Growth and Change

WILEY

growth and change

ORIGINAL ARTICLE OPEN ACCESS

Rural Entrepreneurship in the Face of Geodemographic Challenges

Yolanda Fuertes-Callén | Beatriz Cuéllar-Fernández | Adriana Serrano-Magdalena

Department of Accounting & Finance, Faculty of Economy and Business Administration, University of Zaragoza, Zaragoza, Spain

Correspondence: Yolanda Fuertes-Callén (yfuertes@unizar.es)

Received: 4 December 2023 | **Revised:** 16 September 2024 | **Accepted:** 17 December 2024

Funding: This study was funded by the European Regional Development Fund (ERDF), the Spanish Ministry of Education and Science (codes PID2022-136818NB-I00 and PID2023-146084OB-I00), and the Government of Aragon (codes S38_23R and S33_23R).

Keywords: decision trees | economic geography | entrepreneurial finance | rural studies | start-ups

ABSTRACT

Rural entrepreneurs face significant obstacles due to factors such as sparse population, low density, aging population, remoteness from markets and financial providers, inadequate infrastructure, and challenging geographic conditions. These factors contribute to rural disadvantage, fueling rural depopulation. Governments often attempt to address rural depopulation by supporting entrepreneurship, but eligibility criteria vary widely. This paper develops a model of rural entrepreneurship liability explained by municipal demographic and geographic factors. We analyzed the geographical distribution of all companies established in Spain since 2012, along with the attributes of the municipalities where they are headquartered. While working-age municipal population is the primary factor in determining rural entrepreneurship liability, network coverage, altitude, and population variation also play noteworthy roles. Geographic constraints have a significant negative impact on entrepreneurship, but their influence is mitigated by the working-age municipal population, highlighting the unique challenges faced by rural entrepreneurs. One policy implication of our findings is that governments could consider using this model to allocate grants based on the actual difficulty of undertaking entrepreneurship in each municipality.

1 | Introduction

Rural entrepreneurs face heightened challenges compared to their urban counterparts, referred to as "liability of rurality" (Clausen 2020). Multiple factors are involved in this, including remoteness from markets, customers, and suppliers (Laurin, Pronovost, and Carrier 2020), difficulty in attracting and retaining talent (Agarwal, Rahman, and Errington 2009; Lavesson 2018), and limited infrastructure (Sorenson 2018). These challenges contribute to rural depopulation, prompting public administrations to provide support for rural entrepreneurship, a practice subject to scrutiny (Shane 2009). The criteria for providing such support often lack consistency and effectiveness in assessing these challenges, which motivated our research. For instance, some European public administrations extend aid to rural entrepreneurs residing in municipalities with fewer than 5000 inhabitants. However, this simplistic approach overlooks situations where small municipalities are in proximity to major cities and can benefit from that proximity. Conversely, a declining isolated municipality with 5001 inhabitants would not qualify. Other administrations provide similar grants but with varying criteria and thresholds. The lack of a suitable method to measure rural disadvantage is evident, which is the gap that this study aims to fill. To achieve this, we developed a model to predict business creation in rural areas by considering geographic and demographic characteristics, shedding light on the factors contributing to rural disadvantage and their relative importance.

The main hypothesis regarding the impact of rurality suggests that challenging demographic and geographical conditions hinder the establishment of businesses in rural areas. This hypothesis has garnered substantial support (Clausen 2020; Huiban 2011), although not all businesses benefit from proximity to urban centers (Lavesson 2018). Furthermore, several unresolved questions persist, such as the challenges associated with measuring some determinants of business entry and the potential influence of interaction variables (Freire-Gibb and Nielsen 2014). In addition, despite extensive research on geographical factors, such as distance to metropolitan areas or capital cities (Eriksson and Rataj 2019; Laurin, Pronovost, and Carrier 2020; Müller and Korsgaard 2018; Stearns et al. 1995), the role of other geographical factors, such as municipal altitude and slope, remains less explored. In most studies explaining business entry, researchers typically focus on larger administrative units, such as counties (Argent 2018; Bu and Liao 2022; Deller, Kures, and Conroy 2019), provinces (del Olmo-García et al. 2023), regions (Müller and Korsgaard 2018; Novejarque Civera, Pisá Bó, and López-Muñoz 2021), states (Krichevskiy and Snyder 2015), and even countries (Acs, Desai, and Klapper 2008). However, few studies employ the smallest administrative unit, the municipality. This may be due to the difficulty in obtaining data from small villages with only a few hundred inhabitants, which are often distant from the capital but maintain local administrative autonomy.

Our study differs in how the research questions are approached, the administrative unit analyzed, the explanatory variables examined, and the analytical techniques employed. We address two research questions. The first focuses on determinants of business entry, a well-established topic in entrepreneurship research (Clausen 2020; Huiban 2011; Laurin, Pronovost, and Carrier 2020; Renski 2008). Rural entrepreneurs face heightened challenges compared to their urban counterparts, indicating a liability of rurality (Clausen 2020), which may be attributed to various factors (Agarwal, Rahman, and Errington 2009; Laurin, Pronovost, and Carrier 2020; Lavesson 2018; Sorenson 2018). Although we develop a business entry model identifying explanatory factors, comprising previously unexplored sociodemographic and geographic variables, our primary objectives are to determine their relative importance and detect complex relationships among them. This is achieved through appropriate statistical techniques complementing standard regression analyses. Exploring this question is relevant for understanding specific roadblocks, providing a nuanced understanding of the liability of rurality's explanatory theories, and guiding targeted policies and interventions addressing critical impediments to rural entrepreneurship.

Global rural population growth rates saw their first decline in 2021 (World Bank 2023). Many countries are now promoting rural entrepreneurship as a solution to address rural depopulation and offer various types of aid. However, there is no exact definition of rural liability. Our second research question aims to measure rural liability by moving beyond simplistic population-based categorizations of rural municipalities (e.g., any municipality with fewer than X inhabitants), recognizing the complexities in defining rurality (Pato and Teixeira 2016). While most studies tend to define rurality based on reflective indicators, such as population size, density, and proximity to metropolitan areas (Nelson and Nguyen 2023), our research adopts an alternative approach. We measure rural liability based on its effects, specifically the lack of business creation in rural municipalities, or in other words, "rural entrepreneurship liability." To this end, we used our explanatory model of business entry, which predicts the probability of a municipality creating a business from a set of geographical and sociodemographic explanatory variables. By focusing on the consequences of rurality, this approach captures the true impact of rural challenges more directly. Furthermore, our approach transcends arbitrary thresholds that can be rigid and fail to capture the nuances of rurality. Through the identification of municipalities with low business creation rates, policymakers can gain a better understanding of the challenges faced by potential entrepreneurs in each municipality, enabling the development of more effective targeted policies and interventions.

Most prior studies have focused on larger administrative units such as counties, provinces, regions, or states (Acs, Desai, and Klapper 2008; Argent 2018; Bu and Liao 2022; del Olmo-García et al. 2023; Deller, Kures, and Conroy 2019; Krichevskiy and Snyder 2015; Müller and Korsgaard 2018; Novejarque Civera, Pisá Bó, and López-Muñoz 2021). In contrast, our research provides a more granular and localized perspective by focusing on municipalities, allowing for a deeper understanding of the unique dynamics and challenges present at the municipal level. We have studied the establishment of all companies in each of the 8131 Spanish municipalities over a decade. We examined the municipalities where 407,877 Spanish companies were established from 2012 to 2021, evaluating their sociodemographic and geographical attributes. Most municipalities are undergoing population decline, especially in rural areas. Notably, 6827 of these municipalities have populations below 5000, collectively representing 5.7 million residents, accounting for 12% of the total population. Over the past decade, 78.9% of these municipalities have observed a population decrease. The data are sourced from the Spanish Commercial Register, which is mandatory for all companies regardless of their size, thus minimizing potential biases.

Our study's scope of explanatory factors is comprehensive, encompassing both sociodemographic and geographic variables. These include population size, density, inhabitants within a 50-km radius, population change, unemployment rate, migration rate, average age, distance to provincial capital and municipalities with over 50,000 inhabitants, Internet coverage, and transportation infrastructure amenities. Furthermore, our model incorporates lesser-explored geographic factors like altitude and municipal slope, which have demonstrated significance in other contexts (Chen, Lu, and Liu 2022; He et al. 2023; Yin et al. 2021) but remain unexplored in entrepreneurship literature. The inclusion of this wide set of factors ensures the breadth and depth of our analytical approach.

The techniques employed in our study aim to elucidate the relative importance of each explanatory factor. While linear regression is commonly used, interpreting coefficients can be ambiguous (Bring 1994), prompting recommendations for supplementary techniques (Tonidandel and LeBreton 2011). To address this, we employed the Pratt Relative Importance Index (Pratt 1987), dominance analysis (Budescu 1993), and the Least

Absolute Shrinkage and Selection Operator (LASSO) technique (Tibshirani 1996), a novel approach in business entry entrepreneurship research. Additionally, traditional linear models may inadequately capture potential nonlinear relationships between variables influencing business entry. Incorporating interaction terms in linear regressions is one approach to identify nonlinearities by allowing variable effects to vary based on another variable's value. However, decision trees offer a more flexible alternative for capturing nonlinear relationships without imposing specific functional forms (Biggs, De Ville, and Suen 1991). By partitioning the feature space into smaller regions, decision trees effectively identify interactions and nonlinear patterns, which may be appropriate considering our data's complexity.

Our study makes three contributions to the literature on rural entrepreneurship. Firstly, our findings extend the literature on rural business entry by expanding theoretical models. Our study highlights the adverse effects of combining factors such as low population, poor network coverage, high altitude, steep slopes, and population decline, some of them showing clear nonlinear relationships. In particular, the relationship between unemployment, population size, and business creation has been a controversial research topic (Audretsch and Fritsch 1994; Barboza 2024; Cheng and Li 2010). We elucidate some underlying mechanisms of these relationships. While the organizational ecology perspective (Hannan and Freeman 1977) predicts a positive correlation driven by increased entrepreneurial necessity due to unemployment, our findings reveal a more nuanced picture. We found an overall positive correlation, but a detailed analysis considering population size uncovers an instance of Simpson's paradox (Simpson 1951), which happens when the association between two variables is reversed when the study sample is divided into subsamples. In municipalities with populations between 500 and 3500 inhabitants, unemployment exhibits a negative impact on new firm establishment, aligning with the labor market approach (Evans and Jovanovic 1989). This finding suggests that high unemployment may suppress entrepreneurship within specific population ranges.

Economic geography theories emphasize the impact of geography on firm location decisions (Brekke 2015; Sorenson 2018; Stuart and Sorenson 2003). However, the interaction between geographical and sociodemographic factors in entrepreneurship remains elusive (Freire-Gibb and Nielsen 2014). Our second contribution lies in quantifying the relative importance of these factors for business creation. We developed a parsimonious model with just four key variables—low working-age population, low Internet coverage, high altitude, and population decline—explaining a significant portion of the variance in business entry. Notably, the working-age population emerges as the most critical factor, accounting for 69.8% of the variance. Municipalities facing adverse geodemographic challenges, including remoteness, altitude, and population decline, experience a lower likelihood of business creation. However, this negative impact is mitigated in areas with larger populations. This population buffering effect suggests that a sufficient population base can facilitate entrepreneurship despite geographical limitations. Our study thus unveils a moderating effect of population size on the relationship between geodemographic constraints and entrepreneurial activity, offering a valuable theoretical insight. By examining the interactions between sociodemographic and geographical variables, our study contributes by elucidating the ways in which altitude and slope can exacerbate the challenges faced by rural entrepreneurs.

As a third contribution, we provided a suitable method to measure rural liability at the municipal level. Defining rurality is not an easy task (Pato and Teixeira 2016), with population thresholds often inconsistently applied (Nelson and Nguyen 2023). Our model, considering both geographic and sociodemographic characteristics, predicts business creation to quantify rural liability. We identify precise thresholds for key factors that capture rural disadvantages, surpassing simplistic population-based measures. Furthermore, our model estimates the probability of business creation in each municipality. This predictive power allows us to quantify the municipal-level liability of rurality. By integrating this data, policymakers can estimate business creation probability in a municipality and tailor support strategies. The thresholds and relationships identified within our model hold substantial value for governments, enabling the precise identification of disadvantaged municipalities. This, in turn, facilitates the implementation of targeted and effective public interventions.

2 | Literature Review and Model Selection

2.1 | Rural Entrepreneurship

Rural entrepreneurship has attracted significant research interest due to its critical role in fostering economic development and sustaining rural communities (Gashi Nulleshi and Tillmar 2022). Early seminal work by Wortman (1990) laid the foundation for investigating the value of rural entrepreneurship for economic development. Subsequent systematic literature reviews have identified core topics and findings within this domain (Aguilar 2021; Fortunato 2014; Gashi Nulleshi and Tillmar 2022). Scientometric approaches have facilitated the visualization of interconnections within this literature corpus (Pato and Teixeira 2016; Shrivastava and Kumar Dwivedi 2021). Common research themes in rural entrepreneurship include the conceptualization of its unique characteristics, the role of rural entrepreneurial ecosystems, the influence of context-specific factors on business entry and success, and the evaluation of public policies designed to foster rural entrepreneurship.

A recurrent research topic is the definition and conceptualization of rural entrepreneurship. Delineating rurality is complex (Pato and Teixeira 2016), as demographic criteria are subject to varying thresholds, while socioeconomic or performance typologies based on institutional, social, economic, and environmental indicators are frequently employed. An effective conceptualization of rural entrepreneurship transcends mere firm creation in rural areas, encompassing enterprises that utilize local resources, employ locals, and contribute to rural income generation (McElwee and Smith 2014). Beyond geographic location and agriculture, Aguilar (2021) highlighted rural cultural values, natural resource endowment, rural poverty, and peripheral location as distinctive characteristics. While Babb and Babb (1992) found no major psychological trait differences between rural and urban entrepreneurs, the determinants explaining business resilience diverge across these contexts (Brewton et al. 2010).

Research has adapted the entrepreneurial ecosystems framework (Stam and Van de Ven 2021) to rural contexts (Aguilar 2021; del Olmo-García et al. 2023; Miles and Morrison 2020), acknowledging the necessity of considering contextual nuances for effective application in rural settings. Each rural area possesses unique characteristics concerning its proximity to urban centers, demographic composition, and spatial features (Gashi Nulleshi and Tillmar 2022), thereby justifying the exploration of strategies employed by specific rural ecosystems to overcome constraints and the proposition of initiatives for fostering vibrant entrepreneurial communities. Key components of rural entrepreneurial ecosystems encompass human capital, networks, entrepreneurial culture, financial systems, governance, infrastructure, environmental resources, and markets. A dearth of human capital can impede business growth in rural ecosystems with sparse populations, but it can also attract professionals seeking an improved work-life balance (Miles and Morrison 2020). Inadequate government policies risk disadvantaging rural startups, yet tailored incentives and financial support for rural areas can provide impetus (Barke and Newton 1997). The relative scarcity of investors in rural areas compared to urban centers underscores the critical role of bank access in fostering entrepreneurial ecosystems (del Olmo-García et al. 2023). While small markets in rural areas may restrict customer bases, niche markets and distinctive value propositions can effectively address local needs. Furthermore, advancements in e-commerce and digital finance hold promise for expanding customer reach and supporting revitalization endeavors in rural environments (Xu, Zhong, and Dong 2024).

Numerous studies have investigated the determinants of entrepreneurial success in rural contexts (Habersetzer et al. 2021; Gyimah and Lussier 2021; Miles and Morrison 2020). Habersetzer et al. (2021) highlighted the significance of resource accessibility, including supportive local networks and robust community integration. Gyimah and Lussier (2021) identified critical factors such as capital, industry experience, staffing proficiency, and marketing capabilities. Miles and Morrison (2020) associated success with framework conditions, encompassing natural capital, institutional frameworks, entrepreneurship-enabling cultures, infrastructure, and demand, as well as systemic conditions like the presence of enterprising individuals, entrepreneurial networks, human capital, access to financial resources, and support services like incubation and accelerators.

The literature review underscores the necessity of entrepreneurship policies for bolstering rural entrepreneurship (Pato and Teixeira 2016), with one-size-fits-all and fragmented policies lacking a coordinated vision deemed ineffective (Fortunato 2014). Such efforts may be among the few economic development strategies with demonstrably positive outcomes for rural regions (Stephens and Partridge 2011). However, evaluations of these policies have often lacked rigor and proper controls (Goetz et al. 2010), a shortcoming that persists according to the OECD's (2023) recent framework. Active involvement of rural stakeholders, including communities, enterprises, and economic agencies, is paramount (North and Smallbone 2006). Policymakers and practitioners should consider these factors when designing support measures for rural entrepreneurship and regional development.

The literature review reveals significant shifts in rural entrepreneurship research trends (Aguilar 2021; Pato and Teixeira 2016; Shrivastava and Kumar Dwivedi 2021). Pato and Teixeira (2016) noted an increased focus on economic and business aspects, particularly in hospitality and tourism, while spatial studies' prominence slightly diminished, albeit retaining relevance. Sociological and social science perspectives received less attention, as macro-level topics yielded to micro-level considerations like entrepreneurs' traits and organizational characteristics. The development of theoretical frameworks has received limited attention (Pato and Teixeira 2016); thus, analyzing nascent theories in rural spatial contexts presents an intriguing avenue for future research (Shrivastava and Kumar Dwivedi 2021). Additionally, Aguilar (2021) suggests investigating statistical methods to assess rurality's effect on ecosystem performance, including business survival, financial performance, and growth.

2.2 | Business Creation Theories

Several theories explain the establishment of businesses in specific locations, particularly emphasizing the challenges faced by rural areas (Sternberg 2022). Noteworthy theories in this context include economic geography (Brekke 2015; Isard 1949; Sorenson 2018; Stuart and Sorenson 2003), urban agglomeration (Rosenthal and Strange 2004), organizational ecology (Hannan and Freeman 1977), resource dependence (Pfeffer and Salancik 1978), Marshallian theories of agglomeration (Marshall 1890), knowledge spillover (Audretsch and Feldman 1996; Audretsch and Keilbach 2007), industrial dynamics (Forrester 1997), entrepreneurship ecosystem (Wurth, Stam, and Spigel 2022), and human capital (Unger et al. 2011) theories. These theories tend to support a negative relationship between rurality and entrepreneurship but differ in the justifications they offer.

Geography plays a pivotal role in determining where businesses choose to locate, with early economic geography theories emphasizing factors like proximity to raw materials and consumers (Isard 1949). Minimizing transportation costs remains a key consideration in location decisions. Consequently, large municipalities with strategic geographical advantages, such as proximity to markets, transportation hubs, or resources, tend to be more attractive for business establishment (Sorenson 2018; Stuart and Sorenson 2003). Place dependence refers to the idea that economic activities and businesses are influenced by the geographic location in which they are situated (Brekke 2015). Evolutionary economic geography focuses on how economies and economic structures evolve and develop over time and space, with a particular emphasis on understanding how geographic and spatial factors influence economic development. The theory of agglomeration with heterogeneous agents explores how the concentration of businesses and industries in specific regions influences economic dynamics (Behrens, Duranton, and Robert-Nicoud 2014). This interaction among various entities can foster benefits like knowledge exchange, yet it can also escalate competition and expenses due to congestion and business rivalry. Building upon this theory, Modrego, Atienza, and Hernández (2023), in their examination of Chilean businesses, noted heightened rates of business entry in larger regions. Nevertheless, they found an inconsistent correlation between the growth rates of early-stage businesses and factors linked to agglomeration. This hints that the downsides of agglomeration might outweigh its benefits, suggesting that exploring growth opportunities in peripheral areas may not pose a disadvantage for businesses. In fact, new business survival rates are as high in rural areas as they are in urban areas (Buss and Lin 1990).

The knowledge spillover theory of entrepreneurship posits that knowledge and information are not static within organizations or regions; instead, they can flow or spill over from one entity to another (Audretsch and Keilbach 2007). Therefore, it acknowledges the significant role of the geographical context in entrepreneurship. The proximity of businesses and workers facilitates the exchange of ideas and knowledge, promoting business creation in densely populated areas (Audretsch and Feldman 1996). Conversely, being situated in sparsely populated areas with limited infrastructure can pose challenges for starting any type of business (Huiban 2011).

Industrial dynamics theories (Forrester 1997) focus on how industries evolve over time, emphasizing the roles of innovation and competition in influencing business entry and exit. Industrial clustering theory (Audretsch and Feldman 1996) suggests that businesses in similar industries tend to cluster together, benefiting from shared resources, knowledge spillovers, and supply chain efficiencies. This fact disadvantages isolated rural environments, to the advantage of cities.

The dynamics of agglomeration economies can be analyzed from diverse perspectives. Among the various factors influencing start-up patterns, Marshallian factors such as labor market pooling, knowledge spillovers, and particularly input sharing, appear to have a significant impact on business entry (Dong 2020). The theory of urban agglomeration (Rosenthal and Strange 2004) posits that agglomeration economies, such as knowledge spillovers, access to a large labor market, and input sharing, are less pronounced in rural or small-town areas compared to larger urban centers. In cities, the presence of investors, universities, and suppliers creates an environment conducive to entrepreneurship, whereas rural regions scarcely benefit from these agglomeration economies. However, proximity to urban centers does not always have a uniformly positive impact on business creation; its effects can vary based on business types (Lavesson 2018).

The organizational ecology theory (Hannan and Freeman 1977) explores how organizational populations change over time through stages such as founding, growth, transformation, decline, and dissolution. It posits that organizations compete for resources and that rates of entry, exit, and diversity of organizational forms are influenced more by impersonal forces than individual intervention. In the context of rural entrepreneurship, this theory suggests that villages often have limited resources compared to urban areas. These resource disparities can affect capital access, skilled labor availability, infrastructure, and services. Building upon Stinchcombe's imprinting theory (1965), Clausen (2020) offers additional arguments on the difficulty of setting up businesses in rural areas. Limited access to updated social technology in rural areas can pose challenges for new businesses, hindering access to financial resources, university connections, and community dynamics. These factors associated with imprinting can discourage aspiring entrepreneurs in rural areas.

Resource dependence theory (Pfeffer and Salancik 1978) suggests that environmental uncertainty and external dependencies constrain all types of organizations. This theory focuses on critical resources obtained from external sources and explains how organizations adapt their behavior based on external resource availability. Applied to rural entrepreneurship, it suggests that rural areas often have limited resources compared to urban regions, which can affect various aspects of business creation. The human capital theory (Unger et al. 2011) underscores the importance of a well-educated and skilled labor force for economic development. Specialized skills and knowledge are essential for businesses, and these prerequisites are often challenging to fulfill in rural entrepreneurship. According to the entrepreneurship ecosystem theory (Wurth, Stam, and Spigel 2022), regulatory environments, access to funding, mentorship, and networking opportunities play a key role in supporting entrepreneurial activity. These prospects are less favorable in rural areas, despite the strong social bonds that rural communities develop.

2.3 | A Model for Business Creation

Building on the theoretical framework, we present the following model to explain municipal business creation (Equation 1).

$$Y_i = \sum_{j=1}^{J} \beta_j X_{ij} + \sum_{k=1}^{K} \lambda_k Z_{ik} + \sum_{l=1}^{L} \gamma_l R_{il} + \varepsilon_i \qquad (1)$$

where y_i is business creation for municipality i, X_{ij} is a vector of sociodemographic indicators (e.g., population, density, and age), Z_{ik} is a vector of infrastructure indicators (e.g., availability of nearby bus and train stations and Internet connectivity), and R_{il} is a vector of geographical indicators (e.g., latitude and slope).

Sociodemographic indicators, such as low population and population density, reflect challenges for businesses (Audretsch and Feldman 1996; Clausen 2020). Factors like birth and death rates, migration, gender, and age (Deller, Kures, and Conroy 2019; Kalantaridis and Bika 2006; Rupasingha and Marré 2020) can hinder access to an adequate workforce. Additionally, rural areas grapple with population decline and an aging demographic, impacting labor supply and demand dynamics. Social elements include dimensions such as social capital, culture, and governance (Beer 2014; Bu and Liao 2022), while economic considerations involve unemployment rates (del Olmo-García et al. 2023). The impact of unemployment on business entry varies depending on how business entry is measured (Audretsch and Fritsch 1994). While unemployment theoretically makes hiring easier, the scarcity of employment opportunities often leads rural residents to migrate to urban centers in search of better prospects. In-migration is essential for rural economic development, bringing fresh ideas, connections, and innovative approaches to entrepreneurship (Deller, Kures, and Conroy 2019). A municipality experiencing a negative balance in residential variations may reflect a limited capacity to attract migrants, thereby adversely impacting a company's ability to recruit employees.

The absence of critical infrastructure, including roads, railways, and ports, acts as a deterrent for business initiation in rural areas (Fox and Porca 2001; Huiban 2011). Van Leuven, Low, and Hill (2023) highlighted the significance of transportation infrastructure accessibility for the survival of manufacturing, transportation, wholesale, and warehousing sectors. They emphasized that even the type of highway plays a key role in this aspect. Furthermore, the lack of technological infrastructure, such as communications networks in rural regions, aggravates the challenges for new companies, making the adoption of information and communication technologies difficult (Deller et al. 2022; Galloway, Sanders, and Deakins 2011; Kim and Orazem 2017).

Physical factors, including geographical location, natural resources, and landscape, play a significant role (Eriksson and Rataj 2019; Laurin, Pronovost, and Carrier 2020; Müller and Korsgaard 2018; Stearns et al. 1995). Furthermore, the geographical remoteness from densely populated areas and the capital city creates obstacles in reaching potential consumers and engaging with government decision-making bodies (Huiban 2011). Geographical considerations are of utmost importance, especially in areas with elevated altitudes and steep slopes in mountainous terrain, although it is worth noting that numerous cities are positioned on plateaus or elevated terrain.

3 | Empirical Study

3.1 | Sample and Data

We examined the municipalities where 1,878,585 Spanish companies are headquartered, evaluating the attributes of all 8131 Spanish municipalities. We used data from the Iberian Balance Sheet Analysis System (SABI) database, widely disseminated by Moody's. SABI gathers data from the Spanish commercial register, which is mandatory for all companies, regardless of size. Therefore, SABI covers accurately the population of Spanish companies. SABI retains information even if a company goes bankrupt or is absorbed. We downloaded information for all companies, including the municipality where they are headquartered, industry categorized according to the Statistical Classification of Economic Activities in the European Community (NACE) code, establishment date, active or inactive status, and the most recent year for which financial data were available. The establishment date allowed us to determine the number of companies founded in each municipality since 2012. In Spain, approximately 40,000 companies are established annually. Subsequently, demographic and geographic data for each municipality were sourced from official institutions like the Spanish Institute of Statistics (INE). Table 1 presents the variables and their sources. As some demographic variables had data available only until 2021, we selected that year as the latest available. Specifically, 407,877 Spanish companies were established from 2012 to 2021.

We used two common approaches to measure business creation: the number of new companies per working-age population (NC/POP) and the number of new companies relative to the total number of companies (NC/TC) in the area (Audretsch and Fritsch 1994). The first measure aligns with the labor market approach, following the entrepreneurial choice framework developed by Evans and Jovanovic (1989). The second measure, influenced by the organizational ecology theory of Hannan and Freeman (1977), standardizes the number of new entrants relative to the number of existing firms at the start of the period, or an average of firms at both the beginning and end of the relevant time frame. The choice of approach can significantly affect the results. For instance, the labor market approach suggests a negative influence of unemployment on new firm establishment (Audretsch and Fritsch 1994), while the ecological approach indicates a positive correlation between unemployment and start-up activity.

When using both dependent variables in the context of small municipalities, a statistical problem arises. Since most municipalities have very low populations and few businesses, the creation of even one company in a year can significantly distort municipal business creation rates. This can result in attributing higher business creation rates to a village than to a large capital city, creating a misleading representation of their entrepreneurial activities. These observations will not only behave as outliers, but also as influential observations, which have the potential to distort the findings. Consequently, we opted against the standard panel data methodology, choosing a cross-sectional design instead. This involved computing the median value for each business creation rate from 2012 to 2021, ensuring robustness and reliability. Thus, NC/POP and NC/TC represent the median business creation rates over a 10-year period. We introduced a dummy variable (FOUNDED) that equals 0 when NC/POP is 0 and 1 otherwise. Obviously, using NC/TC would get the same result for FOUNDED.

The independent variables include working-age population (POP), population density (DENS), population change since the year 2000 (POPCHANGE), average age (AGE), unemployment rate of the municipality (UR), net migration to population ratio (MIGRATION), population within a radius of 50 km (POP50KM), distance to the capital (DISTCAP), distance to the nearest municipality with more than 50,000 inhabitants (DIST50000), altitude of the municipality (ALTITUDE), average geographical slope of the entire municipality (SLOPE), percentage of homes in the municipality with 100 Mb coverage (COVERAGE), and number of transportation infrastructures (bus, train, light rail, and metro stops) within a 50 km radius of the municipality (INFRA50). We included three control variables: agricultural business intensity (AGRIC); industrial business intensity (INDUSTRY); and service business intensity (SERVICES). AGRIC is the ratio of the number of agricultural sector companies in a municipality to the total number of companies in that municipality. Likewise for INDUSTRY and SERVICES. We computed the median of the independent and control variables from 2012 to 2021, where they vary.

TABLE 1 | Variables and definitions.

Dependent variables	
NC/POP	New companies in the municipality to population in the municipality. *Source:* Iberian Balance Sheet Analysis System (SABI) database and Spanish Institute of Statistics (INE)
NC/TC	New companies in the municipality to total companies in the municipality. *Source:* SABI database and INE
FOUNDED	A dummy variable that equals 0 when NC/POP is 0 and 1 otherwise
Independent variables	
POP	Population of the municipality in which the company's headquarters are located. We used population from 15 to 64 years old as a proxy for working-age population. *Source:* INE
DENS	Population density of the municipality in which the company's headquarters are located. *Source:* INE
POPCHANGE	Population change of the municipality in which the company's headquarters are located since 2000. *Source:* INE
AGE	Average age of the population in the municipality where the company's headquarters are located. *Source:* INE
UR	Unemployment rate of the municipality in which the company's headquarters are located. *Source:* INE
MIGRATION	Net migration rate. The average balance of people migrating to or from a municipality (i.e., new arrivals—departures), divided by the population of the municipality (in thousands). *Source:* INE
POP50KM	Number of inhabitants in 50 km around the municipality. *Source:* Global Human Settlement Layer (GHSL 2015) database
DISTCAP	Distance in a straight line to the capital of the province. *Source:* Open Street Map
DIST50000	Closest distance to a municipality with more than 50,000 inhabitants. *Source:* Open Street Map
ALTITUDE	Altitude of the municipality above sea level. *Source:* Spanish Geographical Institute (IGN)
SLOPE	Average geographic slope of the entire municipality. *Source:* Zúñiga-Antón et al. (2022)
COVERAGE	Average percentage of homes in the municipality with 100 Mb coverage from 2013 to 2020. *Source:* Spanish Secretariat for Telecommunications and Digital Infrastructures (SETELECO)
INFRA50	Number of bus, train, light rail, and subway stops within a 50 km radius of the municipality. *Source:* Open Street Map
Control variables	
AGRIC	Agricultural business intensity. Agricultural sector companies in a municipality to the total number of companies in that municipality. *Source:* SABI database and INE
INDUSTRY	Industrial business intensity. Industrial sector companies in a municipality to the total number of companies in that municipality. *Source:* SABI database and INE
SERVICES	Service sector business intensity. Service sector companies in a municipality to the total number of companies in that municipality. *Source:* SABI database and INE

3.2 | Exploratory Analysis

There are a total of 8131 municipalities in Spain. We excluded 23 from the sample due to municipal mergers resulting in unavailable data for all years. Table 2 provides a descriptive overview of the variables, including the mean, median, quartiles, minimum, maximum, and standard deviation. Table 3 presents both Pearson and Spearman correlation coefficients. Concerning

TABLE 2 | Descriptive statistics of the variables.

Variable	Obs.	Mean	Median	Q1	Q3	Std. dev.	Min	Max
NC/POP	8108	0.040	0.000	0.000	0.068	0.071	0	1.001
NC/TC	8108	1.243	0.000	0.000	2.77	1.879	0	14.26
POP	8108	5748.1	539.625	165.31	2418.69	46,908.5	4	3,210,898
DENS	8108	175.781	13.795	4.928	55.147	892.846	0.25	22,841.67
AGE	8108	44.3	43.899	42.46	45.86	2.502	35.04	50.10
UR	8108	6.133	5.684	3.806	7.952	3.926	0	100
POPCHANGE	8108	0.026	−0.083	−0.242	0.132	0.561	−0.71	17.87
COVERAGE	8107	0.164	0.000	0.000	0.2300	0.261	0	1
MIGRATION	8108	0.113	0.058	−0.549	0.743	1.408	−9.39	24.86
PEOPLE50	8108	711,393.5	314,707.5	144,524.8	780,409.3	1,109,109	10,800	6,608,981
INFRA50	8108	1023.5	433	109.00	1090.00	1760.2	5	10,46
DISTCAP	8108	43.694	40.206	26.110	58.425	24.233	0	220.54
DIST50000	8108	43.24	37.88	21.76	60.17	28.12	0	220.53
SLOPE	8108	13.207	10.315	332.25	858.00	10.080	0.87	60.39
ALTITUDE	8108	612.96	665.5	5.240	18.630	343.968	1	1695
AGRIC	8108	0.098	0.000	0.000	0.0927	0.204	0	1
INDUSTRY	8108	0.301	0.283	0.000	0.453	0.281	0	1
SERVICES	8108	0.420	0.483	0.000	0.659	0.325	0	1

the dependent variables, business creation is scarce or nonexistent in most municipalities, leading to a right-skewed distribution. The data reveal significant variations in working-age population and density, primarily due to a few highly populated municipalities and a long tail of small municipalities. High correlation coefficients were observed among certain independent variables, prompting the detection of multicollinearity through the variance inflation factor (VIF). Consequently, DENS, PEOPLE50, and DIST50000 were removed from the analysis. The Spearman coefficient shows a stronger correlation between the dependent and independent variables than the Pearson coefficient. This observation suggests the existence of nonlinear relationships between the variables and underscores the potential benefit of variable transformation. We performed a logarithmic transformation of the POP variable. We address nonlinearity in further analyses.

We aimed to explore thresholds for the independent variables that best capture rural disadvantages. A first attempt was made performing a set of univariate logistic regressions, with FOUNDED as the dependent variable and each independent variable as a predictor. Logistic regression was used to estimate the probability that a company will be founded in a municipality, given the value of the predictor. We ordered each municipality from the least to the greatest probability of creating companies. Table 4 provides the threshold for each independent variable, above which the probability of founding a company in a 10-year period exceeds the cutoff point. The cutoff point was 35.2%, which is the percentage of municipalities that created companies between 2012 and 2021. For example, the threshold for working-age population is 1334 inhabitants, for altitude, 559 m, and for the distance to the capital, 42.11 km. Municipalities with values below these thresholds can be considered to be at a severe disadvantage in Spain. However, although all the variables showed statistical significance ($p < 0.05$), the goodness of fit measured by the pseudo R-squared varied greatly, ranging from 0.004 for SLOPE to 0.790 for POP. The table also displays the predictive power for determining whether a municipality will create businesses, with the accuracy or success rate obtained through the bootstrap method. The highest accuracy is achieved in the model that includes POP (90.7%), and the lowest for MIGRATION (62.9%). The table also includes the value of Exp(B), which represents how the odds of an event (in this case, a municipality creating businesses) change for a one-unit change in the predictor variable. Although the results are promising, we next adopt a multivariable approach because it better captures the complex relationships between variables.

3.3 | Multivariable Regression Analysis

A substantial proportion of Spanish municipalities did not experience the foundation of any new businesses during the analyzed years, representing a case of limited dependent variables (LDVs) with a lower bound of zero. This feature requires special consideration in statistical analysis, making traditional regression models less suitable (Cragg 1971). To address this, we employed a two-step modeling approach (Belotti et al. 2015). First, a logistic regression model was employed to forecast the probability of a positive outcome, determining whether a municipality established at least one business. Subsequently, a linear multivariable regression model was applied specifically among municipalities with positive outcomes, that is, those that initiated business ventures. Nevertheless, we conducted an overall linear regression model using the entire data set, which

TABLE 3 | Pearson and Spearman correlation matrix. Pearson (Spearman) correlation coefficients are reported below (above) the diagonal.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
NC/POP (1)	1	0.970**	0.747**	0.673**	−0.276**	0.313**	0.543**	0.666**	0.218**	0.432**	0.359**	−0.226**	−0.387**	−0.059**	−0.494**	0.292**	0.225**	0.377**
NC/TC (2)	0.794**	1	0.760**	0.677**	−0.301**	0.353**	0.535**	0.668**	0.209**	0.439**	0.357**	−0.226**	−0.393**	−0.055**	−0.499**	0.291**	0.215**	0.386**
POP (3)	0.559**	0.701**	1	0.843**	−0.335**	0.515**	0.602**	0.760**	0.184**	0.508**	0.393**	−0.200**	−0.432**	−0.011	−0.608**	0.361**	0.383**	0.523**
DENS (4)	0.175**	0.239**	0.344**	1	−0.349**	0.435**	0.680**	0.729**	0.254**	0.648**	0.557**	−0.331**	−0.573**	−0.083**	−0.676**	0.253**	0.358**	0.464**
AGE (5)	−0.193**	−0.265**	−0.306**	−0.114**	1	−0.223**	−0.475**	−0.371**	−0.199**	−0.456**	−0.319**	0.018	0.228**	−0.142**	0.334**	−0.089**	−0.132**	−0.205**
UR (6)	0.121**	0.237**	0.370**	0.086**	−0.178**	1	0.269**	0.338**	0.018**	0.203**	0.086**	−0.009	−0.207**	−0.082**	−0.308**	0.135**	0.206**	0.266**
POPCHANGE (7)	0.303**	0.35**	0.359**	0.106**	−0.279**	0.102**	1	0.577**	0.426**	0.574**	0.525**	−0.329**	−0.493**	0.021*	−0.483**	0.173**	0.261**	0.378**
COVERAGE (8)	0.480**	0.555**	0.730**	0.431**	−0.27**	0.200**	0.303**	1	0.200**	0.507**	0.411**	−0.242**	−0.401**	0.028*	−0.547**	0.232**	0.269**	0.391**
MIGRATION (9)	0.137**	0.136**	0.124**	0.043**	−0.152**	−0.001	0.293**	0.103**	1	0.289**	0.311**	−0.188**	−0.253**	0.042**	−0.212**	0.024*	0.074**	0.160**
PEOPLE50 (10)	0.319**	0.351**	0.420**	0.300**	−0.335**	0.081**	0.408**	0.432**	0.188**	1	0.911**	−0.524**	−0.775**	0.059**	−0.582**	0.126**	0.245**	0.296**
INFRA50 (11)	0.302**	0.328**	0.393**	0.278**	−0.283**	0.056**	0.400**	0.407**	0.190**	0.970**	1	−0.564**	−0.756**	0.102**	−0.460**	0.067**	0.194**	0.251**
DISTCAP (12)	−0.175**	−0.176**	−0.175**	−0.185**	−0.023**	0.028**	−0.226**	−0.250**	−0.106**	−0.290**	−0.299**	1	0.715**	0.167**	0.208**	−0.057**	−0.088**	−0.105**
DIST50000 (13)	−0.274**	−0.313**	−0.120**	−0.211**	0.163**	−0.129**	−0.300**	−0.343**	−0.166**	−0.448**	−0.433**	0.693**	1	0.073**	0.476**	−0.112**	−0.186**	−0.247**
SLOPE (14)	−0.067**	−0.061**	−0.031**	−0.076**	−0.111**	−0.073**	−0.049**	−0.036**	0.022*	−0.008**	0.043**	0.190**	0.091**	1	0.017	−0.196**	0.033**	0.054**
ALTITUDE (15)	−0.396**	−0.464**	−0.597**	−0.231**	0.314**	−0.227**	−0.238**	−0.484**	−0.146**	−0.320**	−0.268**	0.187**	0.459**	0.033**	1	−0.252**	−0.238**	−0.323**
AGRIC (16)	−0.074**	−0.085**	−0.093**	−0.072**	0.080**	−0.039**	−0.087**	−0.131**	−0.038**	−0.119**	−0.121**	0.012	0.032**	−0.161**	0.015	1	0.065**	0.038**
INDUSTRY (17)	0.082**	0.089**	0.207**	0.007	−0.079**	0.107**	0.084**	0.095**	0.03**	0.098**	0.091**	−0.050**	−0.123**	0.039**	−0.145**	−0.166**	1	−0.083**
SERVICES (18)	0.289**	0.345**	0.504**	0.144**	−0.188**	0.195**	0.239**	0.317**	0.118**	0.225**	0.215**	−0.062**	−0.203**	0.067**	−0.318**	−0.217**	−0.165**	1

*$p < 0.05$, **$p < 0.001$.

is the standard approach. Table 5 presents the results of five multivariable regression models. The table displays unstandardized regression coefficients with *p*-values in parentheses.

The first column of Table 5 shows the results of a multivariable logistic regression model using FOUNDED as the dependent variable, which represents the first step within our two-step modeling approach. All independent variables displayed statistical significance at $p < 0.001$, with the exceptions being DISTCAP, which remained significant at $p < 0.01$, and AGE and POPCHANGE, which failed to achieve statistical significance. The findings underscore the model's strong fit, as indicated by a pseudo *R*-squared value of 0.825. This value suggests a strong ability to predict whether a municipality will create new businesses.

The second and third columns present the results of a multivariable linear regression using the same data set, with NC/POP and NC/TC as dependent variables. The goodness of fit measured by *R*-squared was 0.351 for NC/POP and 0.518 for

TABLE 4 | Rurality disadvantage thresholds, determined by logistic regressions that use the dependent variable FOUNDED (whether the municipality created companies) and each independent variable.

Variable	Threshold	Pseudo *R*-square	Accuracy	Coef	*p*-value	Exp(B)
POP	1334 inhabitants	0.790	90.7%	0.002	0.000	1.002
DENS	56 inhabitants per km^2	0.358	80.1%	0.015	0.032	1.015
POPCHANGE	0.01% increase	0.249	73.0%	2.992	0.000	19.920
AGE	43.9 years	0.107	64.4%	−0.257	0.000	0.773
UR	6.31% unemployed	0.105	65.8%	0.191	0.000	1.210
MIGRATION	0.16% inhabitants	0.025	62.9%	0.210	0.000	1.233
POP50KM	647,608 inhabitants	0.208	71.2%	0.000	0.000	1.000
DISTCAP	42.11 km	0.044	66.3%	−0.017	0.000	0.983
DIST50000	39.36 km	0.156	72.3%	−0.031	0.000	0.970
ALTITUDE	559 m	0.337	75.6%	−0.004	0.000	0.996
SLOPE	13.02°	0.004	78.4%	−0.012	0.000	0.988
COVERAGE	14% coverage	0.516	82.2%	9.111	0.000	9058.4
INFRA50	950 infrastructures	0.179	70.7%	0.001	0.000	1.001

TABLE 5 | Results of the multivariable regression models explaining company creation by sociodemographic and geographic factors.

	Full sample of municipalities			Subset of municipalities that founded companies	
	Model 1 FOUNDED	Model 2 NC/POP	Model 3 NC/TC	Model 4 NC/POP	Model 5 NC/TC
LnPOP	2.475 (0.000)	0.016 (0.000)	0.623 (0.000)	0.013 (0.028)	0.046 (0.050)
AGE	0.016 (0.483)	0.001 (0.573)	−0.025 (0.002)	0.001 (0.318)	−0.014 (0.239)
UR	−0.095 (0.000)	−0.001 (0.000)	−0.008 (0.006)	−0.006 (0.000)	0.053 (0.000)
POPCHANGE	0.192 (0.114)	0.011 (0.001)	0.307 (0.000)	0.001 (0.689)	0.203 (0.002)
COVERAGE	1.311 (0.000)	0.027 (0.000)	0.309 (0.000)	−0.001 (0.759)	−0.301 (0.222)
MIGRATION	0.125 (0.021)	0.002 (0.000)	0.023 (0.031)	0.011 (0.000)	0.092 (0.092)
INFRA50	0.102 (0.001)	0.001 (0.002)	0.009 (0.404)	−0.005 (0.335)	−0.045 (0.634)
DISTCAP	−0.004 (0.088)	−0.001 (0.139)	−0.002 (0.046)	−0.001 (0.035)	−0.003 (0.002)
ALTITUDE	−0.067 (0.000)	−0.006 (0.000)	−0.032 (0.000)	−0.003 (0.000)	−0.016 (0.009)
SLOPE	−0.017 (0.000)	−0.004 (0.000)	−0.006 (0.000)	−0.001 (0.000)	0.001 (0.869)
Business intensity	Yes	Yes	Yes	Yes	Yes
cte	−32.82 (0.000)	−0.047 (0.001)	−1.064 (0.002)	0.102 (0.001)	2.900 (0.000)
N obs.	8107	8107	8107	2854	2854
Pseudo *R*-squared	0.825				
Adj *R*-squared		0.351	0.518	0.120	0.079

Note: Model 1 is a logistic regression with FOUNDED as the dependent variable. Models 2 and 3 are linear regression models, with NC/POP and NC/TC as the dependent variables, respectively. Models 4 and 5 use the same dependent variables but using the subset of municipalities that founded companies. The table displays unstandardized regression coefficients with *p*-values in parentheses. Bootstrapped standard errors were clustered at the individual level.

NC/TC. Moving on to the fourth and fifth columns in Table 5, they depict the outcomes of the second step in the two-step modeling approach, which uses the sample of municipalities that established companies, with NC/POP and NC/TC as the dependent variables. The goodness of fit, as measured by *R*-squared, is significantly lower than that of the previous regression model, with values of 0.120 for NC/POP and 0.079 for NC/TC. COVERAGE, SLOPE, and INFRA50 were no longer significant. Thus, among municipalities that create businesses, the influence of sociodemographic and geographical factors on business creation is less pronounced.

The first research question focused on quantifying the importance of each predictor. In a regression analysis, the magnitude of estimated coefficients indicates how much the dependent variable is expected to change for a one-unit change in the dependent variable, with all other variables held constant. This change is scaled according to the original measurement units of the variable, making it difficult to compare the importance of different variables directly. To address this issue, standardized estimated coefficients (beta weights) are often used. However, when dealing with multiple linear regression models with correlated predictors, beta weights can sometimes lead to an overestimation or underestimation of the importance of an independent variable because beta weights do not consider the shared explained variance between variables in the regression. In other words, when predictors are correlated, they may collectively explain more of the variance in the dependent variable than they do individually. Therefore, the interpretation of standardized coefficients in a regression can be misleading (Bring 1994). Consequently, it is advisable to complement the analysis with other methods to gain insight into the relative importance of the independent variables (Tonidandel and LeBreton 2011). Tables 6 and 7 present unstandardized and standardized regression coefficients, the Pratt Relative Importance Index, dominance analysis results, and LASSO technique results. These analyses were conducted using NC/POP and NC/TC as the dependent variables, respectively.

The Pratt Relative Importance Index (Pratt 1987) assesses the importance of an independent variable by evaluating the proportion of explained variance it contributes, relative to the other independent variables in the model. It is calculated by multiplying the standardized regression coefficient by the zero-order correlation for the predictor and then dividing it by the total *R*-squared. Dominance analysis (Azen and Budescu 2003; Budescu 1993) expands its scope as it quantifies the unique variance contribution of variables in regression models, providing insights into their overall impact. Specifically, the general dominance statistic calculates the average additional unique variance contribution of each independent variable to all possible subsets of predictors within the model. It represents the average difference in model fit between subsets that include a specific variable and those that do not. The standardized statistics expresses the general dominance statistic value as a percentage of the overall fit statistic value (100%).

All four analyses showed that municipal working-age population is the most important factor in explaining business creation. The beta weight for LnPOP is the highest, 0.425, indicating that it made the largest contribution to the regression equation, holding all other predictor variables constant. LnPOP had a Pratt index of 0.698, indicating that 69.8% of the explained variance in NC/POP can be attributed to population. Note that the 10 independent variables explain 35% of the variance in NC/POP. The general dominance column can be interpreted in terms of the fit metric it decomposes. LnPOP has a value of 0.144, indicating that it is associated with an *R*-squared of approximately 14.4% of the variance in NC/POP, considering the predictive model and other variables. The standardized statistic was 0.411, so LnPOP explains 41.1% of the 100% of the explained variance in the dependent variable. Finally, the ranking column provides a rank ordering of independent variables based on their general dominance statistics. The next three strongest predictors of NC/POP are COVERAGE, ALTITUDE, and POPCHANGE, with combined contributions of 15% (7.8%, 4.8%, and 2.7%, respectively) to the total explained variation in

TABLE 6 | Linear regression model using the entire sample of municipalities and NC/POP as dependent variable.

NC/POP	Coef.	Beta	Pratt	General dominance	Standard. general dominance	Ranking	Lasso
LnPOP	0.016***	0.425	0.698	0.144	0.411	1	0.034
AGE	0.001	0.001	−0.001	0.008	0.022	7	
UR	−0.001***	−0.082	−0.029	0.006	0.018	8	
POPCHANGE	0.011***	0.088	0.078	0.027	0.077	4	0.003
COVERAGE	0.027***	0.099	0.140	0.078	0.222	2	0.003
MIGRATION	0.002***	0.032	0.013	0.005	0.013	9	
INFRA50	0.001**	0.027	0.024	0.021	0.059	5	
DISTCAP	−0.001	−0.009	0.006	0.012	0.034	6	
ALTITUDE	−0.006***	−0.087	0.102	0.048	0.136	3	−0.002
SLOPE	−0.004***	−0.051	0.010	0.003	0.008	10	
N obs.	8107						8107
R-squared	0.350						0.328
Lambda							0.048
CV mean pred. error							0.003

$^{**}p < 0.05$, $^{***}p < 0.001$.

TABLE 7 | Regression model using the entire sample of municipalities and NC/TC as dependent variable.

NC/TC	Coef.	Beta	Pratt	General dominance	Standard. general dominance	Ranking	Lasso
LnPOP	0.623***	0.628	0.860	0.243	0.473	1	1.169
AGE	−0.025***	−0.033	0.017	0.017	0.033	6	
UR	−0.008***	−0.018	−0.008	0.016	0.032	7	
POPCHANGE	0.307***	0.092	0.063	0.035	0.069	4	0.171
COVERAGE	0.309***	0.043	0.047	0.099	0.193	2	0.013
MIGRATION	0.023**	0.018	0.005	0.004	0.007	9	
INFRA50	0.009	0.009	0.006	0.022	0.044	5	
DISTCAP	−0.002**	−0.020	0.007	0.012	0.024	8	
ALTITUDE	−0.032***	−0.058	0.007	0.063	0.122	3	−0.128
SLOPE	−0.006***	−0.035	0.032	0.002	0.003	10	
N obs.	8107						8107
R-squared	0.518						0.509
Lambda							0.861
CV mean pred. error							1.736

$^{**}p < 0.05$, $^{***}p < 0.001$.

the dependent variable (35.2%). At the bottom, SLOPE, MIGRATION, UR, and AGE each contribute < 1% of *R*-squared, with a combined total of only 2.2%. The results present in Table 7 for the dependent variable NC/TC are similar to those previously discussed.

Finally, the implementation of the LASSO model selection (Tibshirani 1996) yielded a parsimonious model, presented in the final column. LASSO combines shrinkage and variable selection to streamline linear regression models and avoid overfitting. The LASSO results in the final column of Tables 6 and 7 reveal that a model with four variables (POP, POPCHANGE, COVERAGE, and ALTITUDE) explains the majority of the variance, accounting for 32.8% of the total variation in NC/POP (closely matching the 35.5% of the full model) and 50.9% of the total variation in NC/TC (closely matching the 51.8% of the full model).

3.4 | Decision Tree Analysis

We conducted a Ramsey test to check for nonlinearity between variables. The test results strongly indicated the presence of nonlinear relationships ($p = 0.000$). Instead of introducing complex nonlinear terms that might enhance model fit but complicate result interpretation, we opted to use decision tree techniques to capture these nonlinear relationships and establish precise thresholds that effectively capture rural disadvantage. The chi-squared automatic interaction detection (CHAID; Kass 1980) is the most widely used decision tree algorithm. CHAID uses the chi-square statistic to identify the most suitable splitting variable. We utilized exhaustive CHAID (Biggs, De Ville, and Suen 1991), a modified version that examines all possible combinations for each predictor to determine the most optimal splits. Some of the decision rules are presented in Table 8. These rules help to answer our second research question.

To make Table 8 informative, it is necessary to compare the probability of business creation in a municipality over a 10-year period (35.2%; Node 0) and the probabilities associated with each branch, which include municipalities with similar characteristics. Node 1 presents the outcomes for the initial branch of the decision tree. The findings confirm that POP is the most important variable, serving as the primary splitting criterion for the tree. Specifically, for municipalities with a working-age population ranging from 951 to 1737 residents, the probability of business creation stands at 45.4%. Nodes 2–4 shed light on the roles of variables amenable to public administration intervention—namely, INFRA50 and COVERAGE. In this municipality group ranging from 951 to 1737 residents, the probability of business creation decreases to 36.9% when INFRA50 is ≤ 433, while it increases to 68.5% when INFRA50 exceeds 2167. The probability of business creation rises to 74.6% when COVERAGE surpasses 18.2%.

Nodes 5–14 show the effect of geographic circumstances and the moderating role of population. Where the geography is adverse (extreme altitude, steep slope, distance from the capital), the probability of creating businesses decreases dramatically. However, if there is enough population, geographic difficulties are not so important. For example, Node 6 shows an extreme case: municipalities in high altitude locations (ALTITUDE > 1038) with a working-age population > 951 have a probability of creating businesses of 58.7%. This group includes tourist villages such as Albarracín, with 1043 working-age inhabitants. The same occurs with the slope and the distance to the capital—the mitigating effect of the population minimizes the difficulties.

The nodes from 15 to 20 highlight the influence of demographic circumstances, particularly the role of population. Similarly to geographic challenges, a sufficient population mitigates instances of population decline and negative net migration balance. For example, Node 15 shows that where the population

TABLE 8 | Decision rules for predicting municipalities that establish companies using the exhaustive CHAID algorithm. The dependent variable is FOUNDED.

Node	*N* (%)	Rule	Status	Probability	*N* (%)
0	8108 (100%)	Entire sample	Not-founded	64.8%	5254
			Founded	35.2%	2854
1	811 (10.0%)	IF 951 < POP ≤ 1737 THEN	Not-founded	54.6%	443
			Founded	45.4%	368
2	379 (4.7%)	IF 951 < POP ≤ 1737 AND INFRA50 ≤ 433 THEN	Not-founded	63.1%	239
			Founded	36.9%	140
3	92 (1.1%)	IF 951 < POP ≤ 1737 AND INFRA50 > 2167 THEN	Not-founded	31.5%	29
			Founded	68.5%	63
4	63 (0.8%)	IF 951 < POP ≤ 1737 AND 433 < INFRA50 ≤ 2167 AND COVERAGE > 0.182 THEN	Not-founded	25.4%	16
			Founded	74.6%	47
5	812 (10.0%)	IF ALTITUDE > 1038 THEN	Not-founded	93.1%	756
			Founded	6.9%	56
6	75 (0.9%)	IF ALTITUDE > 1038 AND POP > 951 THEN	Not-founded	41.3%	31
			Founded	58.7%	44
7	810 (10%)	IF ALTITUDE ≤ 95 THEN	Not-founded	13.2%	107
			Founded	86.8%	703
8	57 (0.7%)	IF ALTITUDE ≤ 95 AND POP < 539 THEN	Not-founded	86.0%	49
			Founded	14.0%	8
9	1620 (20.0%)	IF SLOPE > 20.89 THEN	Not-founded	70.9%	1148
			Founded	29.1%	472
10	228 (2.8%)	IF SLOPE > 20.89 AND POP > 3502	Not-founded	1.8%	4
			Founded	98.2%	224
11	1621 (20.0%)	IF SLOPE ≤ 4.49 THEN	Not-founded	61.7%	1000
			Founded	38.3%	621
12	3243 (40.0%)	IF DISTCAP > 47.10 THEN	Not-founded	73.8%	2393
			Founded	26.2%	850
13	313 (3.9%)	IF DISTCAP > 47.10 AND 951 < POP ≤ 1737 THEN	Not-founded	60.4%	189
			Founded	39.6%	124
14	810 (10.0%)	IF DISTCAP ≤ 15.38 THEN	Not-founded	32.2%	261
			Founded	67.8%	549

(Continues)

TABLE 8 | (Continued)

Node	*N* (%)	Rule	Status	Probability	*N* (%)
15	4594 (56.7%)	IF POPCHANGE ≤ −0.035 THEN	Not-founded	85.9%	3950
			Founded	14.1%	647
16	663 (8.2%)	IF POPCHANGE ≤ −0.035 AND POP > 1356 THEN	Not-founded	27.0%	179
			Founded	73.0%	484
17	3511 (43.3%)	IF POPCHANGE > −0.035 THEN	Not-founded	37.1%	1304
			Founded	62.9%	2207
18	1646 (20.3%)	IF MIGRATION ≤ −0.013 THEN	Not-founded	92.1%	1516
			Founded	7.9%	130
19	145 (1.8%)	IF MIGRATION ≤ −0.013 AND POP > 1303 THEN	Not-founded	34.5%	50
			Founded	65.5%	95
20	1501 (18.5%)	IF MIGRATION ≤ −0.013 AND POP ≤ 1303 THEN	Not-founded	97.7%	1466
			Founded	2.3%	35
21	810 (10.0%)	IF UR ≤ 0.023 THEN	Not-founded	96.3%	780
			Founded	3.7%	30
22	3249 (40.1%)	IF UR ≤ 0.050 THEN	Not-founded	83.9%	2725
			Founded	16.1%	524
23	810 (10.0%)	IF UR > 0.050 THEN	Not-founded	52.0%	2529
			Founded	48.0%	2330
24	2432 (30.0%)	IF UR > 0.074 THEN	Not-founded	46.8%	1139
			Founded	53.2%	1239
25	370 (4.6%)	IF 539 < POP ≤ 951 AND UR ≤ 0.057 THEN	Not-founded	75.9%	281
			Founded	24.1%	89
26	346 (4.3%)	IF 539 < POP ≤ 951 AND 0.057 < UR ≤ 0.105 THEN	Not-founded	83.2%	288
			Founded	16.8%	58
27	95 (1.2%)	IF 539 < POP ≤ 951 AND UR > 0.105 THEN	Not-founded	91.6%	87
			Founded	8.4%	8
28	689 (8.5%)	IF 1737 < POP ≤ 3502 AND UR ≤ 0.105 THEN	Not-founded	18.3%	126
			Founded	81.7%	563
29	122 (1.5%)	IF 1737 < POP ≤ 3502 AND UR > 0.105 THEN	Not-founded	33.6%	41
			Founded	66.4%	81

decreases by 3.5%, the probability of creating businesses is 14.1%, a percentage that increases to 73.0% if the working-age population exceeds 1356.

Nodes 21 to 29 show the effect of unemployment on business creation and the role of population. It is a very complex relationship. Unemployment can positively or negatively affect business entry (Audretsch and Fritsch 1994). Nodes 21–24 clearly show that the higher the unemployment rate, the higher the business creation, ranging from 2.5% when unemployment is ≤ 2.3% (Node 21) to 53.2% when unemployment exceeds 7.4%. However, in certain population ranges, the relationship between unemployment and business creation reverses. For example, Nodes 25–27 show that if the working-age population is between 539 and 951, the higher the unemployment, the lower the probability of business creation. Nodes 28 and 29 show another similar case, in this case if the population is between 1737 and 3502. Therefore, this data range exhibits a Simpson's paradox (Simpson 1951). This is not an exceptional finding when it comes to employment (Armstrong and Wattenberg 2014).

4 | Discussion and Conclusions

We developed a comprehensive model of business creation that considers the geographic and demographic characteristics of municipalities. Our model identifies the factors that contribute to rural liability and assesses their relative importance. We analyzed the characteristics of municipalities where all newly established companies in Spain are headquartered between 2012 and 2021. We examined the working-age population, population density, aging demographics, population variation, net migration, unemployment rates, distance from major markets, infrastructure availability (Internet coverage and transit accessibility), and challenging geographical conditions (municipal altitude and slope). Our findings show that low working-age population is the most important factor explaining rural disadvantage, followed by low Internet coverage, high altitude, and population decline. We found that geographic constraints significantly impede entrepreneurship, yet their influence is moderated by the working-age population within the municipality. Unlike ad hoc approaches that employ arbitrary population thresholds to assess rural disadvantage, our model employs municipal-level data to precisely identify these thresholds.

Municipal-level data pose challenges that require appropriate statistical methods. These challenges arise from frequent zeros for the dependent variable, as many municipalities in the sample did not create new businesses during the 10-year study period. Departing from conventional approaches, we employed a two-part regression model (Belotti et al. 2015) to effectively address this issue. The first part involves a logistic regression that predicts whether a municipality will or will not create businesses. Our model fit well, with a pseudo R-squared of 0.825. Therefore, this model can produce an index for rural liability, which can be used to estimate the number of companies that will be established in a municipality, if any. We found that all independent variables except age significantly explain business creation. The findings of two multivariate regressions employing distinct definitions of business creation rate as the dependent variable exhibited remarkable consistency. R-squared values ranged from 0.351 to 0.518.

The second part of the analysis involved a multivariable regression model that used only the sample of municipalities that successfully established businesses. The model fit decreased significantly in this part, with R-squared values ranging from 0.079 to 0.120. This is a significant finding, which suggests that sociodemographic and geographical factors have less influence in explaining business creation within municipalities that effectively establish new ventures, that is, the large municipalities. Adverse geographical conditions, like high altitude or steep slopes, along with sociodemographic factors, primarily pose challenges in small municipalities. Thus, for rural entrepreneurs, when it rains, it pours. This finding extends the literature on rural business entry by expanding theoretical models. Economic geography theories have indeed emphasized the undeniable impact of geography on firm location decisions (Brekke 2015; Sorenson 2018; Stuart and Sorenson 2003). However, the precise interplay between geographical and sociodemographic factors in entrepreneurship has remained elusive, despite some exploration on their interaction with personal factors (Freire-Gibb and Nielsen 2014). In addressing this gap, our study explains the mechanisms underlying these relationships. We found that the constraints to rural entrepreneurship posed by geographical challenges are particularly pronounced in areas with limited population size. Conversely, in regions boasting a substantial populace, elements such as high altitude, steep slopes, or inadequate infrastructure assume relatively diminished significance. Population presence mitigates or overcomes geographical challenges, facilitating the emergence of entrepreneurship. Our empirical study supports these theoretical arguments, thereby unveiling the mitigating role of the working-age population on the relationship between geography and entrepreneurship.

After establishing the significance of geography and demography in explaining rural liability, we quantified their relative importance. In addition to examining the standardized and unstandardized coefficients in our regression analyses, we employed dominance analysis (Budescu 1993) and the Pratt Relative Importance Index (Pratt 1987). Furthermore, we utilized the LASSO technique (Tibshirani 1996) to derive a parsimonious model. The results highlight the relative importance of working-age population, significantly surpassing all other independent variables. When using the Pratt index, 69.8% of the explained variance in business creation can be attributed to population when NC/POP is the dependent variable, and 84.2% when NC/TC is the dependent variable. The following variables in relative importance are Internet coverage, with values of 14.0% and 4.2%, altitude with 10.2% and 7.1%, and depopulation with 7.8% and 6.5%, respectively. Similar results were obtained when using dominance analysis.

Sociodemographic variables have received extensive attention in the literature due to their central role in explaining the rural–urban divide (Deller, Kures, and Conroy 2019; Rosenthal and Strange 2004; Rupasingha and Marré 2020). We should emphasize that within the parsimonious model, only the working-age population and population variation retained significance, while others, such as aging, unemployment, and migration balance, were excluded. On geographical challenges, in addition to examining proximity to metropolitan areas (Deller, Kures, and Conroy 2019; Rupasingha and Marré 2020), particular attention should be paid to the outcomes derived from altitude and the average slope of the municipality. The examination of these variables is noteworthy, as they have not heretofore been subjected to comprehensive analysis. We found that altitude is one of the four variables selected in the parsimonious model generated through the LASSO procedure, exhibiting a relative importance surpassing that of extensively studied variables—both geographic (e.g., distance to the capital) and demographic (e.g., aging, unemployment, or population decline)—in explaining business creation. Conversely, although the average slope of the municipality shows statistically significant associations with business creation, its relative importance is modest and thus is not incorporated into the parsimonious model.

The debate on the relationship between unemployment and business entry has a long pedigree (Audretsch and Fritsch 1994). The ecological approach posits a positive correlation between start-up activity and unemployment, while the labor market approach suggests that unemployment negatively influences the establishment of new firms. In our standalone analysis of unemployment rates and business creation, a positive correlation was observed. The positive relationship may be explained by the fact that small municipalities lack economic dynamism, prompting residents to migrate to larger cities where prospects are better. In rural communities, there is little incentive to wait for local job opportunities, which is why they exhibit lower unemployment rates than cities. However, a counterintuitive pattern emerges in municipalities with populations between 500 and 3500 inhabitants, where the relationship reverses, exhibiting a Simpson's paradox (Simpson 1951). A plausible explanation for the observed relationship may stem from the trade-off between entrepreneurial dynamism and the prospect of securing formal employment. The smallest villages typically lack opportunities in both spheres. In contrast, cities offer both entrepreneurial prospects and formal employment options. It is conceivable that in municipalities within the specified population range, entrepreneurial dynamism is subdued, hindering the creation of new businesses. However, residents in these municipalities may retain a degree of hope in finding employment within existing local businesses. This lingering hope could explain their decision to remain in their municipalities, leading to a rise in unemployment rates rather than pursuing emigration. However, further analysis is necessary to confirm these findings.

4.1 | Practical Implications

Public administrations aim to address the issue of rural depopulation. However, the definitions they frequently adopt regarding rural disadvantage tend to be arbitrary. Our model, constructed from sociodemographic and geographic indicators, can provide a comprehensive map of rural disadvantage in Spain, pinpointing the areas least favorable for entrepreneurship. These findings can contribute to enhancing public policies. By incorporating municipal data into our logistic regression model, government agencies can assess the likelihood of businesses being established and tailor appropriate support strategies. For example, they can estimate the impact of improving broadband access on business creation in a small municipality.

Subsequently, we employed decision trees due to their capacity to obtain easily interpretable rules, culminating in emergent policy recommendations. From these, we derived straightforward rules delineating thresholds and identifying instances of rural disadvantage. Notably, our research indicates that, among various infrastructures, Internet coverage significantly outweighs public transit infrastructure in its influence. This finding aligns with studies by Audretsch, Heger, and Veith (2015) on German start-ups, endorsing Spanish authorities to redirect public efforts towards enhancing such infrastructure in rural areas.

4.2 | Limitations of the Study and Future Lines of Research

Spain's local governance framework is characterized by a large number of very small municipalities, a pattern also found in the United States and several European countries, including Italy and France (Bel and Warner 2015). While rural depopulation represents a global trend, our findings are specifically tied to the Spanish context, which emphasizes the importance of replicating this study within other countries. The data set employed in this analysis is notably extensive, covering all Spanish companies and municipalities. However, the data are derived from the Spanish commercial registry, which may introduce discrepancies as some companies may register their headquarters in locations other than their actual operations. This could potentially introduce bias in studies that compare urban (the location of most headquarters) and rural regions (Hjaltadóttir, Makkonen, and Mitze 2020). Future research could delve into exploring the role of business creation as a tool to combat depopulation, focusing on identifying the most suitable sectors, survival factors, and success determinants. Investigating these aspects could shed light on effective strategies for fostering economic growth and mitigating depopulation in rural areas. Additionally, examining the regional variations and nuances within Spain or comparing these findings with other countries could offer comprehensive insights into the dynamics of rural entrepreneurship and depopulation on a broader scale.

Acknowledgments

This study was funded by the European Regional Development Fund (ERDF), the Spanish Ministry of Education and Science (codes PID2022-136818NB-I00 and PID2023-146084OB-I00), and the Government of Aragon (codes S38_23R and S33_23R).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

Acs, Z. J., S. Desai, and L. F. Klapper. 2008. "What Does 'Entrepreneurship' Data Really Show?" *Small Business Economics* 31, no. 3: 265–281. https://doi.org/10.1007/s11187-008-9137-7.

Agarwal, S., S. Rahman, and A. Errington. 2009. "Measuring the Determinants of Relative Economic Performance of Rural Areas." *Journal of Rural Studies* 25, no. 3: 309–321. https://doi.org/10.1016/j.jrurstud.2009.02.003.

Aguilar, E. C. 2021. "Rural Entrepreneurial Ecosystems: A Systematic Literature Review for Advancing Conceptualisation." *Entrepreneurial Business and Economics Review* 9, no. 4: 101–114. https://doi.org/10.15678/eber.2021.090407.

Argent, N. 2018. "Heading Down to the Local? Australian Rural Development and the Evolving Spatiality of the Craft Beer Sector." *Journal of Rural Studies* 61: 84–99. https://doi.org/10.1016/j.jrurstud.2017.01.016.

Armstrong, Z., and M. Wattenberg. 2014. "Visualizing Statistical Mix Effects and Simpson's Paradox." *IEEE Transactions on Visualization and Computer Graphics* 20, no. 12: 2132–2141. https://doi.org/10.1109/tvcg.2014.2346297.

Audretsch, D. B., and M. P. Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review* 86, no. 3: 630–640.

Audretsch, D. B., and M. Fritsch. 1994. "On the Measurement of Entry Rates." *Empirica* 21, no. 1: 105–113. https://doi.org/10.1007/bf01383974.

Audretsch, D. B., D. Heger, and T. Veith. 2015. "Infrastructure and Entrepreneurship." *Small Business Economics* 44, no. 2: 219–230. https://doi.org/10.1007/s11187-014-9600-6.

Audretsch, D. B., and M. Keilbach. 2007. "The Theory of Knowledge Spillover Entrepreneurship." *Journal of Management Studies* 44, no. 7: 1242–1254. https://doi.org/10.1111/j.1467-6486.2007.00722.x.

Azen, R., and D. V. Budescu. 2003. "The Dominance Analysis Approach for Comparing Predictors in Multiple Regression." *Psychological Methods* 8, no. 2: 129–148. https://doi.org/10.1037/1082-989x.8.2.129.

Babb, E. M., and S. V. Babb. 1992. "Psychological Traits of Rural Entrepreneurs." *Journal of Socio-Economics* 21, no. 4: 353–362. https://doi.org/10.1016/1053-5357(92)90004-q.

Barboza, G. 2024. "Missing Links of Knowledge Spillover Effects on Firm Intensity and Regional Development." *Small Business Economics* 63, no. 4: 1–25. https://doi.org/10.1007/s11187-024-00904-4.

Barke, M., and M. Newton. 1997. "The EU LEADER Initiative and Endogenous Rural Development: The Application of the Programme in Two Rural Areas of Andalusia, Southern Spain." *Journal of Rural Studies* 13, no. 3: 319–341. https://doi.org/10.1016/s0743-0167(97)00027-2.

Beer, A. 2014. "Leadership and the Governance of Rural Communities." *Journal of Rural Studies* 34: 254–262. https://doi.org/10.1016/j.jrurstud.2014.01.007.

Behrens, K., G. Duranton, and F. Robert-Nicoud. 2014. "Productive Cities: Sorting, Selection, and Agglomeration." *Journal of Political Economy* 122, no. 3: 507–553. https://doi.org/10.1086/675534.

Bel, G., and M. E. Warner. 2015. "Inter-Municipal Cooperation and Costs: Expectations and Evidence." *Public Administration* 93, no. 1: 52–67. https://doi.org/10.1111/padm.12104.

Belotti, F., P. Deb, W. G. Manning, and E. C. Norton. 2015. "Twopm: Two-Part Models." *STATA Journal* 15, no. 1: 3–20. https://doi.org/10.1177/1536867x1501500102.

Biggs, D., B. De Ville, and E. Suen. 1991. "A Method of Choosing Multiway Partitions for Classification and Decision Trees." *Journal of Applied Statistics* 18, no. 1: 49–62. https://doi.org/10.1080/02664769100000005.

Brekke, T. 2015. "Entrepreneurship and Path Dependency in Regional Development." *Entrepreneurship & Regional Development* 27, no. 3–4: 202–218. https://doi.org/10.1080/08985626.2015.1030457.

Brewton, K. E., S. M. Danes, K. Stafford, and G. W. Haynes. 2010. "Determinants of Rural and Urban Family Firm Resilience." *Journal of Family Business Strategy* 1, no. 3: 155–166. https://doi.org/10.1016/j.jfbs.2010.08.003.

Bring, J. 1994. "How to Standardize Regression Coefficients." *American Statistician* 48, no. 3: 209–213. https://doi.org/10.1080/00031305.1994.10476059.

Bu, D., and Y. Liao. 2022. "Land Property Rights and Rural Enterprise Growth: Evidence From Land Titling Reform in China." *Journal of Development Economics* 157: 102853. https://doi.org/10.1016/j.jdeveco.2022.102853.

Budescu, D. V. 1993. "Dominance Analysis: A New Approach to the Problem of Relative Importance of Predictors in Multiple Regression." *Psychological Bulletin* 114, no. 3: 542–551. https://doi.org/10.1037//0033-2909.114.3.542.

Buss, T. F., and X. Lin. 1990. "Business Survival in Rural America: A Three-State Study." *Growth and Change* 21, no. 3: 1–8. https://doi.org/10.1111/j.1468-2257.1990.tb00521.x.

Chen, W., Y. Lu, and G. Liu. 2022. "Balancing Cropland Gain and Desert Vegetation Loss: The Key to Rural Revitalization in Xinjiang, China." *Growth and Change* 53, no. 3: 1122–1145. https://doi.org/10.1111/grow.12568.

Cheng, S., and H. Li. 2010. "The Effects of Unemployment on New Firm Formation Revisited: Does Space Matter?" *Regional Science Policy & Practice* 2, no. 2: 97–120. https://doi.org/10.1111/j.1757-7802.2010.01022.x.

Clausen, T. H. 2020. "The Liability of Rurality and New Venture Viability." *Journal of Rural Studies* 73: 114–121. https://doi.org/10.1016/j.jrurstud.2019.12.005.

Cragg, J. G. 1971. "Some Statistical Models for Limited Dependent Variables With Application to the Demand for Durable Goods." *Econometrica: Journal of the Econometric Society* 39, no. 5: 829–844. https://doi.org/10.2307/1909582.

Deller, S., M. Kures, and T. Conroy. 2019. "Rural Entrepreneurship and Migration." *Journal of Rural Studies* 66: 30–42. https://doi.org/10.1016/j.jrurstud.2019.01.026.

Deller, S., B. Whitacre, and T. Conroy. 2022. "Rural broadband speeds and business startup rates." *American Journal of Agricultural Economics* 104, no. 3: 999–1025. https://doi.org/10.1111/ajae.12259.

del Olmo-García, F., I. Domínguez-Fabián, F. J. Crecente-Romero, and M. T. del Val-Núñez. 2023. "Determinant Factors for the Development of Rural Entrepreneurship." *Technological Forecasting and Social Change* 191: 122487. https://doi.org/10.1016/j.techfore.2023.122487.

Dong, Y. 2020. "Determinants of Entry: Evidence From New Manufacturing Firms in the US." *Growth and Change* 51, no. 4: 1542–1561. https://doi.org/10.1111/grow.12443.

Eriksson, R., and M. Rataj. 2019. "The Geography of Starts-Ups in Sweden. The Role of Human Capital, Social Capital and Agglomeration." *Entrepreneurship & Regional Development* 31, no. 9–10: 735–754. https://doi.org/10.1080/08985626.2019.1565420.

Evans, D. S., and B. Jovanovic. 1989. "An Estimated Model of Entrepreneurial Choice Under Liquidity Constraints." *Journal of Political Economy* 97, no. 4: 808–827. https://doi.org/10.1086/261629.

Forrester, J. W. 1997. "Industrial Dynamics." *Journal of the Operational Research Society* 48, no. 10: 1037–1041. https://doi.org/10.1057/palgrave.jors.2600946.

Fortunato, M. W. P. 2014. "Supporting Rural Entrepreneurship: A Review of Conceptual Developments From Research to Practice." *Community*

Development 45, no. 4: 387–408. https://doi.org/10.1080/15575330.2014.935795.

Fox, W. F., and S. Porca. 2001. "Investing in Rural Infrastructure." *International Regional Science Review* 24, no. 1: 103–133. https://doi.org/10.1177/016001701761012971.

Freire-Gibb, L. C., and K. Nielsen. 2014. "Entrepreneurship Within Urban and Rural Areas: Creative People and Social Networks." *Regional Studies* 48, no. 1: 139–153. https://doi.org/10.1080/00343404.2013.808322.

Galloway, L., J. Sanders, and D. Deakins. 2011. "Rural Small Firms' Use of the Internet: From Global to Local." *Journal of Rural Studies* 27, no. 3: 254–262. https://doi.org/10.1016/j.jrurstud.2011.05.005.

Gashi Nulleshi, S., and M. Tillmar. 2022. "Rural Proofing Entrepreneurship in Two Fields of Research." *International Journal of Entrepreneurial Behavior & Research* 28, no. 9: 332–356. https://doi.org/10.1108/ijebr-05-2021-0323.

GHSL. 2015. "Global Human Settlement Layer, Joint Research Center, European Commission." https://ghsl.jrc.ec.europa.eu.

Goetz, S. J., M. D. Partridge, S. C. Deller, and D. A. Fleming. 2010. "Evaluating US Rural Entrepreneurship Policy." *Journal of Regional Analysis and Policy* 40, no. 1: 20–33.

Gyimah, P., and R. N. Lussier. 2021. "Rural Entrepreneurship Success Factors: An Empirical Investigation in an Emerging Market." *Journal of Small Business Strategy* 31, no. 4: 5–19. https://doi.org/10.53703/001c.29470.

Habersetzer, A., M. Rataj, R. H. Eriksson, and H. Mayer. 2021. "Entrepreneurship in Rural Regions: The Role of Industry Experience and Home Advantage for Newly Founded Firms." *Regional Studies* 55, no. 5: 936–950. https://doi.org/10.1080/00343404.2020.1826038.

Hannan, M. T., and J. Freeman. 1977. "The Population Ecology of Organizations." *American Journal of Sociology* 82, no. 5: 929–964. https://doi.org/10.1086/226424.

He, T., R. Y. Chen, Z. Y. Wang, et al. 2023. "Spatial Differentiation and Factors Influencing the Benefits of Industrial Poverty Alleviation in Villages." *Growth and Change* 54, no. 1: 326–345. https://doi.org/10.1111/grow.12658.

Hjaltadóttir, R. E., T. Makkonen, and T. Mitze. 2020. "Inter-Regional Innovation Cooperation and Structural Heterogeneity: Does Being a Rural, or Border Region, or Both, Make a Difference?" *Journal of Rural Studies* 74: 257–270. https://doi.org/10.1016/j.jrurstud.2019.10.008.

Huiban, J. P. 2011. "The Spatial Demography of New Plants: Urban Creation and Rural Survival." *Small Business Economics* 37, no. 1: 73–86. https://doi.org/10.1007/s11187-009-9228-0.

Isard, W. 1949. "A General Theory of Location and Space-Economy." *Quarterly Journal of Economics* 41: 629–658.

Kalantaridis, C., and Z. Bika. 2006. "In-Migrant Entrepreneurship in Rural England: Beyond Local Embeddedness." *Entrepreneurship & Regional Development* 18, no. 2: 109–131. https://doi.org/10.1080/08985620500510174.

Kass, G. V. 1980. "An Exploratory Technique for Investigating Large Quantities of Categorical Data." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 29, no. 2: 119–127. https://doi.org/10.2307/2986296.

Kim, Y., and P. F. Orazem. 2017. "Broadband Internet and New Firm Location Decisions in Rural Areas." *American Journal of Agricultural Economics* 99, no. 1: 1–18. https://doi.org/10.1093/ajae/aaw082.

Krichevskiy, D., and T. Snyder. 2015. "US State Government Policies and Entrepreneurship." *Journal of Entrepreneurship and Public Policy* 4, no. 1: 102–110. https://doi.org/10.1108/jepp-09-2013-0041.

Laurin, F., S. Pronovost, and M. Carrier. 2020. "The End of the Urban-Rural Dichotomy? Towards a New Regional Typology for SME Performance." *Journal of Rural Studies* 80: 53–75. https://doi.org/10.1016/j.jrurstud.2020.07.009.

Lavesson, N. 2018. "How Does Distance to Urban Centres Influence Necessity and Opportunity-Based Firm Start-Ups?" *Papers in Regional Science* 97, no. 4: 1279–1303. https://doi.org/10.1111/pirs.12289.

Marshall, A. 1890. *Principles of Economics*. 1st ed. London: Macmillan.

McElwee, G., and R. Smith. 2014. "Researching Rural Entrepreneurship." In *Handbook of Research on Entrepreneurship*, edited by A. Fayolle, 432–470. Cheltenham: Edward Elgar.

Miles, M., and M. Morrison. 2020. "An Effectual Leadership Perspective for Developing Rural Entrepreneurial Ecosystems." *Small Business Economics* 54, no. 4: 933–949. https://doi.org/10.1007/s11187-018-0128-z.

Modrego, F., M. Atienza, and L. Hernández. 2023. "Agglomeration Factors and the Geography of Growing Early-Stage Businesses in Chile." *Growth and Change* 55, no. 1: 12692. Advance online publication. https://doi.org/10.1111/grow.12692.

Müller, S., and S. Korsgaard. 2018. "Resources and Bridging: The Role of Spatial Context in Rural Entrepreneurship." *Entrepreneurship & Regional Development* 30, no. 1–2: 224–255. https://doi.org/10.1080/08985626.2017.1402092.

Nelson, K. S., and T. D. Nguyen. 2023. "Community Assets and Relative Rurality Index: A Multi-Dimensional Measure of Rurality." *Journal of Rural Studies* 97: 322–333. https://doi.org/10.1016/j.jrurstud.2022.12.025.

North, D., and D. Smallbone. 2006. "Developing Entrepreneurship and Enterprise in Europe's Peripheral Rural Areas: Some Issues Facing Policy-Makers." *European Planning Studies* 14, no. 1: 41–60. https://doi.org/10.1080/09654310500339125.

Novejarque Civera, J., M. Pisá Bó, and J. F. López-Muñoz. 2021. "Do Contextual Factors Influence Entrepreneurship? Spain's Regional Evidences." *International Entrepreneurship and Management Journal* 17, no. 1: 105–129. https://doi.org/10.1007/s11365-019-00625-1.

OECD. 2023. *OECD Framework for the Evaluation of SME and Entrepreneurship Policies and Programmes*. Paris: Organisation for Economic Development and Cooperation. https://doi.org/10.1787/a4c818d1-en.

Pato, M. L., and A. A. Teixeira. 2016. "Twenty Years of Rural Entrepreneurship: A Bibliometric Survey." *Sociologia Ruralis* 56, no. 1: 3–28. https://doi.org/10.1111/soru.12058.

Pfeffer, J., and G. R. Salancik. 1978. *The External Control of Organizations*. New York, NY: Harper & Row.

Pratt, J. W. 1987. "Dividing the Indivisible: Using Simple Symmetry to Partition Variance Explained." In *Proceedings of the Second International Tampere Conference in Statistics, 1987*, 245–260. Tampere, Finland: Department of Mathematical Sciences, University of Tampere.

Renski, H. 2008. "New Firm Entry, Survival, and Growth in the United States: A Comparison of Urban, Suburban, and Rural Areas." *Journal of the American Planning Association* 75, no. 1: 60–77. https://doi.org/10.1080/01944360802558424.

Rosenthal, S. S., and W. C. Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economies." In *Handbook of Regional and Urban Economics*, Vol. 4, 2119–2171. Elsevier. https://doi.org/10.1016/s1574-0080(04)80006-3.

Rupasingha, A., and A. W. Marré. 2020. "Moving to the Hinterlands: Agglomeration, Search Costs and Urban to Rural Business Migration." *Journal of Economic Geography* 20, no. 1: 123–153.

Shane, S. 2009. "Why Encouraging More People to Become Entrepreneurs Is Bad Public Policy." *Small Business Economics* 33, no. 2: 141–149. https://doi.org/10.1007/s11187-009-9215-5.

Shrivastava, U., and A. Kumar Dwivedi. 2021. "Manifestations of Rural Entrepreneurship: The Journey So Far and Future Pathways."

Management Review Quarterly 71, no. 4: 753–781. https://doi.org/10.1007/s11301-020-00199-1.

Simpson, E. H. 1951. "The Interpretation of Interaction in Contingency Tables." *Journal of the Royal Statistical Society: Series B* 13, no. 2: 238–241. https://doi.org/10.1111/j.2517-6161.1951.tb00088.x.

Sorenson, O. 2018. "Social Networks and the Geography of Entrepreneurship." *Small Business Economics* 51, no. 3: 527–537. https://doi.org/10.1007/s11187-018-0076-7.

Stam, E., and A. Van de Ven. 2021. "Entrepreneurial Ecosystem Elements." *Small Business Economics* 56, no. 2: 809–832. https://doi.org/10.1007/s11187-019-00270-6.

Stearns, T. M., N. M. Carter, P. D. Reynolds, and M. L. Williams. 1995. "New Firm Survival: Industry, Strategy, and Location." *Journal of Business Venturing* 10, no. 1: 23–42. https://doi.org/10.1016/0883-9026(94)00016-n.

Stephens, H. M., and M. D. Partridge. 2011. "Do Entrepreneurs Enhance Economic Growth in Lagging Regions?" *Growth and Change* 42, no. 4: 431–465. https://doi.org/10.1111/j.1468-2257.2011.00563.x.

Sternberg, R. 2022. "Entrepreneurship and Geography—Some Thoughts About a Complex Relationship." *Annals of Regional Science* 69, no. 3: 559–584. https://doi.org/10.1007/s00168-021-01091-w.

Stinchcombe, A. L. 1965. "Social Structures and Organizations." In *Handbook of Organizations*, edited by J. G. March, 142–193. Chicago: Rand McNally.

Stuart, T., and O. Sorenson. 2003. "The Geography of Opportunity: Spatial Heterogeneity in Founding Rates and the Performance of Biotechnology Firms." *Research Policy* 32, no. 2: 229–253. https://doi.org/10.1016/s0048-7333(02)00098-7.

Tibshirani, R. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society - Series B: Statistical Methodology* 58, no. 1: 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Tonidandel, S., and J. M. LeBreton. 2011. "Relative Importance Analysis: A Useful Supplement to Regression Analysis." *Journal of Business and Psychology* 26, no. 1: 1–9. https://doi.org/10.1007/s10869-010-9204-3.

Unger, J. M., A. Rauch, M. Frese, and N. Rosenbusch. 2011. "Human Capital and Entrepreneurial Success: A Meta-Analytical Review." *Journal of Business Venturing* 26, no. 3: 341–358. https://doi.org/10.1016/j.jbusvent.2009.09.004.

Van Leuven, A. J., S. A. Low, and E. Hill. 2023. "What Side of Town? How Proximity to Critical Survival Factors Affects Rural Business Longevity." *Growth and Change* 54, no. 2: 352–385. https://doi.org/10.1111/grow.12652.

World Bank. 2023. "World Development Indicators. Rural Population Growth (Annual %), United Nations Population Division's World Urbanization Prospects [Data File]." https://data.worldbank.org/indicator/SP.RUR.TOTL.ZG.

Wortman, M. S., Jr. 1990. "Rural Entrepreneurship Research: An Integration Into the Entrepreneurship Field." *Agribusiness* 6, no. 4: 329–344. https://doi.org/10.1002/1520-6297(199007)6:4<329::aid-agr2720060405>3.0.co;2-n.

Wurth, B., E. Stam, and B. Spigel. 2022. "Toward an Entrepreneurial Ecosystem Research Program." *Entrepreneurship Theory and Practice* 46, no. 3: 729–778. https://doi.org/10.1177/1042258721998948.

Xu, Q., M. Zhong, and Y. Dong. 2024. "Digital Finance and Rural Revitalization: Empirical Test and Mechanism Discussion." *Technological Forecasting and Social Change* 201: 123248. https://doi.org/10.1016/j.techfore.2024.123248.

Yin, J., X. Huang, Y. Dong, M. Zhao, and W. Tan. 2021. "Dual-Level Impact of Regional Context and Individual Attributes on Entrepreneurship Among Return Migrants in China." *Growth and Change* 52, no. 2: 1099–1116. https://doi.org/10.1111/grow.12480.

Zúñiga-Antón, M., J. Guillén, M. Caudevilla, and C. Bentué-Martínez. 2022. "Mapa 174. Zonificación de los municipios españoles sujetos a desventajas demográficas graves y permanentes. StoryMap." https://storymaps.arcgis.com/stories/9dd9b6e20cad403c95e87d4cc493c8fb.