



# What makes you entrepreneurial? Using machine learning to investigate the determinants of entrepreneurship in China

Feng Liu<sup>a</sup>, Xiao Long<sup>a</sup>, Lin Dong<sup>a</sup>, Mingjie Fang<sup>b,\*</sup>

<sup>a</sup> Business School, Shandong University, Weihai, China

<sup>b</sup> Department of Logistics, Service & Operations Management, Korea University Business School, Seoul, South Korea

## ARTICLE INFO

### JEL codes:

L26  
M13  
M21  
O12

### Keywords:

Entrepreneurs  
Entrepreneurship  
Machine learning  
China

## ABSTRACT

Entrepreneurs and entrepreneurship are essential for new wealth creation and economic growth, particularly in developing countries. The objective of this study is to develop a model to identify the determinants of entrepreneurship, and reveal how it can be predicted based on individual characteristics, family environment, and social environment. Employing 16 sets of machine learning algorithms on data collected from the Chinese General Social Survey in 2017, we find the best-performing algorithms (i.e., lasso, ridge, and elastic net regression) and examine the effects of the feature variables on entrepreneurship. Overall, this study provides significant theoretical underpinnings for entrepreneurship research, and offers insights for individuals and policymakers by revealing various drivers of entrepreneurship.

## 1. Introduction

Fostering entrepreneurship continues to be a priority in governmental policy due to its critical importance to job creation, economic development and social welfare (Méndez-Picazo, Galindo-Martín, and Ribeiro-Soriano, 2012; Yu, Dai, Liu, and Yang, 2023). Governments across the world have been providing supportive programs, including both, financial and non-financial support, to accelerate the growth of entrepreneurship (Farinha, Lopes, Bagchi-Sen, Sebastião, and Oliveira, 2020; Kollmann, Hensellek, Jung, and de Cruppe, 2022). For instance, in the last decade, the Chinese government has launched a set of policies to help college students, migrant workers, and unemployed people to set up businesses. These policies include financial support in terms of tax and loan interest rate reduction, growth capital, as well as non-financial support, such as education programs, legal consulting, and approval process simplification.

Indeed, the “invisible hands” of institutions play a critical role in boosting entrepreneurship, and there is a sizable body of research that investigates how both formal and informal institutional factors can influence entrepreneurship and its economic outcomes (e.g., Bruno, Bychkova, and Estrin, 2013; Bruton, Ahlstrom, and Li, 2010; Castaño, Méndez, and Galindo, 2016). While the locus of this stream of research has remained within the boundaries of institutions, other contextual factors may have an impact on entrepreneurship. In fact, scholars have demonstrated that personality traits (Antoncic, Bratkovic Kregar, Singh, and DeNoble, 2015), early life experiences (Churchill, Munyanyi, Smyth, & Trinh, 2021), and education level (Cheng and Smyth, 2021) are closely related to the propensity of becoming entrepreneurs. Although these advances have significantly enhanced our understanding of what determines

\* Corresponding author.

E-mail addresses: [liufeng@sdu.edu.cn](mailto:liufeng@sdu.edu.cn) (F. Liu), [longxiao0315@mail.sdu.edu.cn](mailto:longxiao0315@mail.sdu.edu.cn) (X. Long), [donglin1225@mail.sdu.edu.cn](mailto:donglin1225@mail.sdu.edu.cn) (L. Dong), [mj\\_fang@korea.ac.kr](mailto:mj_fang@korea.ac.kr) (M. Fang).

<https://doi.org/10.1016/j.chieco.2023.102029>

Received 25 August 2022; Received in revised form 20 May 2023; Accepted 13 July 2023

Available online 20 July 2023

1043-951X/© 2023 Elsevier Inc. All rights reserved.

entrepreneurship, some important factors of personal attributes and other contextual drivers, such as family and the social environment, have been insufficiently studied. Undeniably, family and social environment can play a role in shaping individuals' behavior and cognitive skills and, by extension, the foresight and perceptivity of entrepreneurship. However, to date, empirical insights into the complexity of how personal characteristics, family, and social environment collectively affect entrepreneurship, are limited, which motivates this research.

To fill these gaps, we constructed a model that considered various personal attributes, as well as family and social environment features, to predict their impact on individual entrepreneurship. In this study, we used state-of-the-art machine learning (ML) algorithms to test our model. To find the best ML model for entrepreneurship prediction, we trained and tested 16 different sets of algorithms multiple times, using data collected from the Chinese General Social Survey in 2017. Thereafter, we visualized the permutation importance and heterogeneous effects of the features, based on the best-performing ML model.

Our study makes valuable contributions to the literature in two ways. First, we developed an entrepreneurship determinant model and extended our understanding of the drivers of entrepreneurship. Previous studies in this field have been largely limited to exploring how institutional forces affect entrepreneurship (e.g., Bruno et al., 2013; Bruton et al., 2010), whereas our research enriches the literature by investigating how personal characteristics, family, and social environment matter in individual entrepreneurship. In particular, our study not only confirmed the permutation importance of the features required for entrepreneurship, but also revealed their positive and negative impacts on fostering entrepreneurship. Another contribution of the present study is that we identified the best-performing ML models for entrepreneurship prediction. As our model contains various types of variables, using traditional economic statistics for inferential analysis may be problematic, however, ML allows to "learn" the complexity from the data, that is, "presumed patterns" do not need to be presumed, thereby helping eliminate idiosyncratic problems. Because there is currently no consensus on the best algorithm for entrepreneurship predictive modeling, our findings provide important implications for both, practitioners, and scholars for entrepreneurship prediction.

The remainder of this paper is organized as follows. In Section 2, we discuss the relevant literature on personal characteristics, family, and social environment. Next, we describe the sample and methodology. A performance comparison of the ML algorithms and the main results are provided in Section 4. Section 5 ends with a discussion and conclusion, including the implications, limitations, and future research avenues.

## 2. Literature review

### 2.1. Individual characteristics and entrepreneurship

In investigating the role of individual characteristics, the present research is theoretically grounded in demographic and psychological research. Demographic attributes have been an important research focus in the literature, as they are closely related to cognition and behavior (Chowdhury, 2005). Prior literature has demonstrated that entrepreneurial activities involve a variety of planning and strategic choices, and that people with different values and cognitive bases may arrive at different outcomes (Brandstätter, 2011; Frese and Gielnik, 2014). For example, scholars have shown that gender plays a role in people's entrepreneurial beliefs and intentions because it impacts an individual's strategic orientation, risk-taking preference, and performance outcomes (Davis, Babakus, Englis, and Pett, 2010; Santos, Roomi, and Liñán, 2016), some scholars claimed that such differences do not depend on the level of economic development (Minniti and Nardone, 2007). Other studies, such as Lévesque and Minniti (2011) found that the age of a country's population has an inverted U-shaped impact on entrepreneurial decisions, such that "young" and "old" nations tend to show low rates of entrepreneurship. Scholars have also reported the impact of education, although the results are mixed. Conventional views hold that people who are well educated are more likely to become self-employed entrepreneurs because they are believed to be more capable of managing an efficient organization (Evans and Leighton, 1989). Conversely, there are also studies that put forward a different point of view, arguing that individuals with a higher education level are less likely to become entrepreneurs, because of relatively larger opportunity costs (Huang, Tani, and Zhu, 2021).

Additionally, a growing body of literature has begun to examine the complexity between traits and entrepreneurship. For instance, Antoncic et al. (2015) investigated the influence of the "Big Five" personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism) on entrepreneurship manifestation. Wang, Chang, Yao, and Liang (2016) expanded this idea by illustrating the moderating role of self-efficacy in the relationship between personality and entrepreneurship. Uysal, Karadağ, Tuncer, and Şahin (2022) found that ambitious people are more likely to become entrepreneurs because of their aspirations for accomplishment. Interestingly, some scholars have traced one of the underlying causes of personality formation by examining whether, and how, early life experiences influence future entrepreneurship (e.g., Churchill et al., 2021; Cheng, Guo, Hayward, Smyth, and Wang, 2021). The dimensions of personal characteristics in this study have some similarities to these advances, but our analysis considers more potential attributes, such as religion, optimism, and involvement in investment activities, which have received limited attention in existing literature.

### 2.2. Family environment and entrepreneurship

The literature also includes certain aspects of family environment as factors in explaining entrepreneurship (Nicolaou, Shane, Cherkas, Hunkin, and Spector, 2008; Rosado-Cubero, Freire-Rubio, and Hernández, 2022), while the findings related to this relationship are inconclusive. In early entrepreneurship research, the impact of the family and upbringing on the decision to be self-employed appears to play out across generations through wealth, human capital, and a variety of demographic factors, such as

marriage and the number of children (e.g., Laferre' re and McEntee, 1995). Specifically, they proposed that an individual is more likely to start a business if his/her father or father-in-law is an entrepreneur. In addition, it seems that the role of the spouse is also important, since a person is more likely to be an entrepreneur if the spouse is in the labor force. The decision to move from wage labor to self-employment is also influenced by marriage and the number of children. A series of later studies also suggested that entrepreneurship is closely related to family environment, such as parental attitude, identity, and education level (Heineck et al., 2017; Lindquist, Sol, and Van Praag, 2015; Luis-Rico et al., 2020; Rosado-Cubero et al., 2022). To illustrate, individuals whose parents are entrepreneurs – regardless of whether they are biological or adoptive – receive distinctive help, in terms of not only financial capital, but also entrepreneurial skills.

Conversely, several studies have cautioned that some parental attitudes and communication may be detrimental to the offspring's entrepreneurship. Being inconsistent, holding excessive expectations from, and being protective of children, will have a negative impact on entrepreneurship, because children growing in such family environments may not be independent enough, or may have poor ability to withstand setbacks (Staniewski and Awruk, 2021). Scholars have also reported a marginal or non-significant relationship between family factors (e.g., parental attitudes and cohesion) and entrepreneurship (Nicolaou et al., 2008; Staniewski and Awruk, 2021). In this study, we contribute to the literature by further exploring the mechanisms underlying the relationship between family environment and entrepreneurship by considering large-scale variables, such as family financial status, marital status, and parental conditions and occupations.

### 2.3. Social environment and entrepreneurship

In addition to the dimensions of personal characteristics and family environment, we further argue that the social environment also contributes significantly to fostering entrepreneurship. The social environment refers to the broad context for survival and development, including political, economic, and cultural environments, and the network of relations with others (Repetti and Cosmas, 1991). Theoretically and practically, starting a business is connected with the social environment (Littlewood and Holt, 2018; Santos et al., 2016). First, from political and economic perspectives, scholars have confirmed that societies with strong legal systems, less corruption, free economies, and better economic performance, demonstrate higher levels of entrepreneurial activity, as these all have an important influence on people's expectations in assessing returns from entrepreneurship (Castaño, Méndez, and Galindo, 2015). Second, the overall value and culture of society are also of critical importance in fostering entrepreneurship. Previous studies have argued that an individual's interest in entrepreneurship increases if the social environment in which the person lives values entrepreneurial behavior (De Carolis and Saporito, 2006; Morris and Schindehutte, 2005). This view has been confirmed in the planned behavior theory research, which proposes that the perception of entrepreneurship being accepted and approved by significant referents, such as peers in social networks (i.e., social norms), encourages individuals to set a business (Obschonka, Silbereisen, Cantner, and Goethner, 2015). This study contributes to the literature by adopting social equity and social class to reflect the social environment, and exploring their impact on entrepreneurship.

In conclusion, despite the significant insights offered by the previous studies, they are limited in their ability to provide in-depth empirical evidence on the roles of personal characteristics, family and social environments in shaping entrepreneurship. In this sense, our study develops a model that comprehensively considers these three aspects to enrich the literature. Specifically, the individual characteristics include an individual's gender, age, education level, political party membership, religion, optimism, happiness, information sources, and investment activity. The family environment aspect consists of one's family financial situation, marital status (single, married, and divorced), father's and mother's living conditions (if they are alive), and occupations (employee, agricultural worker, family company, freelancer, entrepreneur, retired, househusband/housewife). The last dimension of the model concerns the social environment, which includes the degree of social equity and an individual's social class-related factors.

## 3. Methodology

### 3.1. Sample and data

The data used in this study were collected from a nationwide survey, the Chinese General Social Survey (CGSS), conducted by the National Survey Research Center at Renmin University of China (NSRC) in 2017, which is a commonly used database in academic research (Su, Lien, and Yao, 2022). We believe that the sampling frame is appropriate because, as the exclusive partner of the Inter-university Consortium for Political and Social Research (ICPSR), the NSRC was the initiator of the East Asian Social Survey project, and is a member of the International Social Survey Program (ISSP). Moreover, CGSS is a nationwide survey that includes comprehensive cross-sectional data from various perspectives, such as society, community, family, and individuals, which we believe ensures randomized sampling to avoid statistical bias. After data processing, the dataset used for the analysis contained responses from 3708 Chinese residents.

### 3.2. Variables

**Target variable.** Entrepreneurship was determined from the following item: "Which of the following better describes your current work situation?" We assigned a value of 1 to the response that selected "as an entrepreneur" and 0 to "otherwise."

**Feature variable.** Our feature variables were divided into three main domains: personal characteristics, family environment, and social environment. First, we considered nine items that reflect a person's individual characteristics: age, gender, education level,

**Table 1**  
Description of the variables and the descriptive statistics.

Types	Variables	Description	Mean	Std. Dev.
Target variable	Entrepreneurship	1 if the individual is an entrepreneur, and 0, otherwise	0.205	0.404
	Gender (PC1)	A binary variable that is set to 1 if the individual is male, otherwise, 0.	0.567	0.496
	Age (PC2)	Age at the time when the data is collected	40.271	11.850
	Education level (PC3)	2 if the respondent has a post graduate degree, 1 if the respondent's highest degree is bachelor's degree, and 0, otherwise.	0.385	0.541
	Political party membership (PC4)	Dummy variable, which equals 1 if the individual is a party member of the Chinese Communist Party, and 1, otherwise.	0.126	0.332
Personal characteristics	Religion (PC5)	Dummy variable, 0 if the individual does not have religious beliefs, and 1 otherwise.	0.092	0.289
	Optimism (PC6)	The degree to which a person is optimistic ranges from 1 = "very pessimistic" to 5 = "very optimistic."	3.886	0.776
	Happiness (PC7)	The degree to which an individual feels happy ranges from 0 = "very unhappy" to 10 = "very happy."	2.571	5.596
	Information sources (PC8)	Dummy variable which equals 0 if the individual acquires information primarily from traditional sources such as newspapers, radio and magazines, and 1 otherwise.	0.981	0.138
	Investment activity (PC9)	If the person is currently engaged in investment activity, it is coded as 1, otherwise it is 0.	0.833	0.373
	Financial situation (FE1)	The degree of the family financial status compared to local average ranges from 1 = "far lower" to 5 = "far higher."	2.695	0.697
	Single (FE2)	If the individual's marital status is "single," it is coded as 1, otherwise, 0.	0.185	0.388
	Married (FE3)	If the individual's marital status is "married," it is coded as 1, otherwise, 0.	0.785	0.411
	Divorced (FE4)	If the individual's marital status is "divorced," it is coded as 1, otherwise, 0.	0.030	0.171
	Father was an employee (FE5)	If the individual's father was employed by others when the individual was fourteen years old, it is coded as 1, otherwise, 0.	0.378	0.485
	Father was an agricultural worker (FE6)	If the father of the individual was an agricultural worker when the individual was fourteen years old, it is coded as 1, otherwise, 0.	0.475	0.499
	Father worked in a family company (FE7)	If the father of the individual worked in a family company when he or she was fourteen years old, it is coded as 1, otherwise, 0.	0.004	0.061
	Father was a freelancer (FE8)	If the father of the individual was a freelancer when the individual was fourteen years old, it is coded as 1, otherwise, 0.	0.013	0.114
	Father was an entrepreneur (FE9)	If the father of the individual was an entrepreneur when the individual was fourteen years old, it is coded as 1, otherwise, 0.	0.079	0.270
	Father was retired (FE10)	If the father of the individual had retired when the individual was fourteen years old, it is coded as 1, otherwise, 0.	0.009	0.095
	Father was a househusband (FE11)	If the father of the individual only did the housework at home when the individual was fourteen years old, it is coded as 1, otherwise, 0.	0.001	0.028
	Father passed away (FE12)	If the father of the individual had passed away when the individual was fourteen years old, it is coded as 1, otherwise, 0.	0.029	0.167
	Mother was an employee (FE13)	If the mother of the individual was employed by others when the individual was fourteen years old, it is coded as 1, otherwise, 0.	0.233	0.423
	Mother was an agricultural worker (FE14)	If the mother of the individual was an agricultural worker when the individual was fourteen years old, it is coded as 1, otherwise, 0.	0.523	0.500
	Mother worked in a family company (FE15)	If the mother of the individual worked in a family company when the individual was fourteen years old, it is coded as 1, otherwise, 0.	0.006	0.077
Mother was a freelancer (FE16)	If the mother of the individual was a freelancer when the individual was fourteen years old, it is coded as 1, otherwise, 0.	0.008	0.087	
Mother was an entrepreneur (FE17)	If the mother of the individual was an entrepreneur when the individual was fourteen years old, it is coded as 1, otherwise, 0.	0.050	0.219	
Mother was retired (FE18)	If the mother of the individual had retired when the individual was fourteen years old, it is coded as 1, otherwise, 0.	0.008	0.090	
Mother was a housewife (FE19)	If the mother of the individual was a housewife when the individual was fourteen years old, it is coded as 1, otherwise, 0.	0.133	0.340	
Family environment	Mother passed away (FE20)	If the mother of the individual had passed away when the individual was fourteen years old, it is coded as 1, otherwise, 0.	0.017	0.130
	Social equity (SE1)	The degree of social equity in a person's location ranges from 1 = "very unequal" to 5 = "very equal".	3.060	1.040
	Social class (now) (SE2)	The social class in which the individual currently lives, from 1 = "very low" to 10 = "very high."	4.389	1.610
	Social class (ten years ago) (SE3)	The social class the individual was in ten years ago, from 1 = "very low" to 10 = "very high."	3.579	1.729
Social environment	Social class (after ten years) (SE4)	The social class the individual is in after 10 years, from 1 = "very low" to 10 = "very high."	5.463	1.942
	Social class (fourteen) (SE5)	Social class of the individual at age 14, from 1 = "very low" to 10 = "very high."	3.507	1.786

political party membership, religion, optimism, happiness, information sources, and investment activity. Numerous indicators were used to reflect the family environment. We considered the financial situation, marital status (i.e., single, married, and divorced), and parents' condition (whether they are alive) and occupation. Specifically, we created several dummy variables to distinguish between fathers and mothers' occupations in order to specify their impact on an individual's entrepreneurial activities. Finally, the social environment is reflected by the degree of social equity and social class in which an individual lives. The detailed definitions and descriptive statistics of the variables are presented in Table 1.

### 3.3. Machine learning

This study utilizes ML algorithms to build a model for entrepreneurship prediction. In recent years, ML has attracted much attention in the literature for its superior predictive performance (Shrestha, He, Puranam, and von Krogh, 2021), and has been applied to various disciplines (e.g., Leippold, Wang, and Zhou, 2022; Melançon, Grangier, Prescott-Gagnon, Sabourin, and Rousseau, 2021; Williamson, Drencheva, and Battisti, 2020). Compared to traditional statistical models, the merits of ML are that it does not require presumed inference, but “learns” patterns from the data, and that the amplified set of procedures in ML allows better adaptability to complex functions and avoids overfitting problems (Shrestha et al., 2021; Wang, Hao, Ma, and Jiang, 2011; Xu, Pan, and Xia, 2020). This virtue can be particularly useful for this study, as our sample contains a variety of features and different types of variables, which are typically challenging to conduct inferential analysis. Moreover, because the primary goal of ML is to accurately predict and replicate results, the findings generated are unlikely to be untenable because of sample idiosyncrasy.

Sixteen sets of different types of machine-learning algorithms were employed in this study. For linear models, we employed Lasso regression (Lasso), ridge regression (Ridge), elastic net regression (ElasticNet), logistic regression (LR), and linear discriminant analysis (LDA). The tree models, decision tree (DT), random forest (RF), adaptive boosting (AdaBoost), and light gradient boosting (LightGBM), were also trained and tested in our study. In addition, support vector machine (SVM) with different kernels, linear kernel function (SVM-L), radial basis kernel function (SVM-R) and polynomial kernel function (SVM-P), were employed in this research. Three different Naïve Bayes (NB) algorithms (NB-Gaussian and NB-Bernoulli), k-nearest neighbor (KNN), and multilayer perceptron (MLP) were utilized for entrepreneurship prediction. Online Appendix A provides a detailed description of these algorithms.

## 4. Results

### 4.1. Performance comparison of ML models

To explore the determinants of entrepreneurship, we divided our sample into two sets (i.e., training and test sets). The training set was used for hyperparameter tuning and model fitting, while the test set was used to evaluate prediction performance. The training and test results for the algorithms are presented in Table 2. As there is currently no consensus on the best data-splitting method, 11 splitting strategies were applied in this study. For example, 80:20 indicates that we used 80% of the data to train the model and 20% to test the performance of the model. To determine the best-performing algorithm for entrepreneurship prediction, we used the mean square error (MSE) for performance evaluation, which is calculated as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

where  $y_i$  is the actual output data,  $\hat{y}_i$  is the fitted output data and  $n$  is the total number of samples. The smaller the MSE value, the better the performance of the model (Makridakis, Wheelwright, and Hyndman, 2008). As can be seen from Table 2, the MSE values of Lasso, Ridge and ElasticNet are the lowest (MSE = 0.145). As a result, Lasso, Ridge, and ElasticNet demonstrate the best performance in

**Table 2**  
MSE prediction performance.

Splitting	80:20	78:22	76:24	74:26	72:28	70:30	68:32	66:34	64:36	62:38	60:40	Average MSE
Lasso	0.141	0.142	0.143	0.141	0.143	0.145	0.146	0.147	0.148	0.147	0.148	0.145
Ridge	0.141	0.142	0.143	0.142	0.143	0.145	0.147	0.147	0.148	0.148	0.149	0.145
ElasticNet	0.141	0.141	0.143	0.141	0.143	0.141	0.143	0.145	0.146	0.147	0.148	0.145
KNN	0.241	0.235	0.228	0.230	0.234	0.236	0.242	0.241	0.240	0.243	0.245	0.238
NB-Bernoulli	0.185	0.188	0.190	0.187	0.192	0.196	0.198	0.199	0.200	0.199	0.200	0.194
NB-Gaussian	0.398	0.382	0.375	0.366	0.364	0.362	0.364	0.358	0.363	0.364	0.367	0.369
DT	0.185	0.188	0.190	0.187	0.192	0.196	0.198	0.199	0.200	0.199	0.201	0.194
RF	0.185	0.188	0.190	0.187	0.192	0.196	0.198	0.199	0.200	0.199	0.200	0.194
MLP	0.182	0.186	0.194	0.189	0.192	0.193	0.198	0.197	0.199	0.200	0.202	0.194
SVM-L	0.158	0.160	0.162	0.159	0.163	0.167	0.168	0.169	0.170	0.169	0.170	0.165
SVM-R	0.158	0.160	0.162	0.159	0.163	0.167	0.168	0.169	0.170	0.169	0.170	0.165
SVM-P	0.158	0.160	0.162	0.159	0.163	0.167	0.168	0.169	0.170	0.169	0.170	0.165
LightGBM	0.154	0.160	0.155	0.153	0.149	0.151	0.148	0.153	0.152	0.152	0.155	0.153
LR	0.185	0.188	0.190	0.187	0.192	0.196	0.198	0.199	0.200	0.199	0.200	0.194
AdaBoost	0.178	0.180	0.183	0.180	0.185	0.190	0.192	0.192	0.193	0.192	0.193	0.187
LDA	0.185	0.188	0.190	0.187	0.192	0.196	0.198	0.199	0.199	0.198	0.199	0.194

entrepreneurship prediction.

#### 4.2. Empirical results

Based on Reddit’s “Explain Like I’m Five” (ELI5) (Pflugfelder, 2017), we revealed the permutation importance of these feature variables in predicting the target variable using the 80:20 splitting (i.e., the sample splitting with lowest MSE), and the results are shown in Fig. 1. The color gradient indicates the importance of the feature variable, from highest to lowest (1 to 0). Although there are some slight differences in the output of the feature importance for the three algorithms, the results suggest that education level (PC3), financial situation (FE1), political party membership (PC4), marital status (single, FE2), and mother’s occupation (mother was an employee, FE13) are the top five factors that influence an individual’s entrepreneurship, while father’s occupation (father was a freelancer, FE8), social class (ten years ago, SE3), social class (after ten years, SE4), optimism (PC6), and mother’s condition (Mother passed away, FE20) are relatively less important. Note that some variables are not shown in Fig. 1, because of insignificant contribution to the prediction output.

To this end, we have identified the best algorithms for entrepreneurship prediction, and the importance of these predictors, while we are interested in taking a closer look at the respective positive/negative impacts of features. The results of coefficients in the Ridge Model are plotted in Fig. 2. Red indicates that an independent variable is positively associated with entrepreneurship, whereas blue indicates a negative relationship. Note that because the results of the three models are similar, we only select the results of the Ridge model to save space (Online Appendix B presents the results of the Lasso and ElasticNet Models).

Overall, in descending order of positive determinants, father’s occupation (father worked in a family company, FE7), mother’s occupation (mother worked in a family company, FE15), mother’s condition (mother passed away, FE20), father’s occupation (father was a househusband, FE11), financial situation (FE1), marital status (married, FE3), mother’s occupation (mother was an agricultural worker, FE14), father’s occupation (father was an entrepreneur, FE9), mother’s occupation (mother was an entrepreneur, FE17), father’s occupation (father was an agricultural worker, FE6), investment activity (PC9), mother’s occupation (mother was a housewife, FE19), marital status (divorced, FE4), social class (now, SE2), social class (ten years ago, SE3), social class (after ten years, SE4), social equity (SE1), optimism (PC6), and age (PC2) were positively related to entrepreneurship. Conversely, political party membership (PC4) had the strongest negative impact on entrepreneurship, followed by educational level (PC3), father’s condition (father passed away, FE12), father’s occupation (father was retired, FE10), marital status (single, FE2), father’s occupation (father was a freelancer, FE8), mother’s occupation (mother was an employee, FE13), information sources (PC8), father’s occupation (father was an employee, FE5), social class (fourteen, SE5), mother’s occupation (mother was a freelancer, FE16), mother’s occupation (mother was retired, FE18), religion (PC5), gender (PC1), and happiness (PC7).

### 5. Discussion and conclusion

Academics and policymakers alike have promoted the benefits of nurturing entrepreneurship as a way to improve economic development in an era of economic reform. The present research aims to develop an entrepreneurship determinant model, and reveal how entrepreneurship can be predicted from an individual’s personal characteristics, family environment, and social environment. Employing 16 sets of ML algorithms on data collected from a large-scale survey data, we identified the best-performing algorithms and examined the positive/negative effects of the features. Our research makes several contributions to the theory and practice of entrepreneurial research, which we discuss below.

#### 5.1. Theoretical contributions

This study contributes to the literature in several ways. The first contribution is related to the use of ML models. Previous literature

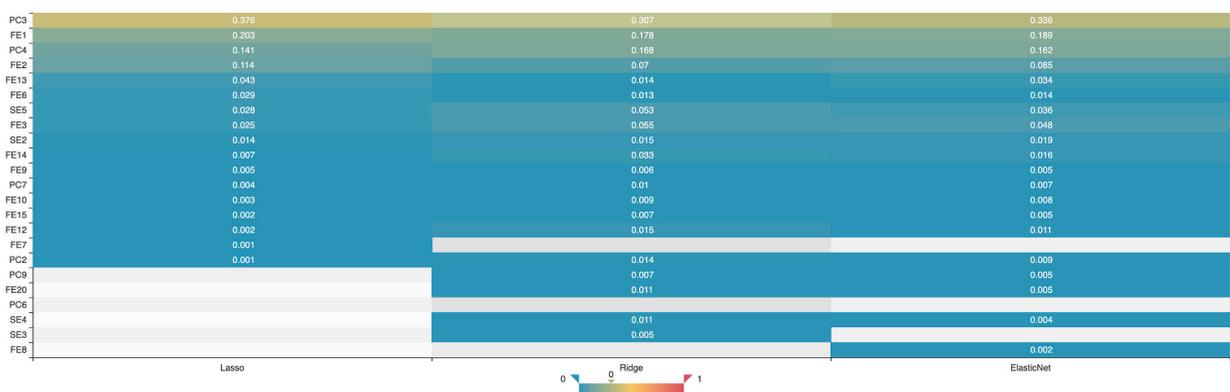


Fig. 1. The importance degree of feature variables for Lasso, Ridge, and ElasticNet.

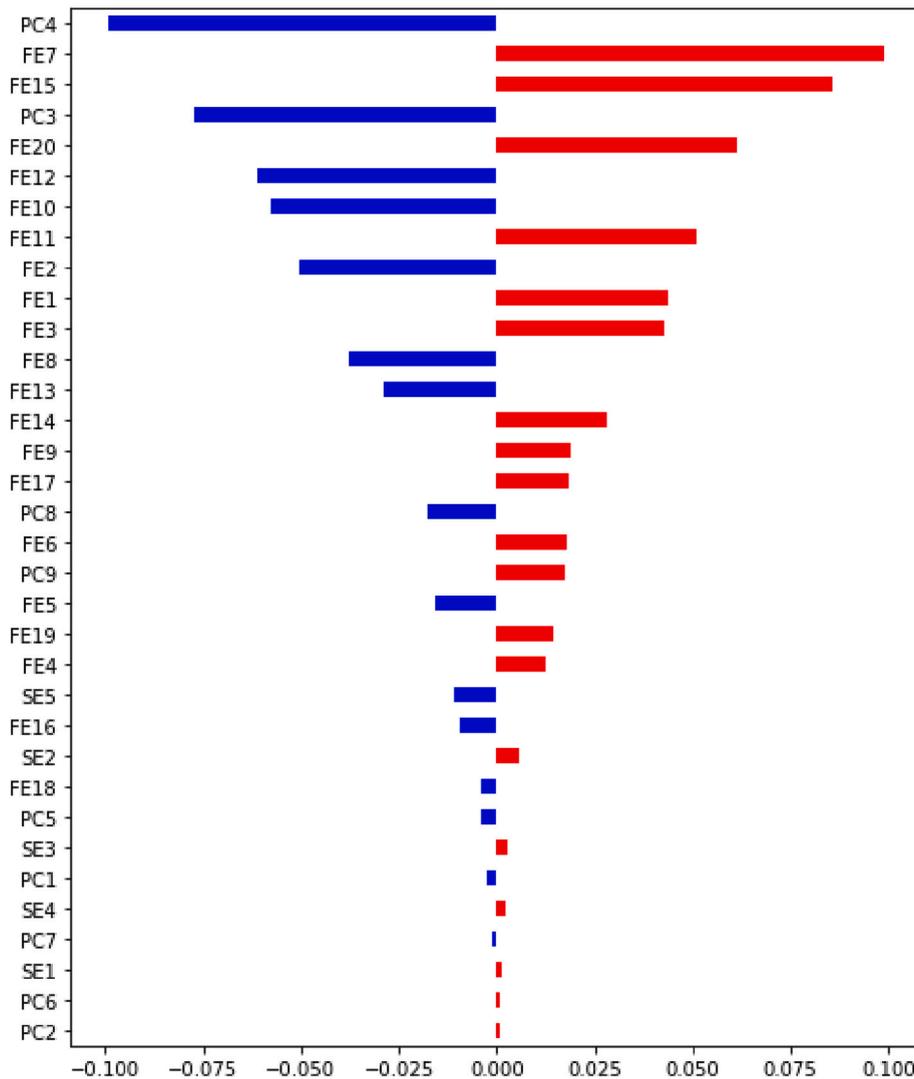


Fig. 2. The results of entrepreneurship determinant effects for Ridge.

has argued that addressing the overfitting issue when using the ordinary least squares (OLS) model can be challenging, as the process of adding or removing variables in inferential analysis may lead to overly complex and sample-dependent models (Shrestha et al., 2021). This may explain the stringent requirements for robustness testing in academia. ML can help fix this issue by allowing us to find appropriate functional forms while alleviating the risk of overfitting. As there is currently no consensus on the best-performing model for predicting entrepreneurship, we consider 16 sets of algorithms that have been proven by previous studies to have good predictive performance, to construct our model. By training and testing these algorithms based on 11 splitting strategies, we identified the best models for entrepreneurship prediction (i.e., Lasso, Ridge, and ElasticNet), visualized the results, and interpreted the findings.

Second, this study contributes to entrepreneurship literature by identifying the determinants of entrepreneurship from multiple aspects that have not been well considered in previous studies. In this study, we show that an individual's personal characteristics, family, and social environment are important drivers of entrepreneurship. Regarding personal characteristics, we find that individuals who engage in investment activities, and are older and optimistic, are more likely to be entrepreneurs. Other factors such as gender, educational level, political party membership, religion, happiness, and information sources were found to affect entrepreneurship negatively. The positive role of investment activities implies that people who invest, may be capable of understanding market trends, and choosing promising industries to start a business, while more optimistic people may not easily give up, even if they encounter setbacks in the process of starting a business or investment, making them more likely to become entrepreneurs (Rao, Mei, and Zhu, 2016). Note that there are divergent results for happiness and optimism, which may be because people who always feel happy have a higher level of satisfaction with the current state of their lives, and therefore, they may be more risk-averse.

Interestingly, previous studies have argued that party members are more capable of seizing opportunities in the market and utilizing their political resources to obtain benefits in terms of regulations, taxes and subsidiaries (Beck, Lu, and Yang, 2015; Liu, Tang,

and Tian, 2013; Yang, Huang, Deng, and Bordignon, 2020). We challenge this view by demonstrating the negative relationship between Chinese Communist Party (CCP) membership and an individual's entrepreneurship. We infer that this may be because, in recent years, the CCP has increasingly regulated the business activity of party members, which greatly limited the existence of "free-riders" (Yan and Xu, 2022). We also find that more educated individuals are less likely to choose to become entrepreneurs, implying that well-educated individuals are expected to have higher earnings and more attractive working conditions (Van der Sluis, Van Praag, and Vijverberg, 2008), thereby making them less likely to start a business; this is consistent with the relevant literature (Cheng and Smyth, 2021; Huang et al., 2021). In addition, while some studies have shown that gender can be an important contextual determinant of entrepreneurship (Cheng and Smyth, 2021; Santos, Neumeier, and Morris, 2019), we extend their advances by directly showing that gender matters in terms of entrepreneurship, and that males are more entrepreneurial than females. We further verify that individuals' information access preferences influence their entrepreneurship; for example, people whose information sources are traditional, such as newspapers, radio, and magazines, are less likely to become entrepreneurs.

Our third contribution highlights the crucial role of family environment in influencing entrepreneurs. We found that individuals whose families have better financial conditions are more likely to become entrepreneurs. This is consistent with the view that the more affluent a family is, the greater its opportunities of embarking on entrepreneurial ventures, and succeeding in them (Nair and Pandey, 2006). Moreover, we took a closer look at how individuals' marital status, and parents' occupations and conditions, may influence entrepreneurship. The results suggest that married people are more entrepreneurial. Based on previous studies, we speculate that, to some extent, this may be due to the fact that marriage specializes tasks between couples, thereby increasing people's productivity (Wilmoth and Koso, 2002). Regarding the role of parents' occupation in nurturing entrepreneurship, a common tenor of the prior literature is that if one's parents are entrepreneurs, it would subconsciously influence the child's identity choices and career plans (Falck, Heblich, and Luedemann, 2012). Our results suggest that such an impact may be limited to the individual's father, but not to both parents. Notwithstanding, we partially agree with their view by showing the negative impact of parental occupation as an employee, freelancer, or retired. We also documented that parental occupation, as agricultural and family business employees, and househusbands/housewives, has a catalytic effect on children's entrepreneurship tendencies in the future. We speculate that this may be because people whose parents are employees, or have retired, find fewer opportunities for higher income and more attractive working conditions, and entrepreneurship is the best option to make a difference in their lives.

Our fourth contribution extends the understanding of the role of the social environment in nurturing entrepreneurship. Previous studies in this field have mainly focused on the perspective of social ties, social values, norms, and culture (Krueger, Liñán, and Nabi, 2013; Santos et al., 2016; Ülhoi, 2005). This study contributes to this stream of literature by investigating how social equity and social class may determine entrepreneurship. Specifically, social equity and social class positively impact entrepreneurship. To this end, our study holistically examined the critical role of personal characteristics, family, and social environment in nurturing entrepreneurship, answering Cheng, Li, Sun, and Wang's (2021) call for a new understanding of the causes of entrepreneurship in China.

## 5.2. Practical implications

The current research not only aids in expanding the theory and literature, but also has important implications for practice. First, our findings suggest that participation in investment activities positively influences entrepreneurship. Thus, participation in investment activities is crucial for becoming an entrepreneur. By accumulating investment experience, people can be more sensitive to market trends; thus, when they find an idea worthy of implementation, they start a business to turn the idea into a new product or service. Therefore, when people are conceiving entrepreneurial strategies but have not devised the execution scheme, we may suggest that they engage in some investment activities to learn the laws of industrial economics, firm growth, and profits. In terms of implications for policy makers, this research advances the understanding of how family and social environment can be critical determinants of entrepreneurship, and equips them with theories and empirical supporting evidence that explain why policies should be managed with great care. For instance, the positive impact of social equity requires government efforts to ensure the fair and equitable management of all institutions and resources. We argue that it is often not the risk that dissuades entrepreneurs, but inequity. Consequently, actions should be taken to regulate violations, such as corruption and speculation.

## 5.3. Limitations and directions

This study has several limitations, which suggest avenues for future research. The first limitation is that the data used in this study were obtained from residents in China, which may make the findings inapplicable to other economies and/or regions with different cultural backgrounds. Therefore, it is recommended that studies include data from other countries to strengthen the generalizability of our findings. Moreover, since entrepreneurial nurturing is not a one-off project, panel data analysis can be used to capture the dynamic nature of the relationship between predictors and entrepreneurship, to help entrepreneurs adjust their strategies from time to time. Second, although our study considers a relatively large set of variables in determining entrepreneurship, it is fair to infer that there are other factors driving the likelihood of an individual becoming an entrepreneur. Future research should consider more contextual factors and investigate how they impact entrepreneurship.

## Funding statement

This work was supported by the Humanities and Social Sciences Foundation of the Ministry of Education of China [Grant No. 21YJC630076].

## Declaration of Competing Interest

None.

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Acknowledgements

We gratefully acknowledge insightful suggestions from the editors and the anonymous reviewers, which substantively improved this article. We would also like to thank the members of Star-lights Machine Learning Research Team for their comments on earlier versions of the manuscript.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chieco.2023.102029>.

## References

- Antonic, B., Bratkovic Kregar, T., Singh, G., & DeNoble, A. F. (2015). The big five personality–entrepreneurship relationship: Evidence from Slovenia. *Journal of Small Business Management*, 53(3), 819–841. <https://doi.org/10.1111/jsbm.12089>
- Beck, T., Lu, L., & Yang, R. (2015). Finance and growth for microenterprises: Evidence from rural China. *World Development*, 67, 38–56. <https://doi.org/10.1016/j.worlddev.2014.10.008>
- Brandstätter, H. (2011). Personality aspects of entrepreneurship: A look at five meta-analyses. *Personality and Individual Differences*, 51(3), 222–230. <https://doi.org/10.1016/j.paid.2010.07.007>
- Bruno, R. L., Bytchkova, M., & Estrin, S. (2013). Institutional determinants of new firm entry in Russia: A cross-regional analysis. *Review of Economics and Statistics*, 95(5), 1740–1749. [https://doi.org/10.1162/REST\\_a\\_00322](https://doi.org/10.1162/REST_a_00322)
- Brunton, G. D., Ahlstrom, D., & Li, H. L. (2010). Institutional theory and entrepreneurship: Where are we now and where do we need to move in the future? *Entrepreneurship Theory and Practice*, 34(3), 421–440. <https://doi.org/10.1111/j.1540-6520.2010.00390.x>
- Castano, M. S., Méndez, M. T., & Galindo, M.Á. (2015). The effect of social, cultural, and economic factors on entrepreneurship. *Journal of Business Research*, 68(7), 1496–1500. <https://doi.org/10.1016/j.jbusres.2015.01.040>
- Castano, M. S., Méndez, M. T., & Galindo, M.Á. (2016). The effect of public policies on entrepreneurial activity and economic growth. *Journal of Business Research*, 69(11), 5280–5285. <https://doi.org/10.1016/j.jbusres.2016.04.125>
- Cheng, Z., Guo, W., Hayward, M., Smyth, R., & Wang, H. (2021). Childhood adversity and the propensity for entrepreneurship: A quasi-experimental study of the Great Chinese Famine. *Journal of Business Venturing*, 36(1), Article 106063. <https://doi.org/10.1016/j.jbusvent.2020.106063>
- Cheng, Z., Li, J., Sun, G., & Wang, W. (2021). Call for papers: Entrepreneurship in China: Causes, outcomes and implications. *China Economic Review*. Retrieved from <https://www.sciencedirect.com/journal/china-economic-review/about/call-for-papers#entrepreneurship-in-china-causes-outcomes-and-implications> Retrieved from.
- Cheng, Z., & Smyth, R. (2021). Education and migrant entrepreneurship in urban China. *Journal of Economic Behavior and Organization*, 188, 506–529. <https://doi.org/10.1016/j.jebo.2021.05.040>
- Chowdhury, S. (2005). Demographic diversity for building an effective entrepreneurial team: Is it important? *Journal of Business Venturing*, 20(6), 727–746. <https://doi.org/10.1016/j.jbusvent.2004.07.001>
- Churchill, S. A., Munyanyi, M. E., Smyth, R., & Trinh, T. A. (2021). Early life shocks and entrepreneurship: Evidence from the Vietnam war. *Journal of Business Research*, 124, 506–518. <https://doi.org/10.1016/j.jbusres.2020.11.009>
- Davis, P. S., Babakus, E., Englis, P. D., & Pett, T. (2010). The influence of CEO gender on market orientation and performance in service small and medium-sized service businesses. *Journal of Small Business Management*, 48(4), 475–496. <https://doi.org/10.1111/j.1540-627X.2010.00305.x>
- De Carolis, D. M., & Saporito, P. (2006). Social capital, cognition, and entrepreneurial opportunities: A theoretical framework. *Entrepreneurship Theory and Practice*, 30(1), 41–56. <https://doi.org/10.1111/j.1540-6520.2006.00109.x>
- Evans, D., & Leighton, L. (1989). Some empirical aspects of entrepreneurship. *American Economic Review*, 79(3), 519–535.
- Falck, O., Heblich, S., & Luedemann, E. (2012). Identity and entrepreneurship: Do school peers shape entrepreneurial intentions? *Small Business Economics*, 39(1), 39–59. <https://doi.org/10.1007/s11187-010-9292-5>
- Farinha, L., Lopes, J., Bagchi-Sen, S., Sebastião, J. R., & Oliveira, J. (2020). Entrepreneurial dynamics and government policies to boost entrepreneurship performance. *Socio-Economic Planning Sciences*, 72, Article 100950. <https://doi.org/10.1016/j.seps.2020.100950>
- Frese, M., & Gielnik, M. M. (2014). The psychology of entrepreneurship. *Annual Review of Organizational Psychology and Organizational Behavior*, 1(1), 413–438. <https://doi.org/10.1146/annurev-orgpsych-031413-091326>
- Heineck, G., Viinikainen, J., Böckerman, P., Hintsanen, M., Raitakari, O., & Pehkonen, J. (2017). Born entrepreneur? Adolescents' personality characteristics and self-employment in adulthood. *Journal of Business Venturing Insights*, 8, 9–12.
- Huang, B., Tani, M., & Zhu, Y. (2021). Does higher education make you more entrepreneurial? Causal evidence from China. *Journal of Business Research*, 135, 543–558. <https://doi.org/10.1016/j.jbusres.2021.06.044>
- Kollmann, T., Hensellek, S., Jung, P. B., & de Cruppe, K. (2022). How bricoleurs go international: A European cross-country study considering the moderating role of governmental entrepreneurship support programs. *Journal of Technology Transfer*. <https://doi.org/10.1007/s10961-022-09948-8> (Ahead-of-print).
- Krueger, N., Liñán, F., & Nabi, G. (2013). Cultural values and entrepreneurship. *Entrepreneurship and Regional Development*, 25(9–10), 703–707. <https://doi.org/10.1080/08985626.2013.862961>
- Laferrière, A., & McEntee, P. (1995). Self-employment and intergenerational transfers of physical and human capital: An empirical analysis of French data. *Economic and Social Review*, 27(1), 43–54.
- Leippold, M., Wang, Q., & Zhou, W. (2022). Machine learning in the Chinese stock market. *Journal of Financial Economics*, 145(2), 64–82. <https://doi.org/10.1016/j.jfineco.2021.08.017>
- Lévesque, M., & Minniti, M. (2011). Age matters: How demographics influence aggregate entrepreneurship. *Strategic Entrepreneurship Journal*, 5(3), 269–284. <https://doi.org/10.1002/sej.117>

- Lindquist, M. J., Sol, J., & Van Praag, M. (2015). Why do entrepreneurial parents have entrepreneurial children? *Journal of Labor Economics*, 33(2), 269–296. <https://doi.org/10.1086/678493>
- Littlewood, D., & Holt, D. (2018). Social entrepreneurship in South Africa: Exploring the influence of environment. *Business and Society*, 57(3), 525–561. <https://doi.org/10.1177/0007650315613293>
- Liu, Q., Tang, J., & Tian, G. G. (2013). Does political capital create value in the IPO market? Evidence from China. *Journal of Corporate Finance*, 23, 395–413. <https://doi.org/10.1016/j.jcorpfin.2013.10.002>
- Luis-Rico, M. I., Escolar-Llamazares, M. C., De la Torre-Cruz, T., Herrero, Á., Jiménez, A., Arranz Val, P., ... Jiménez-Eguizabal, A. (2020). The association of parental interest in entrepreneurship with the entrepreneurial interest of Spanish youth. *International Journal of Environmental Research and Public Health*, 17(13), 4744. <https://doi.org/10.3390/ijerph17134744>
- Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). *Forecasting Methods and Applications*. New York: John Wiley & Sons.
- Melançon, G. G., Grangier, P., Prescott-Gagnon, E., Sabourin, E., & Rousseau, L. M. (2021). A machine learning-based system for predicting service-level failures in supply chains. *INFORMS Journal on Applied Analytics*, 51(3), 200–212. <https://doi.org/10.1287/inte.2020.1055>
- Méndez-Picazo, M. T., Galindo-Martín, M.A., & Ribeiro-Soriano, D. (2012). Governance, entrepreneurship and economic growth. *Entrepreneurship and Regional Development*, 24(9–10), 865–877. <https://doi.org/10.1080/08985626.2012.742323>
- Minniti, M., & Nardone, C. (2007). Being in someone else's shoes: The role of gender in nascent entrepreneurship. *Small Business Economics*, 28(2–3), 223–238. <https://doi.org/10.1007/s1187-006-9017-y>
- Morris, M., & Schindehutte, M. (2005). Entrepreneurial values and the ethnic enterprise: An examination of six subcultures. *Journal of Small Business Management*, 43(4), 453–479. <https://doi.org/10.1111/j.1540-627X.2005.00147.x>
- Nair, K. R. G., & Pandey, A. (2006). Characteristics of entrepreneurs: An empirical analysis. *Journal of Entrepreneurship*, 15(1), 47–61. <https://doi.org/10.1177/097135570501500104>
- Nicolaou, N., Shane, S., Cherkas, L., Hunkin, J., & Spector, T. D. (2008). Is the tendency to engage in entrepreneurship genetic? *Management Science*, 54(1), 167–179. <https://doi.org/10.1287/mnsc.1070.0761>
- Obschonka, M., Silbereisen, R. K., Cantner, U., & Goethner, M. (2015). Entrepreneurial self-identity: Predictors and effects within the theory of planned behavior framework. *Journal of Business and Psychology*, 30(4), 773–794. <https://doi.org/10.1007/s10869-014-9385-2>
- Pflugfelder, E. H. (2017). Reddit's "Explain Like I'm Five": Technical descriptions in the wild. *Technical Communication Quarterly*, 26(1), 25–41. <https://doi.org/10.1080/10572252.2016.1257741>
- Rao, Y., Mei, L., & Zhu, R. (2016). Happiness and stock-market participation: Empirical evidence from China. *Journal of Happiness Studies*, 17(1), 271–293. <https://doi.org/10.1007/s10902-014-9594-4>
- Repetti, R. L., & Cosmas, K. A. (1991). The quality of the social environment at work and job satisfaction. *Journal of Applied Social Psychology*, 21(10), 840–854. <https://doi.org/10.1111/j.1559-1816.1991.tb00446.x>
- Rosado-Cubero, A., Freire-Rubio, T., & Hernández, A. (2022). Entrepreneurship: What matters most. *Journal of Business Research*, 144, 250–263. <https://doi.org/10.1016/j.jbusres.2022.01.087>
- Santos, F. J., Roomi, M. A., & Liñán, F. (2016). About gender differences and the social environment in the development of entrepreneurial intentions. *Journal of Small Business Management*, 54(1), 49–66. <https://doi.org/10.1111/jsbm.12129>
- Santos, S. C., Neumeyer, X., & Morris, M. H. (2019). Entrepreneurship education in a poverty context: An empowerment perspective. *Journal of Small Business Management*, 57(sup1), 6–32. <https://doi.org/10.1111/jsbm.12485>
- Shrestha, Y. R., He, V. F., Puranam, P., & von Krogh, G. (2021). Algorithm supported induction for building theory: How can we use prediction models to theorize? *Organization Science*, 32(3), 856–880. <https://doi.org/10.1287/orsc.2020.1382>
- Staniewski, M. W., & Awruk, K. (2021). Parental attitudes and entrepreneurial success. *Journal of Business Research*, 123, 538–546. <https://doi.org/10.1016/j.jbusres.2020.10.039>
- Su, Y. S., Lien, D., & Yao, Y. (2022). Economic growth and happiness in China: A Bayesian multilevel age-period-cohort analysis based on the CGSS data 2005–2015. *International Review of Economics and Finance*, 77, 191–205. <https://doi.org/10.1016/j.iref.2021.09.018>
- Ullhøi, J. P. (2005). The social dimensions of entrepreneurship. *Technovation*, 25(8), 939–946. <https://doi.org/10.1016/j.technovation.2004.02.003>
- Uysal, Ş. K., Karadağ, H., Tuncer, B., & Şahin, F. (2022). Locus of control, need for achievement, and entrepreneurial intention: A moderated mediation model. *International Journal of Management Education*, 20(2), Article 100560. <https://doi.org/10.1016/j.ijme.2021.100560>
- Van der Sluis, J., Van Praag, M., & Vijverberg, W. (2008). Education and entrepreneurship selection and performance: A review of the empirical literature. *Journal of Economic Surveys*, 22(5), 795–841. <https://doi.org/10.1111/j.1467-6419.2008.00550.x>
- Wang, G., Hao, J., Ma, J., & Jiang, H. (2011). A comparative assessment of ensemble learning for credit scoring. *Expert Systems with Applications*, 38(1), 223–230. <https://doi.org/10.1016/j.eswa.2010.06.048>
- Wang, J., Chang, C., Yao, S., & Liang, C. (2016). The contribution of self-efficacy to the relationship between personality traits and entrepreneurial intention. *Higher Education*, 72(2), 209–224. <https://doi.org/10.1007/s10734-015-9946-y>
- Williamson, A. J., Drencheva, A., & Battisti, M. (2020). Entrepreneurial disappointment: Let down and breaking down, a machine-learning study. *Entrepreneurship Theory and Practice*. <https://doi.org/10.1177/1042258720964447>
- Wilmoth, J., & Koso, G. (2002). Does marital history matter? Marital status and wealth outcomes among preretirement adults. *Journal of Marriage and Family*, 64(1), 254–268. <https://doi.org/10.1111/j.1741-3737.2002.00254.x>
- Xu, F., Pan, Z., & Xia, R. (2020). E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework. *Information Processing and Management*, 57(5), Article 102221. <https://doi.org/10.1016/j.ipm.2020.102221>
- Yan, Y., & Xu, X. (2022). Does entrepreneur invest more in environmental protection when joining the communist party? Evidence from Chinese private firms. *Emerging Markets Finance and Trade*, 58(3), 754–775. <https://doi.org/10.1080/1540496X.2020.1848814>
- Yang, J., Huang, J., Deng, Y., & Bordignon, M. (2020). The rise of red private entrepreneurs in China: Policy shift, institutional settings and political connection. *China Economic Review*, 61, Article 101431. <https://doi.org/10.1016/j.chieco.2020.101431>
- Yu, W., Dai, S., Liu, F., & Yang, Y. (2023). Matching disruptive innovation paths with entrepreneurial networks: A new perspective on start-ups' growth with Chinese evidence. *Asian Business and Management*, 22, 878–902. <https://doi.org/10.1057/s41291-022-00177-3>