of the financial sector. Finally, my model includes a variety of fixed effects (at the industry (ξ_i), county (π_c) and year levels (τ_t)), and I cluster the errors at the county level (ε_c). This baseline model will be further enriched by an interaction term (for Hypothesis 2) and run on different subsamples (for Hypothesis 3).

4. Results

Descriptives

Figure 2 below shows the spatial distribution of projects across the US. The map confirms the patterns highlighted by previous scholars, with successful projects concentrated in larger agglomerations but also present in relatively small counties, such as the Midwest (Breznitz and Noonan 2020; Mollick 2014).

Figure 2: Spatial distribution of KS projects



More interestingly, my data reveal the distribution of CF projects across different industries. Figure 3 below provides information on the top 20 industries by the number of successful campaigns. Interestingly, the number of projects does not always correspond to the amount of funds received. For example, the industry gathering performing artists (711) is only fourth in terms of capital raised, with publishing industries (511), video industries (512) and miscellaneous manufacturing (339) having fewer projects but much higher pledges. More in general, the average pledged amount for successful projects in the period 2009-2016 is approximately 22,000 USD, with the median being approximately 5,100 USD.

Figure 3: Distribution of KS projects across 3-digit NAICS industries

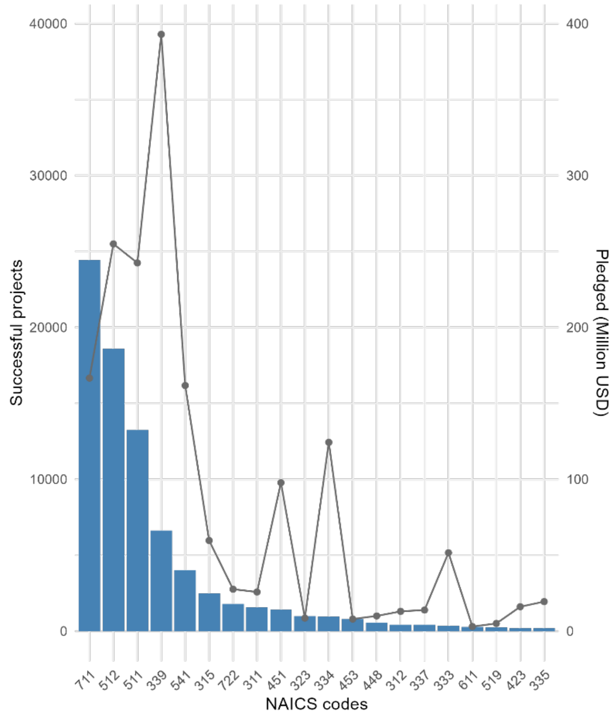


Table 1: Descriptive statistics					
Statistic	N	Mean	St. Dev.	Min	Max
Entry 3y	757,475	0.007	0.086	0	1
CF (L)	1,463,638	0.015	0.122	0	1
CF (VL)	1,463,638	0.003	0.059	0	1
CF Projects (L)	1,463,638	0.029	1.254	0	488
Amount raised	1,463,638	1,230.307	62,003.350	0.000	21,681,025.000
# Empl.	1,463,638	523.197	3,672.582	0	428,881
Relatedness density	1,463,638	5.937	6.924	0.000	87.080
Perm. VC	1,463,638	0.019	0.137	0	1
Tot. est.	1,463,638	3,189.706	9,719.385	1	268,594
Tot. emp.	1,463,638	49,563.940	160,463.300	1	4,006,031
Fin. Share GDP	1,437,683	0.208	0.083	0.000	0.784
GDP	1,437,683	6,209,156.000	23,981,808.000	4,040	575,689,171

Tables 1 and 2 provide a further overview of the descriptive statistics of the data and the correlations between the variables. Two main points emerge from the descriptive statistics. First, the number of observations for which I have non-missing values for entries (*Entry 3y*) is a relatively small portion of the observations. This is due to the stringent strategy in defining specializations using the top 10 percentile and avoiding counting entries multiple times (assigning *NA* after the first entry). In this way, county-industry pairs that have already achieved specialization are excluded by design and the estimation sample focuses on observations where entry remains possible[7]. Second, successful and large crowdfunding campaigns (*CF (L)* and *CF (VL))* are relatively rare. Alternative measures of CF *(# CF Projects*, *Amount raised*) confirm that the distribution of CF is rather skewed. Focusing now on the correlation table, it is interesting to note how entries and crowdfunding appear to be pretty much uncorrelated.

Table 2: Correlation table

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Entry 3y	(1)	1	.	.	.	.	.	.	.	.	.	.	.
CF (L)	(2)	.00	1	.	.	.	.	.	.	.	.	.	.
CF (VL)	(3)	.00	.47	1	.	.	.	.	.	.	.	.	.
CF Projects (L)	(4)	.01	.19	.26	1	.	.	.	.	.	.	.	.
Amount raised	(5)	.00	.16	.29	.65	1	.	.	.	.	.	.	.
Empl.	(6)	-.00	.14	.13	.14	.14	1	.	.	.	.	.	.
Relatedness density	(7)	.05	.03	.03	.03	.02	.07	1	.	.	.	.	.
Perm. VC	(8)	-.00	.21	.20	.09	.10	.27	.07	1	.	.	.	.
Tot. est.	(9)	-.01	.20	.18	.15	.14	.49	.01	.36	1	.	.	.
Tot. emp.	(10)	-.01	.21	.19	.15	.14	.50	.02	.38	.98	1	.	.
Fin. Share GDP	(11)	.01	.04	.02	.01	.01	.04	.04	.05	.10	.08	1	.
GDP	(12)	-.01	.20	.19	.17	.15	.49	.03	.38	.94	.97	.07	1

<u>Analysis on local economic outcomes</u>

Before testing the relationship between CF and diversification, it is important to evaluate whether CF is associated with relevant outcome variables, such as the number of jobs and

[7] Missing values for *Entry 3y* are concentrated in county-industry cells with higher relatedness density and employment, consistent with cases where the focal industry is already part of the local specialization portfolio and thus excluded by design. The mean density (4.96, 6.99) and mean employment (428, 626) differ slightly across non-missing and missing observations, but this is expected given the sample construction and does not introduce systematic bias into the estimation.

establishments in the targeted industry and county in the following years. To this end, I estimate the relationship between the amount of CF received and different outcome variables via a fixed effect panel regression[8]. The analysis suggests that industries receiving CF typically see an increase in the number of jobs and plants active in the industry-county group. In particular, the estimates reported in Table 3 indicate that receiving an additional 10,000 through CF is associated with an increase in the number of jobs, from approximately 3 after one year to almost 13 after three years. This result appears relatively high and is likely affected by selection and sorting effects (e.g., if successful CF campaigns are located in relatively large industries). Considering the case of establishments, the impact of crowdfunding is positive overall. On average, on the basis of the point estimates, raising between 50 and 60,000 USD leads to one extra employment-generating plant in the same industry-county three years later. Interestingly, comparing the size of the coefficients for the total number of plants (column 2) and plants with fewer than 5 employees (column 3) suggests that CF may impact not only micro-firms but also larger ones. While giving just a broad indication, this basic analysis provides some interesting insights, with positive relationships between CF and different sectoral outcomes suggesting that crowdfunding may function as a local entrepreneurial stimulus. As the theoretical mechanism behind diversification conceptually relies on entrepreneurial dynamics, the findings of this preliminary analysis are in line with a possible role for CF in fostering diversification.

[8] This exercise should be taken with some caution and not interpreted as a thorough benchmarking exercise but rather as offering an indication of the association between CF and economic outcomes at the sectoral level. Besides, in this model I exploit the variation within industry-county groups over time. In the main regressions, because entry of a new industry can occur only once, the fixed effects included are by industry, county and year individually. This choice is in line with previous studies (Cortinovis et al. 2017; Steijn et al. 2023; Boschma and Capone 2015).

Table 3: Crowdfunding and regional economic outcomes			
	Number of employees	Number of establishments	Number of micro establishments (<5 employees)
Amount raised (USD 10,000) $_{t-1}$	3.185+	0.082+	0.050
	(1.742)	(0.044)	(0.032)
Amount raised (USD 10,000) $_{t-2}$	8.326*	0.099**	0.074**
	(4.082)	(0.033)	(0.026)
Amount raised (USD 10,000) $_{t-3}$	12.817*	0.198**	0.161**
	(5.894)	(0.066)	(0.057)
Num.Obs.	873662	873662	873662
R2 Adj.	0.988	0.998	0.997
R2 Within Adj.	0.027	0.008	0.008
NAICS3 X County FE	X	X	X
Year FE	X	X	X
Clustered SE	County	County	County

Clustered standard errors in parenthesis
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Regression analysis

The results of the estimation of the regression model discussed above are reported in Table 4 below. The estimated coefficients suggest that the relationship between crowdfunding and entries is positive and significant. This suggests that, as funds and information are made available to entrepreneurs in a given industry and region through CF, the probability of the target industry entering the portfolio of specialization of the county increases. For example, receiving a large CF project (*CF (L)*) is associated with an increase in the probability of the industry entering of 0.003 (0.3 percentage points). This point estimate, arguably rather small, is nonetheless in line with previous studies in the literature[9] and it is somewhat expected, given the rarity of entry events.

[9] For example, Steijn et al. (2023), one of the few studies focusing on technological diversification in the US from an EEG perspective, find that major economic crisis reduces entry dynamics by only 0.032 percentage points. In the case of Europe, Uhlbach and Balland (2022) find similarly that the impact of participating in a Framework programme is relatively small: when including no control variables, a 10% increase of project participation is associated with 2% (relative to the mean) increase in the probability of entry.

Table 4: Baseline regressions					
	Entry	Entry	Entry	Entry	Entry
CF (L)		0.003**			
		(0.001)			
CF (VL)			0.006+		
			(0.004)		
CF Projects (L)				0.001+	
				(0.000)	
Amount raised (USD 10,000, log+1)					0.003*
					(0.001)
Tot. Empl. (log)	0.003	0.003	0.003	0.003	0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Relatedness density	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Fin. Share GDP	-0.012	-0.012	-0.012	-0.012	-0.012
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
GDP (log)	-0.004**	-0.004**	-0.004**	-0.004**	-0.004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Perm. VC (dummy)	0.002**	0.002**	0.002**	0.002*	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Num.Obs.	744323	744323	744323	744323	744323
R2 Adj.	0.011	0.011	0.011	0.011	0.011
R2 Within Adj.	0.001	0.001	0.001	0.001	0.001
County FE	X	X	X	X	X
NAICS3 FE	X	X	X	X	X
Year FE	X	X	X	X	X
Clustered SE	County	County	County	County	County

Clustered standard errors in parenthesis
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

As shown in Table 4, my findings are robust to the specific measure used to measure CF, whether consider only very large (at least 50,000 USD) crowdfunding projects (*CF (VL)*) or whether I use measures capturing the intensity of CF (*CF projects (L)*, *Amount raised*). These findings support Hypothesis 1, for which the provision of capital through CF is positively associated with industry entry. Interestingly, the relatedness between the focal industry and the current industrial structure plays a role in fostering entry, confirming the findings in the extant evolutionary economic geography literature (Boschma 2017; Boschma and Capone 2015; Content et al. 2022; Cortinovis et al. 2017).

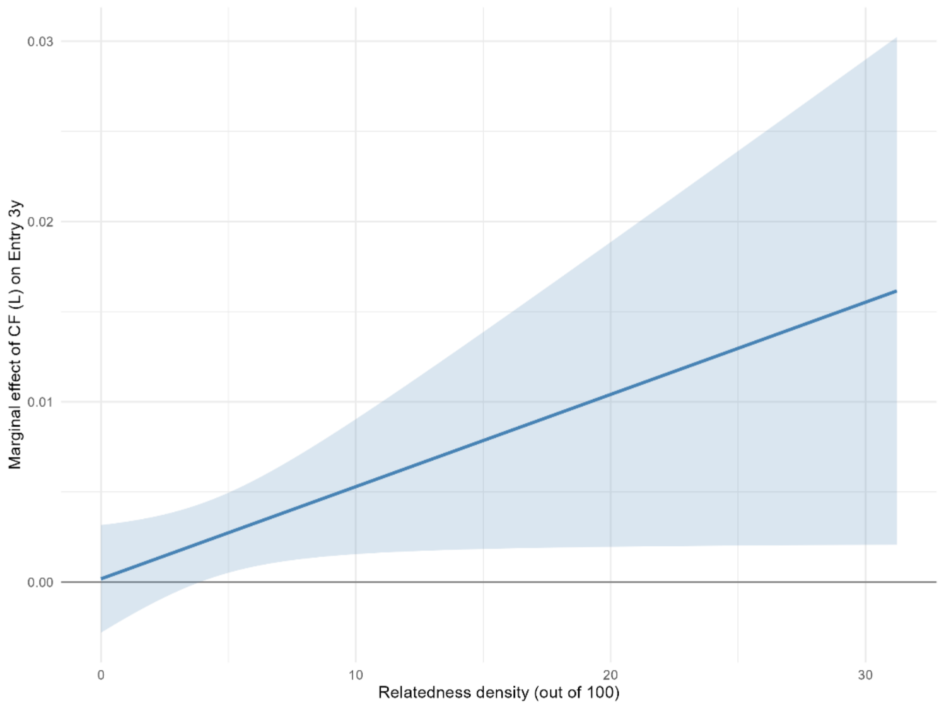
Table 5:Regressions with interaction terms				
	Entry	Entry	Entry	Entry
CF (L)	0.000			
	(0.001)			
CF (VL)		-0.004		
		(0.003)		
CF Projects (L)			0.001	
			(0.001)	
Amount raised (USD 10,000, log+1)				-0.001
				(0.001)
Relatedness density	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
CF (L) * Rel. density	0.001+			
	(0.000)			
CF (VL) * Rel. density		0.001+		
		(0.001)		
CF Projects (L) * Rel. density			0.000	
			(0.000)	
Amount raised (USD 10,000, log+1)* Rel. density				0.000*
				(0.000)
Num.Obs.	744323	744323	744323	744323
R2 Adj.	0.011	0.011	0.011	0.011
R2 Within Adj.	0.001	0.001	0.001	0.001
Control variables	X	X	X	X
County FE	X	X	X	X
NAICS3 FE	X	X	X	X
Year FE	County	County	County	County
Clustered SE	744323	744323	744323	744323

Clustered standard errors in parenthesis

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

When focusing on the role of relatedness density in strengthening or weakening the effect of CF, my findings, reported in Table 5, provide further insights. The direct positive relationship between successful CF campaigns and industrial diversification is no longer significant, and the interaction terms with relatedness density fully takes up the relationship between CF and industrial entry. The positive and significant interaction terms in Table 5 indicate that the impact of CF on diversification is magnified when the industry receiving CF is relatively well connected to other specializations in the county. Figure 4 below shows the marginal effects of *CF (L)* at different levels of relatedness density and suggests the impact of CF becomes positive and significantly different from zero when the level of relatedness density is around the median level (3.80). This finding confirms Hypothesis 2, which is based on the regional branching argument (Boschma 2017; Boschma and Frenken 2011), and suggests that provision crowdfunding to industries too weakly related to the local industry may not be sufficient to spur diversification.

Figure 4: Marginal effect of CF (L) on Entry 3y across levels of relatedness density



Access to credit, CF and diversification

To test Hypothesis 3, I study the relationship between CF and the industrial diversification of US counties across different subsamples, defined on the basis of their level of credit insecurity. If CF can facilitate the emergence of a new industry, it is likely that it will play an especially important role in areas characterized by limited credit access (Cumming et al. 2021; Lee and Brown 2017; Kim and Hann 2019). To explore heterogeneity across different levels of credit insecurity, I use data from credit access insecurity from the Federal Reserve Bank of New York (Hamdani et al. 2019), which groups US counties into five different categories on the basis of the availability of credit to their residents across three years (2007, 2012 and 2017). Using information for 2007 and 2012, I group counties into three groups[10] and compare the relationship between CF and entry across areas at the highest and lowest levels of credit constraints. In Table 6, I focus only on *CF (L)* and *CF (VL)* and compare the results across these two definitions of CF.

Overall, the findings concerning credit insecurity offer some new insights. First, compared with the first three columns of Table 6, CF is significantly and positively associated with entry only in areas subject to high credit insecurity (first column, high CI). The point estimate of the coefficient for *CF (L)* is also larger than the one in Table 3 (0.006 vs 0.003), suggesting a stronger relation in this subsample. This suggests that CF plays a role in fostering industry entry for those areas that face the most difficulties in obtaining credit. In contrast, the same variable does not produce a statistically significant result for areas facing low credit insecurity (second column, low CI). Introducing the interaction terms between CF and relatedness density (second two columns in Table 6) confirms that CF is directly related to entry in high-CI counties and is not mediated by relatedness density (third column). These characteristics may help explaining this finding, with the strength of the direct effects supporting the idea of CF easing capital constraints in these communities (Kim and Hann 2019). As highlighted by Hamdani et al. (2019), counties facing high credit insecurity are typically economically weaker (e.g., higher poverty rate, lower levels of educational attainment, lower population and more rural) and characterized by a greater share of demographic groups (e.g., African-American and Hispanic) usually ignored by funding and relying more on CF (Mollick and Robb 2016). This finding suggests that the idea of CF possibly playing a role in reducing territorial disparities. Moreover, the nonsignificant interaction term may be connected to the limited scope of relatedness in these counties, a compatible explanation with economically less dynamic and structurally weaker regions (Hamdani et al. 2019). Interestingly, when looking at counties facing low credit insecurity (fourth column in Table 6), I find the opposite result, namely, that CF is positively related to entry only through relatedness density.

When considering larger CF projects (*CF (VL)*), the results are somewhat different. The results for counties facing high credit insecurity do not produce a statistically significant coefficient, either directly (column 5) or in interaction with relatedness (column 7). Columns 6 and 8, pertaining to areas in the group characterized by low credit insecurity, instead overall confirm the previous

[10] Hamdani et al. (2019) leverages individual credit information and aggregates it at the local level. This approach allows them to divide US counties in five tiers: credit-assured, credit-likely, mid-tier, credit-at-risk and credit-insecure. I aggregate the tiers into three groups: high credit insecurity (credit-insecure and credit-at-risk tiers), medium credit insecurity (mid-tier) and low credit insecurity (credit-assured and credit likely tiers).

results. In those counties, the provision of crowdfunding is not associated with diversification directly but only when industries related to the current industrial structure of the area are targeted. Interestingly, the presence of very large projects in industries that are only weakly related to local specializations seems to reduce—rather than increase—the probability of entry in those industries in counties facing low credit insecurity. Overall, the results presented in Table 6 support Hypothesis 3, as CF appears to have a stronger impact on diversification when firms and individuals face high credit insecurity.

Table 6:Regressions across different levels of Credit Insecurity (CI)

	High CI - Entry	Low CI - Entry	High CI - Entry	Low CI - Entry	High CI - Entry	Low CI - Entry	High CI - Entry	Low CI - Entry
CF (L)	0.006*	0.002	0.005*	-0.003				
	(0.002)	(0.002)	(0.003)	(0.003)				
CF (VL)					0.004	0.010	0.002	-0.012+
					(0.005)	(0.007)	(0.005)	(0.007)
Relatedness density	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
CF (L) *Density			0.000	0.001+				
			(0.000)	(0.000)				
CF (VL) *Density							0.000	0.003+
							(0.001)	(0.002)
Num.Obs.	249756	322486	249756	322486	249756	322486	249756	322486
R2 Adj.	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009
R2 Within Adj.	0.001	0.002	0.001	0.002	0.001	0.002	0.001	0.002
Control variables	X	X	X	X	X	X	X	X
County FE	X	X	X	X	X	X	X	X
NAICS3 FE	X	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X	X
Clustered SE	County	County	County	County	County	County	County	County

Clustered standard errors in parenthesis

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Matching, event study and two-stage least squares

One important consideration is that the observed relationship between CF and entry may reflect differences in local or sectoral characteristics rather than a direct effect of crowdfunding. For example, successful CF campaigns might be more frequent in industries or areas with greater structural dynamism or diversification potential. Given the spatial distribution of the CF (Breznitz and Noonan 2020), this is a valid concern. To address this point, I use a matching algorithm to build a sample of comparable "treated" (for which *CF (L)* is equal to 1 at some point in the period of analysis) and "control" industry-county groups are truly comparable. The matching exercise, which uses both propensity score and exact matching, is carried out using information on the period before the treatment started (2005-2008) and is based on the same variables used as controls, with the addition of the number of plants in the industry-county-year group and the total number of plants by county and year. For the exact part of the matching, I condition the pair of treated and untreated units to belong to the same US state but different counties[11]. Appendix 1 shows the overall good performance of the matching exercise, which reduces the difference in the means between treated and control units. The propensity score matching leads to a reduction in the sample size, which now consists of approximately 64,000 observations. Using this balanced sample for the analysis, I estimate again the baseline model and the one including interaction effects. The results are presented in the first two columns of Table 7 and provide further support for the previous findings discussed in Tables 3 and 4.

Table 7: Regressions on matched samples and two-stage least squares

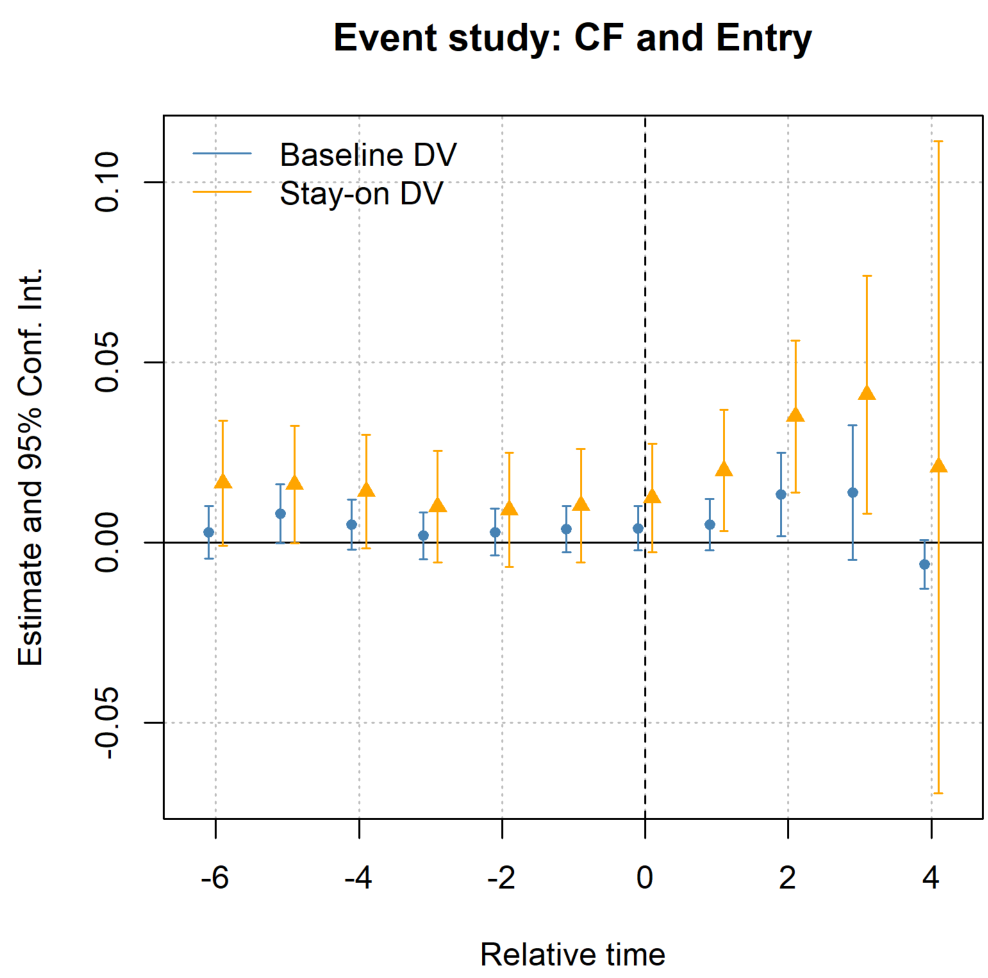
	Matched sample	Matched sample	IV1	IV2	IV1 and IV2
CF (L)	0.003*	-0.001	0.052**	0.044*	0.047***
	(0.002)	(0.002)	(0.019)	(0.018)	(0.014)
Relatedness density	0.000***	0.000***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CF (L) *Density		0.001*			
		(0.000)			
Num.Obs.	64219	64219	743795	744320	743795
R2 Adj.	0.023	0.023	0.010	0.011	0.010
R2 Within Adj.	0.001	0.002	0.001	0.001	0.001
Control variables	X	X	X	X	X
County FE	X	X	X	X	X
NAICS3 FE	X	X	X	X	X
Year FE	X	X	X	X	X
Clustered SE	County	County	County	County	County
First-stage F			1871.65	3626.17	2746.85
Sargan-Hansen stat.					0.081
Sargan-Hansen p-value					0.776

Clustered standard errors in parenthesis
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[11] I have tested different alternative in the matching strategy (e.g. conditioning on belonging to the same county or same 2-digit industry and state) which provide comparable results. The preference for this specific setup for the matching is motivated by the better balance achieved, as reported in Appendix 1.

Building on the matched sample analysis, I also examine the temporal patterns of entry relative to those of CF via an event-study framework. I follow a two-stage DiD approach (Butts and Gardner 2022; Gardner 2022). Since I am not exploiting the within-unit variation, this event-study simply illustrates how entry evolves around the occurrence of crowdfunding activity, while accounting for structural differences across counties, industries, and years. In addition, the approach chosen for the baseline dependent variable (setting value to 1 in the year when entry occurs and to NA in all subsequent years) mechanically reduce the number of observations in later years and bias the estimated coefficients downwards. Therefore, I also estimate the event study under a "stay-on" assumption, where the dependent variable remains equal to 1 in all years following entry. The findings are summarized in Figure 5 below, where the blue line represents the baseline dependent variable and the orange one the stay-on dependent variable. Two main points emerge from the event-study analysis. The first is a confirmation that, after receiving support through CF, the probability of acquiring a new specialization in the target industry significantly increases, in particular between 1 and 3 years after the campaign (with the baseline dependent variable being significant only for the second year after the campaign). The second pertains to the magnitude of the coefficients, which, in both event studies, are markedly larger than those in the baseline models. The likely reason is that, by summarizing the relationship between CF and entry in one single coefficient over time, the baseline dilutes the stronger association that the event study detects in certain years. On the basis of the event study estimates, the probability of entry increases between 1.3% (with the more baseline approach) and almost 4% (under the stay-on assumption).

Figure 5: Event study plot



I further check the validity of my findings using use two instrumental variables, which are presented in last three columns of Table 7. The in-depth discussion of the construction of the instruments is included in the Appendix 1. The first instrument follows the approach by Sorenson et al. (2016). Specifically, I use information on successful CF campaigns in the same county and same year but in sectors that are unrelated to the focal one. Focusing on unrelated industries ($\varphi_{i,j} < 0.5$, so less related than at random, according to the relatedness measure (see the section *Construction of the variables*)) ensures the exogeneity of the instrument. The second instrument instead uses information on successful projects in the focal industry and year at state-level to predict the presence of successful CF campaigns. To ensure the exclusion restriction holds in this second instrument, I exclude the focal county as well as the adjacent counties from the count of successful projects at state level. Both instruments leverage the presence of relevant information on crowdfunding and the legitimacy of the crowdfunding ecosystem at different levels, which should facilitate an entrepreneur in gathering resources on CF without directly affecting the probability of entry in the focal industry (since the information comes either from unrelated local sectors (instrument 1) or from campaigns in the same sector but located further away (instrument 2)). The coefficients for the two instruments in the first stage regressions (Table 10 in Appendix 1) are positive and highly significant. As reported in the last three columns of Table 7 at the bottom of the table, the IV approach appears to be valid: the first-stage F statistics are well above 10 and the Sargan-Hansen test fails to reject the null hypothesis of the validity of instruments. Moving to the variable of interest, the estimated coefficients for *CF (L)* fully support my previous results, being positive and significant. Interestingly, the point estimates are larger than those of the baseline and similar in size to the coefficients of the event study, confirming the point of the event study about the baseline results being downward biased.

Additional robustness checks

To test the validity of my findings, I carry out several other robustness checks, particularly focusing on three additional issues. First, I consider whether considering the impact of crowdfunding as permanent (so that all the years after a successful large project are considered affected by CF) may affect the results. Using a time-varying definition of CF *(CF (L, varying))* does not affect the findings reported in the first two columns of Table 8. Second, I test whether my findings are robust to studying entry by nonoverlapping periods[12]. The results reported in the second two columns of Table 8 are based on two nonoverlapping periods (2009-2012, 2013-2016) and confirm the previous findings on the positive association between CF and entry into a new industry in the county. Similarly, I explore whether the results are robust when a longer time window is used to define industry entry. In the fifth and sixth columns of Table 6, I take the whole period for defining entries, i.e., comparing specializations in 2009 with 2016[13]. The inclusion of county fixed effects implies that, in this specific setting, control variables vary by county-year drop

[12] To maximize the number of observations, the analysis has thus far used overlapping time periods, essentially comparing specialization on the basis of a moving time window. Essentially, entry is assessed by comparing specializations present in a region in 2009 versus 2012, then 2010 vs 2013, 2011 vs. 2014, etc.), following Cortinovis et al. (2017; 2024).

[13] The measure of exposure is computed considering the whole period (e.g., an industry-county with the first large successful campaign in 2014 will be considered treated from 2014 onwards)

out. Nevertheless, the results concerning the relationship between CF and entry are robust and exhibit even larger coefficients than the baseline. Lastly, I check whether using a logit specification rather than a linear probability model to account for the binary nature of the dependent variable influences the results. Changing the functional form of the model does not affect the findings.

Table 8: Additional robustness checks on variable definition and model specification

	Time varying CF	Time varying CF	No overlap	No overlap	Entry - 2009/2016	Entry - 2009/2016	Logit model	Logit model
CF (L, varying)	0.004*	-0.002						
	(0.001)	(0.002)						
CF (L)			0.006*	-0.002			0.619***	0.190
			(0.003)	(0.004)			(0.156)	(0.246)
Relatedness density	0.001***	0.001***	0.002***	0.002***	0.004***	0.003***	0.058***	0.057***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)
CF (L, varying) *Density		0.001*						
		(0.000)						
CF (L) *Density				0.001*				0.042**
				(0.001)				(0.016)
LP CF					0.008**	0.000		
					(0.003)	(0.003)		
LP CF *Density						0.001*		
						(0.001)		
Num.Obs.	744323	744323	327211	327211	163732	163732	590761	590761
R2 Adj.	0.011	0.011	0.038	0.038	0.041	0.042	0.008	0.008
R2 Within Adj.	0.001	0.001	0.006	0.007	0.012	0.012	0.014	0.014
Control variables	X	X	X	X	X	X	X	X
County FE	X	X	X	X	X	X	X	X
NAICS3 FE	X	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X	X
Clustered SE	County	County	County	County	County	County	County	County

Clustered standard errors in parenthesis
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

5. Conclusions

The literature on industrial diversification has shed light on how regional economies evolve, mainly emphasizing the role of entrepreneurship and cognitive proximity. New industries emerge when entrepreneurs recombine existing local capabilities with new ones. Yet, the role of factors, and particularly entrepreneurial finance, remains underexplored. This paper addresses this gap by examining how CF supports industrial diversification in the United States. Although CF represents a limited share of entrepreneurial finance investment, its conceptual and empirical interest connects to its diversification-enabling role. In this sense, the focus is not on how much capital CF mobilizes but rather if and how it activates mechanisms — financial, cognitive, and relational — that support the emergence of new industries. Drawing on EEG, I conceptualize CF as an enabling factor that alleviates financing constraints, provides access to knowledge and activates network relations. These mechanisms help entrepreneurs establish new industries locally. Empirically, I test whether CF is associated with subsequent industry entry, whether relatedness density mediates this relationship, and whether credit access conditions shape CF's impact.

My empirical analysis indicates that the provision of a CF is positively and significantly associated with industry entry in the following three years. This is an important finding since we know very little about the role of entrepreneurial finance (Clayton et al. 2024; Sorenson et al. 2016) in the diversification process. Interestingly, the relationship between CF and diversification appears to be generally driven by entries of industries that have received funds through CF but that are also related to the existing industrial structure (Boschma 2017; Hidalgo et al. 2018). Moreover, when exploring the heterogeneity of these relationships across counties with different levels of credit insecurity (Hamdani et al. 2019; Lee and Brown 2017), my analysis indicates that CF plays a stronger and more direct (i.e., not mediated by relatedness density) role in areas where access to credit is particularly difficult (Kim and Hann 2019). This result is particularly interesting since it suggests that CF may facilitate the diversification of structurally and financially weaker areas, thus potentially contribute to reducing territorial disparities. Overall, my findings support the intuition that crowdfunding can contribute to the diversification and structural change of local economies, with the event study analysis and 2SLS regressions suggesting that the association between crowdfunding and diversification may be stronger (between 1.5 and 4 percentage points higher probability of entry) than what the baseline regression suggested (0.3 percentage points higher probability of entry).

On the basis of these findings, this paper has three main contributions and several possible avenues for further research. First, this paper contributes to the literature on EEG by highlighting the importance of the provision of capital resources for diversification. Specifically, by showing that even relatively small-scale projects can meaningfully support the early stages of industry formation, this study expands the current focus of EEG by integrating a dimension of entrepreneurial finance. Building on recent efforts (Clayton et al. 2024; Uhlbach et al. 2022), understanding how different types of entrepreneurial finance, from venture capital to equity-based crowdfunding, can contribute to diversification would be an important step.

Second, the role of the industrial structure emerges as an important factor in translating the resources made available through CF platforms into local economic impact. This conclusion, while not new (Boschma and Capone 2015; Content et al. 2022; Cortinovis et al. 2017; Hidalgo et al. 2018), is relevant, especially in the context of policy (Foray 2014; McCann and Ortega-Argilés 2015, 2016). In terms of future research, exploring the role of other types of local capabilities, such as informal institutions and social capital (Butticè et al. 2017; Cai et al. 2021; Giudici et al. 2018; Cortinovis 2025), may help in understanding what type of local capabilities matter in connection with CF (Cornejo Costas et al. 2025; Painter et al. 2020).

Finally, the impact of CF is especially strong in areas that are most affected by credit insecurity (Hamdani et al. 2019; Kim and Hann 2019). This finding suggests a possibly important role for CF in providing access to financial resources in areas that are structurally weaker and face lower access to capital. While on the basis of the analysis, it is not possible to make any claim about CF addressing spatial disparities, the results point to the possibility that alternative forms of entrepreneurial finance may alleviate the conditions of the structurally disadvantaged regions. Connecting to recent works on development traps (Diemer et al. 2022; Iammarino et al. 2020; Balland and Boschma 2024), this suggests a possible role for CF investments (and possibly entrepreneurial finance in general) in stimulating diversification in credit insecure, and thus potentially trapped, regions. Future research should consider this aspect, also in connection to the growing literature on innovation in the periphery (Castaldi et al. 2025; Fitjar and Rodríguez-Pose 2011).

Despite the contributions of this study, several limitations should be acknowledged. First, while the theoretical framework proposes mechanisms linking CF to industrial diversification, the level of observation in the empirical analysis is too coarse to directly identify and observe the specific these mechanisms at play. The findings are nevertheless consistent with the proposed theoretical arguments and provide systematic evidence on how crowdfunding activity is associated with patterns of industrial entry across regions. Second, the analysis reveals robust and stable associations between CF activity and subsequent industrial entry thanks to the inclusion of fixed effects, a propensity score matching exercise, an event-study analysis and 2SLS regressions. However, the empirical settings do not allow for micro-level causal inference and the estimated relationships should not be interpreted in a causal sense. Future research could build on these findings by exploiting exogenous variation in crowdfunding campaigns to identify causal effects and to further clarify the role of CF in shaping firm behaviour, industrial dynamics, and regional economic trajectories. Addressing these challenges will be essential for deriving more robust policy implications and for assessing how crowdfunding and entrepreneurial finance may be leveraged in support of regional development objectives.

Acknowledgements

I am grateful to Camilla Lenzi, Raquel Ortega-Argiles, Mabel Sanchez-Barrioluengo, Annalisa Caloffi, Niccolò Innocenti, Luciana Lazzeretti, and Francesco Capone, as well as participants at the 'Rethinking Clusters' Conference (Padova), NARSC 2023, GeoInno 2024, the Manchester Institute of Innovation Research (MIOIR) seminar, and the DELOS Seminar (Florence) for valuable comments and suggestions on earlier versions of this work. I also thank the editor and three anonymous referees for their constructive and thoughtful suggestions, which substantially improved the paper. This publication is part of the project 'Crowdfunding and Industrial Development" with file number 406.XS.01.014 of the research programme Open Competition which is financed by the Dutch Research Council (NWO) under the grant nwo.nl/en/projects/406xs01014.

References

Acs, Zoltan J., David B. Audretsch, and Erik E. Lehmann. 2013. "The Knowledge Spillover Theory of Entrepreneurship." *Small Business Economics* 41 (4): 757–74. https://doi.org/10.1007/s11187-013-9505-9.

Agrawal, Ajay, Christian Catalini, and Avi Goldfarb. 2014. "Some Simple Economics of Crowdfunding." *Innovation Policy and the Economy* 14 (January): 63–97. https://doi.org/10.1086/674021.

Alessandrini, Pietro, Michele Fratianni, and Alberto Zazzaro. 2009. "The Changing Geography of Banking and Finance: The Main Issues." In *The Changing Geography of Banking and Finance: The Main Issues*, edited by Alberto Zazzaro, Michele Fratianni, and Pietro Alessandrini. Springer US. https://doi.org/10.1007/978-0-387-98078-2_1.

Antonietti, Roberto, and Sandro Montresor. 2021. "Going beyond Relatedness: Regional Diversification Trajectories and Key Enabling Technologies (KETs) in Italian Regions." *Economic Geography* 97 (2): 187–207. https://doi.org/10.1080/00130095.2021.1920390.

Audretsch, David B. 2002. "The Dynamic Role of Small Firms: Evidence from the U.S." *Small Business Economics* 18 (1): 13–40. https://doi.org/10.1023/A:1015105222884.

Audretsch, David B., Alessandra Colombelli, Luca Grilli, Tommaso Minola, and Einar Rasmussen. 2020. "Innovative Start-Ups and Policy Initiatives." *Research Policy*, Innovative start-ups and policy initiatives, vol. 49 (10): 104027. https://doi.org/10.1016/j.respol.2020.104027.

Bai, John, Wang Jin, and Sifan Zhou. 2024. "Proximity and Knowledge Spillovers: Evidence from the Introduction of New Airline Routes." *Management Science* 70 (11): 7464–85.

Balland, Pierre-Alexandre. 2017. "Economic Geography in R: Introduction to the EconGeo Package." SSRN Scholarly Paper No. 2962146. Rochester, NY, April 27. https://doi.org/10.2139/ssrn.2962146.

Balland, Pierre-Alexandre, and Ron Boschma. 2021. "Complementary Interregional Linkages and Smart Specialisation: An Empirical Study on European Regions." *Regional Studies* 55 (6): 1059–70. https://doi.org/10.1080/00343404.2020.1861240.

Balland, Pierre-Alexandre, and Ron Boschma. 2024. "An Evolutionary Approach to Regional Development Traps in European Regions." *Papers in Evolutionary Economic Geography (PEEG)*, no. 24.20.

Bao, Jiayi. 2025. "Do Makerspaces Affect Entrepreneurship? If so, Who, How, and When?" *Strategic Management Journal* 46 (2): 502–33. https://doi.org/10.1002/smj.3664.

Bargoni, Augusto, Alberto Ferraris, Stefano Bresciani, and Mark Anthony Camilleri. 2022. "Crowdfunding and Innovation: A Bibliometric Review and Future Research Agenda." *European Journal of Innovation Management* ahead-of-print (ahead-of-print). https://doi.org/10.1108/EJIM-06-2022-0310.

Bathelt, Harald, and Rachael Gibson. 2015. "Learning in 'Organized Anarchies': The Nature of Technological Search Processes at Trade Fairs." *Regional Studies* 49 (6): 985–1002. https://doi.org/10.1080/00343404.2013.783691.

Beer, Andrew, Sarah Ayres, Terry Clower, Fabian Faller, Alessandro Sancino, and Markku Sotarauta. 2019. "Place Leadership and Regional Economic Development: A Framework for Cross-Regional Analysis." *Regional Studies* 53 (2): 171–82. https://doi.org/10.1080/00343404.2018.1447662.

Belleflamme, Paul, Thomas Lambert, and Armin Schwienbacher. 2014. “Crowdfunding: Tapping the Right Crowd.” *Journal of Business Venturing* 29 (5): 585–609. https://doi.org/10.1016/j.jbusvent.2013.07.003.

Binz, Christian, Bernhard Truffer, and Lars Coenen. 2016. “Path Creation as a Process of Resource Alignment and Anchoring: Industry Formation for On-Site Water Recycling in Beijing.” *Economic Geography* 92 (2): 172–200. https://doi.org/10.1080/00130095.2015.1103177.

Block, Joern H., Massimo G. Colombo, Douglas J. Cumming, and Silvio Vismara. 2018. “New Players in Entrepreneurial Finance and Why They Are There.” *Small Business Economics* 50 (2): 239–50. https://doi.org/10.1007/s11187-016-9826-6.

Boschma, Ron. 2017. “Relatedness as Driver of Regional Diversification: A Research Agenda.” *Regional Studies* 51 (3): 351–64. https://doi.org/10.1080/00343404.2016.1254767.

Boschma, Ron, and Gianluca Capone. 2015. “Institutions and Diversification: Related versus Unrelated Diversification in a Varieties of Capitalism Framework.” *Research Policy* 44 (10): 1902–14. https://doi.org/10.1016/j.respol.2015.06.013.

Boschma, Ron, Lars Coenen, Koen Frenken, and Bernhard Truffer. 2017. “Towards a Theory of Regional Diversification: Combining Insights from Evolutionary Economic Geography and Transition Studies.” *Regional Studies* 51 (1): 31–45. https://doi.org/10.1080/00343404.2016.1258460.

Boschma, Ron, and Koen Frenken. 2010. “Technological Relatedness and Regional Branching.” In *Beyond Territory*. Routledge.

Boschma, Ron, and Koen Frenken. 2011. “Technological Relatedness, Related Variety and Economic Geography.” Handbook of Regional Innovation and Growth. *Handbook of Regional Innovation and Growth*, August 31. https://www.elgaronline.com/view/edcoll/9781848444171/9781848444171.00028.xml.

Breschi, S. 2001. “Knowledge Spillovers and Local Innovation Systems: A Critical Survey.” *Industrial and Corporate Change* 10 (4): 975–1005. https://doi.org/10.1093/icc/10.4.975.

Breznitz, Shiri M., and Douglas S. Noonan. 2020. “Crowdfunding in a Not-so-Flat World.” *Journal of Economic Geography* 20 (4): 1069–92. https://doi.org/10.1093/jeg/lbaa008.

Brown, Ross, José Liñares-Zegarra, and John O. S. Wilson. 2019. “Sticking It on Plastic: Credit Card Finance and Small and Medium-Sized Enterprises in the UK.” *Regional Studies* 53 (5): 630–43. https://doi.org/10.1080/00343404.2018.1490016.

Bruton, Garry, Susanna Khavul, Donald Siegel, and Mike Wright. 2015. “New Financial Alternatives in Seeding Entrepreneurship: Microfinance, Crowdfunding, and Peer–to–Peer Innovations.” *Entrepreneurship Theory and Practice* 39 (1): 9–26. https://doi.org/10.1111/etap.12143.

Butticè, Vincenzo, Massimo G. Colombo, and Mike Wright. 2017. “Serial Crowdfunding, Social Capital, and Project Success.” *Entrepreneurship Theory and Practice* 41 (2): 183–207. https://doi.org/10.1111/etap.12271.

Butts, Kyle, and John Gardner. 2022. “Did2s: Two-Stage Difference-in-Differences.” *The R Journal* 14 (3): 162–73. https://doi.org/10.32614/RJ-2022-048.

Cai, Wanxiang, Friedemann Polzin, and Erik Stam. 2021. “Crowdfunding and Social Capital: A Systematic Review Using a Dynamic Perspective.” *Technological Forecasting and Social Change* 162 (January): 120412. https://doi.org/10.1016/j.techfore.2020.120412.

Castaldi, Carolina, Nicola Cortinovis, and Milene S. Tessarin. 2025. "Out of Sight? Revealing Creativity-Led Innovation in Rural Regions." *Papers in Evolutionary Economic Geography (PEEG)*, Papers in Evolutionary Economic Geography (PEEG).

Chai, Sen, and Richard B. Freeman. 2019. "Temporary Colocation and Collaborative Discovery: Who Confers at Conferences." *Strategic Management Journal* 40 (13): 2138–64. https://doi.org/10.1002/smj.3062.

Clayton, Paige, Maryann Feldman, and Benjamin Montmartin. 2024. "Entrepreneurial Finance and Regional Ecosystem Emergence." *Small Business Economics* 62 (4): 1493–521. https://doi.org/10.1007/s11187-023-00827-6.

Colombo, Massimo G., Chiara Franzoni, and Cristina Rossi–Lamastra. 2015. "Internal Social Capital and the Attraction of Early Contributions in Crowdfunding." *Entrepreneurship Theory and Practice* 39 (1): 75–100. https://doi.org/10.1111/etap.12118.

Colombo, Massimo G., and Kourosh Shafi. 2021. "Receiving External Equity Following Successfully Crowdfunded Technological Projects: An Informational Mechanism." *Small Business Economics* 56 (4): 1507–29. https://doi.org/10.1007/s11187-019-00259-1.

Content, Jeroen, Nicola Cortinovis, Koen Frenken, and Jacob Jordaan. 2022. "The Roles of KIBS and R&D in the Industrial Diversification of Regions." *The Annals of Regional Science* 68 (1): 29–64. https://doi.org/10.1007/s00168-021-01068-9.

Cornejo Costas, Benjamin, Nicola Cortinovis, and Andrea Morrison. 2025. "How External Linkages and Informal Institutions Enable Green Innovation in EU Regions." *Papers in Evolutionary Economic Geography (PEEG)*, no. 25.03.

Cornejo Costas, Benjamin, Nicola Cortinovis, and Andrea Morrison. 2025. "How Migrant Inventors and Informal Institutions Enable Green Innovation in EU Regions." *Cambridge Journal of Regions, Economy and Society*, November 11, rsaf042. https://doi.org/10.1093/cjres/rsaf042.

Cortinovis, Nicola. 2025. "Does Crowdfunding Foster Rural Entrepreneurship?" *Papers in Evolutionary Economic Geography (PEEG)*, no. 25.35.

Cortinovis, Nicola, Jing Xiao, Ron Boschma, and Frank G. van Oort. 2017. "Quality of Government and Social Capital as Drivers of Regional Diversification in Europe." *Journal of Economic Geography* 17 (6): 1179–208. https://doi.org/10.1093/jeg/lbx001.

Cortinovis, Nicola, D. Zhang, and R. Boschma. 2024. "Regional Diversification and Intra-Regional Wage Inequality in the Netherlands." *Regional Studies*, April 2, 1–15. https://doi.org/10.1080/00343404.2024.2319339.

Cowling, Marc, Ross Brown, and Neil Lee. 2021. "The Geography of Business Angel Investments in the UK: Does Local Bias (Still) Matter?" *Environment and Planning A: Economy and Space* 53 (5): 1180–200. https://doi.org/10.1177/0308518X20984484.

Cumming, Douglas, Michele Meoli, and Silvio Vismara. 2021. "Does Equity Crowdfunding Democratize Entrepreneurial Finance?" *Small Business Economics* 56 (2): 533–52. https://doi.org/10.1007/s11187-019-00188-z.

Cuvero, Marco, Maria L. Granados, Alan Pilkington, and Richard Evans. 2023. "Start-ups' Use of Knowledge Spillovers for Product Innovation: The Influence of Entrepreneurial Ecosystems and Virtual Platforms." *R&D Management* 53 (4): 584–602. https://doi.org/10.1111/radm.12567.

Dam, Alje van, and Koen Frenken. 2022. "Variety, Complexity and Economic Development." *Research Policy*, Special Issue on Economic Complexity, vol. 51 (8): 103949. https://doi.org/10.1016/j.respol.2020.103949.

De Luca, Valerio V., Alessandro Margherita, and Giuseppina Passiante. 2019. "Crowdfunding: A Systemic Framework of Benefits." *International Journal of Entrepreneurial Behavior & Research* 25 (6): 1321–39. https://doi.org/10.1108/IJEBR-11-2018-0755.

Diemer, Andreas, Simona Iammarino, Andrés Rodríguez-Pose, and Michael Storper. 2022. "The Regional Development Trap in Europe." *Economic Geography* 98 (5): 487–509. https://doi.org/10.1080/00130095.2022.2080655.

Dunkley, Emma. 2016. "Crowdfunding Rides to the Rescue of Many SMEs." Special Report. *Financial Times*, February 9. https://www.ft.com/content/a1b69b96-475b-11e5-af2f-4d6e0e5eda22.

Eck, Nees Jan van, and Ludo Waltman. 2009. "How to Normalize Cooccurrence Data? An Analysis of Some Well-Known Similarity Measures." *Journal of the American Society for Information Science and Technology* 60 (8): 1635–51. https://doi.org/10.1002/asi.21075.

Eckert, Fabian, Teresa Fort, Peter Schott, and Natalie Yang. 2020. *Imputing Missing Values in the US Census Bureau's County Business Patterns*. No. W26632. National Bureau of Economic Research. https://doi.org/10.3386/w26632.

Eldridge, Derek, Tahir M. Nisar, and Mariateresa Torchia. 2021. "What Impact Does Equity Crowdfunding Have on SME Innovation and Growth? An Empirical Study." *Small Business Economics* 56 (1): 105–20. https://doi.org/10.1007/s11187-019-00210-4.

Esposito, Christopher, and Frank Van Der Wouden. 2022. "Learning Amid Technological Disruption: Evidencefrom U.S. Inventors, 1836-1999." *Academy of Management Proceedings* 2022 (1): 17898. https://doi.org/10.5465/AMBPP.2022.17898abstract.

Fanjul, Ana P., Liliana Herrera, and Maria F. Munoz-Doyague. 2023. "Fostering Rural Entrepreneurship: An Ex-Post Analysis for Spanish Municipalities." *Technological Forecasting and Social Change* 197 (December): 122915. https://doi.org/10.1016/j.techfore.2023.122915.

Feldman, Maryann P. 2023. "Realizing the Promise of Evolutionary Economic Geography: Ecosystem Perspectives." *Cambridge Journal of Regions, Economy and Society* 16 (3): 587–92. https://doi.org/10.1093/cjres/rsad028.

Fitjar, Rune Dahl, and Andrés Rodríguez-Pose. 2011. "Innovating in the Periphery: Firms, Values and Innovation in Southwest Norway." *European Planning Studies* 19 (4): 555–74. https://doi.org/10.1080/09654313.2011.548467.

Foray, Dominique. 2014. *Smart Specialisation: Opportunities and Challenges for Regional Innovation Policy*. Routledge. https://doi.org/10.4324/9781315773063.

Frenken, Koen, and Ron Boschma. 2015. "Geographic Clustering in Evolutionary Economic Geography." In *Handbook of Research Methods and Applications in Economic Geography*, edited by Charlie Karlsson, Martin Andersson, and Therese Norman. Edward Elgar Publishing. https://doi.org/10.4337/9780857932679.00021.

Gardner, John. 2022. "Two-Stage Differences in Differences." arXiv:2207.05943. Preprint, arXiv, July 13. https://doi.org/10.48550/arXiv.2207.05943.

Giudici, Giancarlo, Massimiliano Guerini, and Cristina Rossi-Lamastra. 2018. "Reward-Based Crowdfunding of Entrepreneurial Projects: The Effect of Local Altruism and Localized Social Capital on Proponents' Success." *Small Business Economics* 50 (2): 307–24. https://doi.org/10.1007/s11187-016-9830-x.

Grillitsch, Markus, and Markku Sotarauta. 2020. “Trinity of Change Agency, Regional Development Paths and Opportunity Spaces.” *Progress in Human Geography* 44 (4): 704–23. https://doi.org/10.1177/0309132519853870.

Gwet, Kilem Li. 2014. *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement among Raters*. Fourth edition. Advances Analytics, LLC.

Hamdani, Kausar, Claire Kramer Mills, Edison Reyes, and Jessica Battisto. 2019. “Unequal Access to Credit.” *Federal Reserve Bank of New York*. https://www.federalreserve.gov/econres/feds/credit-scores-and-committed-relationships.htm.

Hausmann, Ricardo, Jason Hwang, and Dani Rodrik. 2007. “What You Export Matters.” *Journal of Economic Growth* 12 (1): 1–25. https://doi.org/10.1007/s10887-006-9009-4.

Hausmann, Ricardo, and Dani Rodrik. 2003. “Economic Development as Self-Discovery.” *Journal of Development Economics*, 14th Inter-American Seminar on Economics, vol. 72 (2): 603–33. https://doi.org/10.1016/S0304-3878(03)00124-X.

Hervé, Fabrice, and Armin Schwienbacher. 2018. “Crowdfunding and Innovation.” *Journal of Economic Surveys* 32 (5): 1514–30. https://doi.org/10.1111/joes.12274.

Hidalgo, César A., Pierre-Alexandre Balland, Ron Boschma, et al. 2018. “The Principle of Relatedness.” In *Unifying Themes in Complex Systems IX*, edited by Alfredo J. Morales, Carlos Gershenson, Dan Braha, Ali A. Minai, and Yaneer Bar-Yam. Springer Proceedings in Complexity. Springer International Publishing. https://doi.org/10.1007/978-3-319-96661-8_46.

Hidalgo, César A., B. Klinger, A. L. Barabási, and R. Hausmann. 2007. “The Product Space Conditions the Development of Nations.” *Science* 317 (5837): 482–87. https://doi.org/10.1126/science.1144581.

Hornuf, Lars, Matthias Schmitt, and Eliza Stenzhorn. 2018. “Equity Crowdfunding in Germany and the United Kingdom: Follow-up Funding and Firm Failure.” *Corporate Governance: An International Review* 26 (5): 331–54. https://doi.org/10.1111/corg.12260.

Iammarino, Simona, Andrés Rodríguez-Pose, and Michael Storper. 2020. *Falling into the Middle-Income Trap?: A Study on the Risks for EU Regions to Be Caught in a Middle Income Trap : Final Report.* Final Report. Publications Office. https://data.europa.eu/doi/10.2776/02363.

Jacobs, Wouter, Hans R. A. Koster, and Frank van Oort. 2014. “Co-Agglomeration of Knowledge-Intensive Business Services and Multinational Enterprises.” *Journal of Economic Geography* 14 (2): 443–75. https://doi.org/10.1093/jeg/lbs055.

Jara-Figueroa, Christian, Bogang Jun, Edward L. Glaeser, and César A. Hidalgo. 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* 115 (10). https://doi.org/10.1073/pnas.1800475115.

Johan, Sofia, and Jon Taylor. 2019. “Does Crowdfunding Democratize Success? Revisiting the Effects of Agglomeration and Localized Knowledge Spillover on Creative Projects.” SSRN Scholarly Paper No. 3484487. Rochester, NY, November 10. https://doi.org/10.2139/ssrn.3484487.

Kärnä, Anders, and Andreas Stephan. 2022. “Do Firms in Rural Regions Lack Access to Credit? Local Variation in Small Business Loans and Firm Growth.” *Regional Studies* 56 (11): 1919–33. https://doi.org/10.1080/00343404.2021.2016681.

Kim, Keongtae, and Il-Horn Hann. 2019. “Crowdfunding and the Democratization of Access to Capital—An Illusion? Evidence from Housing Prices.” *Information Systems Research*, ahead of print, February 22. World. https://doi.org/10.1287/isre.2018.0802.

Klein, Daniel. 2018. “Implementing a General Framework for Assessing Interrater Agreement in Stata.” *The Stata Journal* 18 (4): 871–901. https://doi.org/10.1177/1536867X1801800408.

Kogler, Dieter F., Adam Whittle, Keungoui Kim, and Balázs Lengyel. 2023. “Understanding Regional Branching: Knowledge Diversification via Inventor and Firm Collaboration Networks.” *Economic Geography* 99 (5): 471–98. https://doi.org/10.1080/00130095.2023.2242551.

Kuppuswamy, Venkat, and Barry L. Bayus. 2015. “A Review of Crowdfunding Research and Findings.” In *Handbook of New Product Development Research*. Edward Elgar Publishing.

Lee, Neil, and Ross Brown. 2017. “Innovation, SMEs and the Liability of Distance: The Demand and Supply of Bank Funding in UK Peripheral Regions.” *Journal of Economic Geography* 17 (1): 233–60. https://doi.org/10.1093/jeg/lbw011.

Leland, Jonathan. 2022. “Kickstarter Data, Global, 2009-2023: Version 3.” Version v3. ICPSR - Interuniversity Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR38050.V3.

Macht, Stephanie A., and Jamie Weatherston. 2014. “The Benefits of Online Crowdfunding for Fund-Seeking Business Ventures.” *Strategic Change* 23 (1–2): 1–14. https://doi.org/10.1002/jsc.1955.

Markard, Jochen, and Bernhard Truffer. 2008. “Technological Innovation Systems and the Multi-Level Perspective: Towards an Integrated Framework.” *Research Policy* 37 (4): 596–615. https://doi.org/10.1016/j.respol.2008.01.004.

Martínez-Climent, Carla, Leonardo Mastrangelo, and Domingo Ribeiro-Soriano. 2021. “The Knowledge Spillover Effect of Crowdfunding.” *Knowledge Management Research & Practice* 19 (1): 106–16. https://doi.org/10.1080/14778238.2020.1768168.

McCann, Philip, and Raquel Ortega-Argilés. 2015. “Smart Specialization, Regional Growth and Applications to European Union Cohesion Policy.” *Regional Studies* 49 (8): 1291–302. https://doi.org/10.1080/00343404.2013.799769.

McCann, Philip, and Raquel Ortega-Argilés. 2016. “Smart Specialisation, Entrepreneurship and SMEs: Issues and Challenges for a Results-Oriented EU Regional Policy.” *Small Business Economics* 46 (4): 537–52. https://doi.org/10.1007/s11187-016-9707-z.

McKenzie, David, and Christopher Woodruff. 2008. “Experimental Evidence on Returns to Capital and Access to Finance in Mexico.” *The World Bank Economic Review* 22 (3): 457–82.

Mollick, Ethan. 2014. “The Dynamics of Crowdfunding: An Exploratory Study.” *Journal of Business Venturing* 29 (1): 1–16. https://doi.org/10.1016/j.jbusvent.2013.06.005.

Mollick, Ethan. 2016. “Containing Multitudes: The Many Impacts of Kickstarter Funding.” *SSRN Electronic Journal*, ahead of print. https://doi.org/10.2139/ssrn.2808000.

Mollick, Ethan, and Venkat Kuppuswamy. 2016. “Crowdfunding: Evidence on the Democratization of Start-up Funding.” In *Revolutionizing Innovation: Users, Communities, and Open Innovation*. MIT Press. https://direct.mit.edu/books/edited-volume/4075/chapter/169013/Crowdfunding-Evidence-on-the-Democratization-of.

Mollick, Ethan, and Alicia Robb. 2016. “Democratizing Innovation and Capital Access: The Role of Crowdfunding.” *California Management Review* 58 (2): 72–87. https://doi.org/10.1525/cmr.2016.58.2.72.
Moritz, Alexandra, and Joern H. Block. 2016. “Crowdfunding: A Literature Review and Research Directions.” In *Crowdfunding in Europe: State of the Art in Theory and Practice*, edited by Dennis Brüntje and Oliver Gajda. FGF Studies in Small Business and Entrepreneurship. Springer International Publishing. https://doi.org/10.1007/978-3-319-18017-5_3.
Neffke, Frank, Matté Hartog, Ron Boschma, and Martin Henning. 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, Frank, Martin Henning, and Ron Boschma. 2011. “How Do Regions Diversify over Time? Industry Relatedness and the Development of New Growth Paths in Regions.” *Economic Geography* 87 (3): 237–65. https://doi.org/10.1111/j.1944-8287.2011.01121.x.
O’Donoghue, Dan, and Bill Gleave. 2004. “A Note on Methods for Measuring Industrial Agglomeration.” *Regional Studies* 38 (4): 419–27. https://doi.org/10.1080/03434002000213932.
Painter, Deryc T., Frank van der Wouden, Manfred D. Laubichler, and Hyejin Youn. 2020. “Quantifying Simultaneous Innovations in Evolutionary Medicine.” *Theory in Biosciences* 139 (4): 319–35. https://doi.org/10.1007/s12064-020-00333-3.
Pominova, Mariya, Todd Gabe, and Andrew Crawley. 2022. “The Stability of Location Quotients.” *Review of Regional Studies* 52 (3). https://doi.org/10.52324/001c.66197.
Qiao, Yibo, Andrea Ascani, and Andrea Morrison. 2024. “External Linkages and Regional Diversification in China: The Role of Foreign Multinational Enterprises.” *Environment and Planning A: Economy and Space* 56 (4): 1077–101. https://doi.org/10.1177/0308518X231223621.
Sewaid, Ahmed, Shaker A. Zahra, and Sebastian Aparicio. 2026. “From Novice to Serial Entrepreneur: How Institutions Influence Relaunching Decisions on International Crowdfunding Platforms.” *Technovation* 149 (January): 103355. https://doi.org/10.1016/j.technovation.2025.103355.
Sorenson, Olav, Valentina Assenova, Guan-Cheng Li, Jason Boada, and Lee Fleming. 2016. “Expand Innovation Finance via Crowdfunding.” *Science* 354 (6319): 1526–28. https://doi.org/10.1126/science.aaf6989.
Sotarauta, Markku, Heli Kurikka, Jari Kolehmainen, Sami Sopanen, and Sami Moisio. 2025. “Place Leadership in the Geography of Opportunities: Survival of Small Industrial Localities When Industrial Giants Fall.” *Regional Studies* 59 (1): 2479108. https://doi.org/10.1080/00343404.2025.2479108.
Steijn, Mathieu P. A. 2021. “Improvement on the Association Strength: Implementing a Probabilistic Measure Based on Combinations without Repetition.” *Quantitative Science Studies* 2 (2): 778–94. https://doi.org/10.1162/qss_a_00122.
Steijn, Mathieu P. A., Pierre-Alexandre Balland, Ron Boschma, and David L. Rigby. 2023. “Technological Diversification of U.S. Cities during the Great Historical Crises.” *Journal of Economic Geography* 23 (6): 1303–44. https://doi.org/10.1093/jeg/lbad013.
Tian, Zheng. 2013. “Measuring Agglomeration Using the Standardized Location Quotient with a Bootstrap Method.” *The Journal of Regional Analysis & Policy* 43 (2): 186–97.

Ughetto, Elisa, Marc Cowling, and Neil Lee. 2019. “Regional and Spatial Issues in the Financing of Small and Medium-Sized Enterprises and New Ventures.” *Regional Studies* 53 (5): 617–19. https://doi.org/10.1080/00343404.2019.1601174.

Uhlbach, Wolf-Hendrik, Pierre-Alexandre Balland, and Thomas Scherngell. 2022. “Public R&D Funding and New Regional Specialisations: The Contingent Role of Technological Relatedness.” *Industry and Innovation* 29 (4): 511–32. https://doi.org/10.1080/13662716.2022.2043147.

Wald, Andreas, Merete Holmesland, and Kalanit Efrat. 2019. “It Is Not All About Money: Obtaining Additional Benefits Through Equity Crowdfunding.” *The Journal of Entrepreneurship* 28 (2): 270–94. https://doi.org/10.1177/0971355719851899.

Walthoff-Borm, Xavier, Tom Vanacker, and Veroniek Collewaert. 2018. “Equity Crowdfunding, Shareholder Structures, and Firm Performance.” *Corporate Governance: An International Review* 26 (5): 314–30. https://doi.org/10.1111/corg.12259.

Yeung, Henry Wai-chung. 2021. “Regional Worlds: From Related Variety in Regional Diversification to Strategic Coupling in Global Production Networks.” *Regional Studies* 55 (6): 989–1010. https://doi.org/10.1080/00343404.2020.1857719.

Yu, Sandy, Scott Johnson, Chiayu Lai, Antonio Cricelli, and Lee Fleming. 2017. “Crowdfunding and Regional Entrepreneurial Investment: An Application of the CrowdBerkeley Database.” *Research Policy* 46 (10): 1723–37. https://doi.org/10.1016/j.respol.2017.07.008.

Appendix 1

Matching strategy and outcomes

Using propensity score and exact matching is very effective in reducing the difference between treated and untreated units, as represented in Figure 1.A below. As the figure shows, the mean differences across all the variables are close to 0 and much smaller than when the full sample is considered. The internal validity of my matching exercise is confirmed by further analysis of the difference between the means of treated and untreated units, which become rather small (and mostly statistically insignificant) after the matching exercise (Table 9). Two variables, *Perm. VC* (dummy) and # *Plants* still reported a significant difference between treated and untreated plants. While this is not ideal, the difference between treated and untreated after the matching procedure is remarkably small, to a point that may well be statistically significant but economically negligible (e.g., in the matched sample, 15% of treated county-industry groups received some venture capital, whereas 13% of the untreated ones did).

Figure 6: Results of the matching exercise

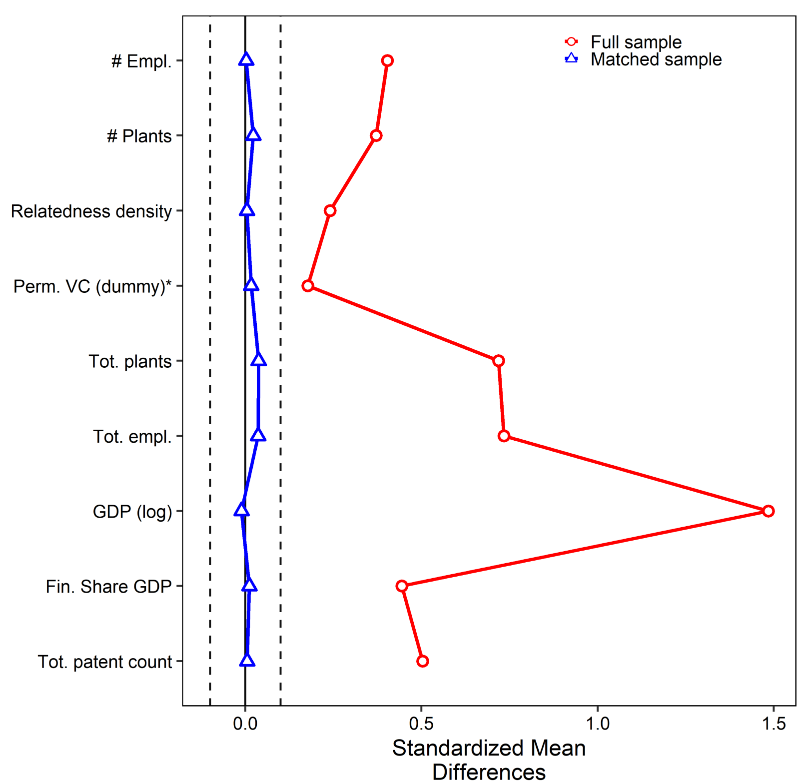


Table 9: Overview of the sample balance before and after matching					
Variable	Mean Treated	Mean Controls	T Statistic	p-value	Sample
Entry 3y	0.01	0.02	-6.94	0	Full sample
# Plants	272.99	32.44	35.24	0	Full sample
Perm. VC (dummy)	0.14	0.01	58.23	0	Full sample
Density	944.38	102.06	50.56	0	Full sample
Tot. Plants	8.91	6.75	235.72	0	Full sample
Tot. empl.	11.59	9.21	231.55	0	Full sample
GDP	16.27	13.91	218.41	0	Full sample
Fin. Share GDP	0.22	0.19	64.57	0	Full sample
Entry 3y	0.01	0.01	-0.14	0.89	Matched sample
# Plants	299.39	279.45	2.37	0.02	Matched sample
Perm. VC (dummy)	0.15	0.13	3.98	0	Matched sample
Density	7.23	7.2	0.35	0.73	Matched sample
Tot. Plants	9.3	9.3	0.4	0.69	Matched sample
Tot. empl.	12.03	12.03	0.19	0.85	Matched sample
GDP	16.71	16.73	-0.99	0.32	Matched sample
Fin. Share GDP	0.22	0.22	1.03	0.3	Matched sample

The table reports covariate balance between treated and control county-industry-year observations before (full sample) and after (matched sample) the matching. Treated observations are county-industry-year cells with at least one large successful crowdfunding campaign (a single campaign having raised at least 5,000 USD). Control observations are county-industry-year cells with no crowdfunding activity. For the matching, I have used a combination of propensity score and exact matching, using the variables reported for the propensity score part (estimated via a logit) measured in the years before the analysis (2005-2008) and conditioning on being the in the same US state but in a different county for the exact matching part. After the matching the differences between treated and control group markedly decrease, signalling the matching is effective in balancing the sample.

Identification strategy through instrumental variables

To address the endogeneity of crowdfunding activity at the industry-location level, I construct two instrumental variables. Both instruments exploit variation in the presence of large successful crowdfunding campaigns that is plausibly exogenous to the entry decisions of firms in a given industry-county-year. At the same time, as noted above, I do not interpret this analysis in a strictly causal sense, since it is still possible for the exclusion restrictions to be violated.

The first instrument follows the same logic as Sorenson et al. (2016) as it aims at capturing information on crowdfunding from unrelated industries within the same county-year. Theoretically, the presence of crowdfunding campaigns in unrelated industries in the same location should reflect the presence of relevant information on crowdfunding and the legitimacy of the local crowdfunding ecosystem. In formal terms, I build the instrument as follows:

$$Z_{i,c,t}^{1} = \begin{cases} 1 \ if \displaystyle\sum_{j \neq i, 0<\varphi_{i,j}<0.5} CF_{j,c,t} \geq 1 \\ 0 \ otherwise \end{cases}$$

where $Z_{i,c,t}$ takes value 1 if, in the same county and year but in industries which are unrelated to the focal one there is at least one large successful campaigns. To define unrelated industries, I consider industries pairs whose relatedness $\varphi_{i,j}$ is lower than 0.5. It is important to recall a value of 1 in the relatedness score represent a level of relatedness equal to the random benchmark. The instrument therefore uses information on industry relatively distant in the industry space from the focal one. This conservative choice is key for the exclusion restriction: focusing on cognitively distant industries, I mitigate the risk that the instrument captures local industry-cluster dynamics that independently drive entry in the focal industry (e.g. through spinoffs or knowledge spillovers from strongly related industries which may impact the quality and substance of the ideas behind campaigns). I use the relatedness matrix for 2008 – i.e. the year before the first year of the analysis – to further prevent any contamination of the instrument.

The second instrument uses a similar approach but focuses on geographical distance. More specifically, this instrument leverages large crowdfunding campaigns having occurred in the same industry, year and state as the endogenous variable, but excluding the specific county and its adjacent neighbours. This geographical restriction should then ensure the instrument is plausibly exogenous to entry dynamics in the same industry in focal county. Beyond a certain distance, it unlikely that the successful crowdfunding campaigns will affect local industrial entry through channels other than the provision of information and legitimacy thanks to experienced entrepreneurs and backers with industry-specific knowledge are relatively close by. In formal terms,

$$Z_{i,c,t}^{2} = \begin{cases} 1 \ if \displaystyle\sum_{a \in S(c), \ a \notin \mathrm{N(c)}} CF_{i,a,t} \geq 1 \\ 0 \ otherwise \end{cases}$$

where a represent each county of the state S to which the focal county c belongs which is not part of the set N of counties adjacent (based on queen contiguity) to the focal county c.

Table 10: First-stage regressions

	FS IV1	FS IV2	FS IV1 and 2
Tot. Empl. (log)	0.015***	0.016***	0.014***
	(0.003)	(0.003)	(0.002)
Relatedness density	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)
Fin. Share GDP	0.002	0.004	0.005
	(0.006)	(0.007)	(0.006)
GDP (log)	-0.004**	-0.005**	-0.004**
	(0.001)	(0.002)	(0.001)
Perm. VC (dummy)	0.052***	0.052***	0.051***
	(0.004)	(0.004)	(0.004)
IV1	0.020***		0.020***
	(0.001)		(0.001)
IV2		0.025***	0.025***
		(0.002)	(0.002)
Num.Obs.	743795	744320	743795
R2 Adj.	0.112	0.114	0.116
R2 Within Adj.	0.009	0.011	0.013
County FE	X	X	X
NAICS4 FE	X	X	X
Year FE	X	X	X
Clustered SE	County	County	County
First-stage F	1871.65	3626.17	2746.85
Sargan-Hansen stat.			0.081
Sargan-Hansen p-value			0.776

Clustered standard errors in parenthesis

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix 2: Data development and validation

Given the use of new technologies to classify about 170,000 Kickstarter projects into different industry codes, it is important to discuss the data development procedure and data validation performed. In terms of data development, I provided the chatGPT API with the title, blurb (short introductory description of the project under the project's title) and the relevant category and subcategory in which the project is listed and ask the algorithm to classify the project in one of the 311 4-digit codes of 2017 NAICS classification. The process used to develop the data is more thoroughly discussed by (removed to maintain anonymity), and the python script used to develop the data is available in this GitHub repository (link removed for anonymity). Using the information developed with the help of AI, the information at the KS project level has been aggregated by county, year and 3-digit NAICS codes. Data validation was also performed on the basis of 3-digit data.

With respect to data validation, I perform three main tests. First, I check the reliability of the prediction made by ChatGPT (version 3.5, produced in September 2024) and approach the issue as a standard interrater reliability question (Gwet 2014; Klein 2018). Essentially, I consider ChatGPT as a possible rater facing a classification problem and compare its performance to that of other human raters. To this end, given the impossibility of rating 170,000 projects, I created a sample of more than 500 projects, stratified according to the distribution of Kickstarter categories to maximize the representativity of the sample. I then proceed and classify each of the projects into the best-fitting 3-digit code out of the 86 possible codes of the NAICS 2017 classification. Importantly, the human rater—in this case, myself—used the same information available to ChatGPT (title, blurb, category and subcategory) to perform the classification job and had no knowledge of the industry code selected by chatGPT. I then calculate some basic descriptive statistics (Figure 2.A), in particular, a confusion matrix (Panel A) showing the overall good level of overlap between human and electronic raters and three standard measures of reliability across raters: Cohen's kappa and Gwet's AC1 and percentage agreement (Panel B). With respect to the confusion matrix, it is interesting to note that the human rater and machine mostly agree, with the main exception being for cases in which ChatGPT provided the code 511 and the human rater code slightly different codes (e.g., 512, 711). A data inspection indicates that these projects are creative projects—in particular music album production—which are inherently difficult to classify, as they both imply some music performance (NAICS code 711) as well as the publication of the album (NAICS code 511). Concerning the graph in Panel B, in 65% of the cases, the human and ChatGPT agreed on the same code, with both Cohen's kappa and Gwet's AC1 suggesting a substantial level of agreement, according to the Landis-Koch benchmark (Gwet, 2014).

Figure 7: Data validation comparing human and machine

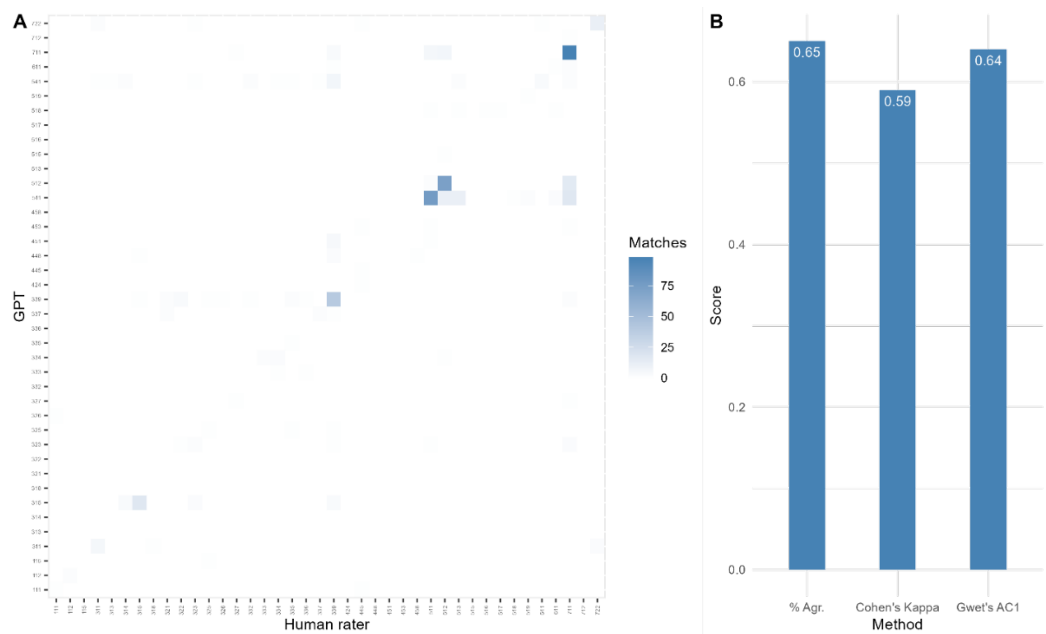
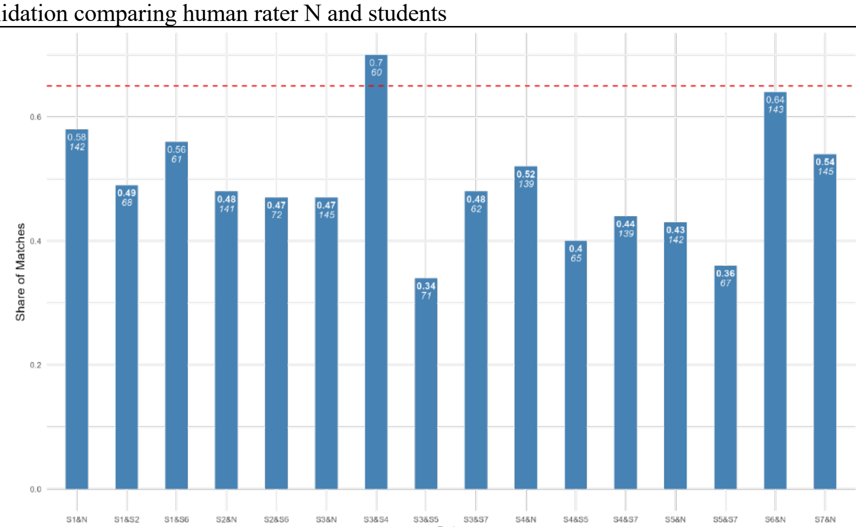


Figure 8: Data validation comparing human rater N and students



To obtain a more credible benchmark against which to evaluate the score, I perform a second test to determine whether different humans would agree more or less than myself and chatGPT. To this end, 7 third-year bachelor economics students were asked to match a list of projects to the relevant 2017 NAICS codes. Each student had a specific set of approximately 140 projects, which were randomly drawn from the same sample of 500 hand-coded projects. The sample each student had been built in such a way that at least some overlap with the sample of someone else occurred. Figure 3.A above reports the share of overlap between different pairs of students and between students and myself (N), including both the share of agreement (number on the top, in bold if significantly different from 0.65, the agreement score between me and ChatGPT) and the number of total cases (in italics). The red dashed line represents the overlap between the human rater and the GPT (0.65). The highest level of overlap reached is between students S3 and S4 (70% of the cases of 63 projects evaluated), and the lowest is between students S3 and S5 (34% of the cases out of 74 projects). Interestingly, the only sets for which a statistically significant difference exists are all agreeing less than myself and ChatGPT. In only one case, and arguably the one with the fewest projects, humans performed better but not in a statistically significant way. Overall, this provides more confidence in the quality of the developed database, as the chat GPT did not underperform statistically or substantially in its classification task.

While the previous two validation tests (human rater vs. machine, comparison between human vs. machine and human vs. human) aimed at checking how accurate chatGPT was, the last test aimed at checking whether the data developed provided a plausible picture. In other words, I want to check whether the industry-location-year groups identified by ChatGPT effectively exist. To this end, I compare the data I developed with the County Business Patterns database (Eckert et al. 2020). Figure 4.A below reports my findings. Overall, in about 98% of the cases, the industry-county-year group to which a project was allocated existed in the official data, as shown by the waffle graph in Panel A (mostly green, with only two orange squares). Checking the accuracy of the data over the years of my sample, I see a slight decline over time, with the overall consistency remaining higher than 97%.

Figure 9: Accuracy of generated data

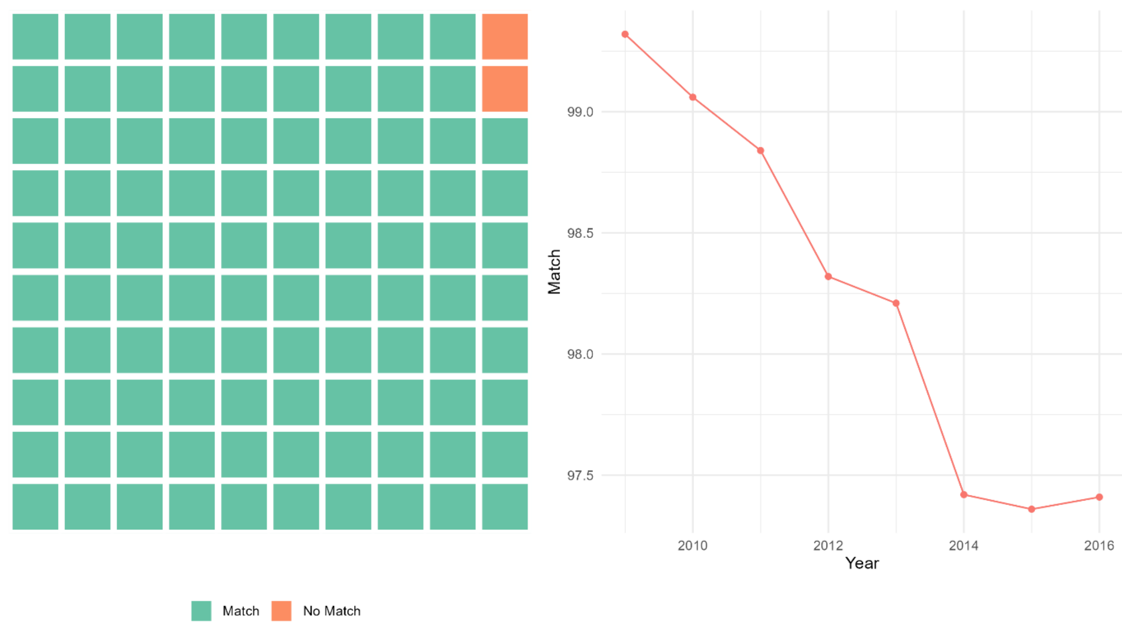


Table 11: Comparison of Kickstarter data and KIUS data

Size group	Succ. proj	Succ. proj. in KIUS	Coverage
<1k	14101	8948	0.63
1k<=x<10k	67155	46996	0.7
10k<=x<20k	16889	12233	0.72
20k<=x<100k	16008	11563	0.72
100k<=x<1M	3450	2418	0.7
x<=1M	210	167	0.8