



ARTICLE



<https://doi.org/10.1057/s41599-024-03695-x>

OPEN

What drives intercity venture capital investment? A comparative analysis between multiple linear regression and random forest

Delin Du¹ , Jiaoe Wang^{1,2}  & Jianjun Li³

Venture capital (VC) significantly contributes to the development of regional economies and fosters innovation. Analyzing the factors that influence VC investments holds key importance. This study employs two methods to ascertain the relative significance of different factors at the city level: the Lindeman, Merenda, and Gold (LMG) approach in multiple linear regression (MLR) and variable importance in random forest (RF) machine learning. The findings reveal that several factors, including economy, finance, innovation, location, and policy, significantly influence VC investments. Both the MLR and RF models highlight the preeminence of economic and financial variables, followed closely by the city's potential for innovation. Moreover, spatial heterogeneity exists in the importance of these variables. In the economically developed and densely populated eastern regions of China, the financial environment of cities emerges as the most crucial, whereas in the central and western regions, the economy and innovation, respectively, take precedence. This research contributes to a deeper understanding of the distribution of VC investments and offers valuable insights for the development of regional policies.

¹Key Laboratory of Regional Sustainable Development Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China. ²College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China. ³School of Business, East China University of Science and Technology, Shanghai 200237, China. email: wangje@igsrr.ac.cn

Introduction

The swift evolution of technology in recent decades has revolutionized business models, industries, and entire economies. Start-ups have been pivotal in driving this transformation. Scholars increasingly emphasize the role of start-ups as catalysts for economic growth (Ressin 2022). Venture capital (VC) has emerged as a critical financial resource for supporting start-ups and early-stage enterprises. It has played a pivotal role in fostering innovation, propelling technological advancements, and driving economic expansion, consequently capturing the attention of researchers and policymakers. Understanding the intricacies of the factors that influence VC investment decision-making is essential for facilitating regional development.

Existing studies have analyzed the factors that influence VC investments, encompassing both the attributes of the involved parties and external elements (Gompers et al. 2020; Guzella et al. 2024; Zhou et al. 2023). The former refer to the qualities of the VC firms and start-ups themselves. For instance, the reputation, experience, and industry specialization of VC firms significantly influence their investment decisions (Hsu 2004; Liu et al. 2023). Concerning the targets of investment, considerations including growth potential and the traits of entrepreneurs and management teams are widely recognized as crucial determinants (Pintado et al. 2007). The external elements that influence VC investments encompass a range of broader economic and market conditions, institutions, and rules, industry trends, and technological advancements, among other factors (Dai and Nahata 2016; Yang et al. 2024).

The trend of financial globalization has fostered the expansion of intercity VC activities. Notably, substantial VC flows from New York and Chicago to other high-technology regions in the US have been observed (Florida and Smith 1993), while Beijing, Shanghai, and Shenzhen exhibit widespread geographical coverage in terms of VC flows in China (Pan et al. 2016). Greater information asymmetry and heightened supervision and management costs are often encountered in intercity VC investments compared to local investments (Cumming and Dai 2010). Consequently, the dynamics of institutional relationships, cultural factors, distance, and transportation between cities have garnered attention (Félix et al. 2023; Martin et al. 2005; Wu et al. 2022; Yang et al. 2024). Robust economies and financial markets, coupled with promising potential for innovation and growth, significantly attract investors (Chen et al. 2010; Fang 2018). This is also evidenced by the concentration of VC activities in financial centers and high-tech industrial areas, where there is an abundance of skilled professionals, banks, legal firms, and other commercial institutions (Florida and Smith 1993).

These different factors influence capital flow in varying ways, making it crucial to understand the relative importance of these factors for comprehending the spatial distribution of VC investments. For enterprises, clarifying determinants informs VC investment decisions and strategic planning. Some studies have investigated the importance of factors that attract venture capitalists and entrepreneurs through interviews and surveys (Macmillan et al. 1985; Pintado et al. 2007). They found that VC firms prioritize the management team over business-related characteristics like product or technology (Gompers et al. 2020). For regions and cities, identifying key determinants helps policymakers and stakeholders allocate resources effectively, enhancing regional attractiveness to VC and guiding comprehensive development strategies to improve financial capital distribution.

However, existing studies have primarily focused on the attributes of investors and recipients of investments. There is a research gap in the limited understanding of the relative importance of city-level factors that influence VC investments.

Recognizing the significance of these factors is crucial for urban governance and marketing. Exploring this research not only provides valuable insights for enterprise positioning and investment decisions but is also pertinent to regional studies and policy development. Therefore, this study aims to offer a comprehensive examination of the factors influencing VC investments at the city level.

This study investigated the relative importance of various explanatory variables and their detailed relationships with intercity VC investments using variable importance metrics in regression models. To account for the potential linear and nonlinear characteristics of different variables, we employed two distinct methodologies: multiple linear regression (MLR) and the random forest (RF) algorithms. These approaches provide effective means of assessing the relative importance of variables (Liu et al. 2021). MLR identifies direct variable contributions, and, when combined with other predictors through variance decomposition, provides insights (Grömping 2015). RF excels at examining nonlinear relationships between dependent and independent variables (Breiman 2001). Integrating the results from both models offers a comprehensive understanding of the factors that influence VC investments.

Given China's significant role in the global VC market, the study focused on China as a case study. With over three decades of evolution, VC investments have become a prominent force in the financial market and have extended their reach across numerous cities throughout the nation. As per the Q4'22 Venture Pulse Report (<https://kpmg.com/xx/en/home/campaigns/2023/01/q4-venture-pulse-report-global.html>), the fourth quarter of 2022 witnessed China dominating the majority of megadeals exceeding \$500 million. The findings have implications for investment strategies and policy interventions, thereby contributing to the sustainable growth of the VC ecosystem.

This paper is organized as follows. The subsequent section provides an overview of the literature, followed by the data and methods in Section 3. Section 4 presents the results, encompassing a comparative analysis of the MLR and RF models, alongside an examination of the spatial heterogeneity of the determinants. Finally, Section 5 concludes the paper by summarizing the results and analyzing the limitations of the study.

Literature review

Influencing factors of VC investments. Many studies have analyzed the factors influencing VC investments, including the characteristics of investors, investees, and the external environment, as summarized in Table 1. Investors include venture capitalists, VC firms, or VC funds, whose characteristics—such as reputation, network, age, location, and collaboration—impact both project evaluation and selection (e.g., Falik et al. 2016; Hochberg et al. 2010; Wray 2012), as well as entrepreneurs' decisions (e.g., Hsu 2004). For example, reputable VC firms exhibit less local bias due to their ability to overcome information asymmetry (Cumming and Dai 2010). According to Hsu (2004), investments from high-reputation VC firms are three times more likely to be accepted, and these firms can acquire start-up equity at a discount.

As for investees, the characteristics of start-ups, such as the management team, product, industry sector, innovation capability, scale, and political background, can significantly influence the future development and returns (e.g., Blank and Carmeli 2021; Zhou et al. 2023). The entrepreneur's personality, experience, and entrepreneurship can also affect the likelihood of a project's selection and success (e.g., Gompers et al. 2020; Pintado et al. 2007). Blank and Carmeli (2021) highlighted a

Table 1 Influencing factors of VC investments.

Objects	Categories	Main characteristics	Examples
Investor	Venture Capitalist/VC firm/VC fund	<ul style="list-style-type: none"> • Reputation • Network • VC firm age • Fund location 	Cumming and Dai 2010; Dimov and Murray 2008; Falik et al. 2016; Wray 2012
Investee	Start-up	<ul style="list-style-type: none"> • Snydication • Management team • Production and market acceptance • Industry sector • Technology/Innovation • Stage/Size • Political background 	Blank and Carmeli 2021; Gompers et al. 2020; Macmillan et al. 1985; Zhou et al. 2023
	Entrepreneur	<ul style="list-style-type: none"> • Personality • Experience • Entrepreneurship 	Blank and Carmeli 2021; Macmillan et al. 1985; Pintado et al. 2007
External factors	Economy	<ul style="list-style-type: none"> • GDP • Service industries • Economic freedom 	Ning et al. 2015; Wang and Wang 2012; Wu et al. 2022
	Finance	<ul style="list-style-type: none"> • IPO • Concentration of VC firms • Exit opportunity 	Du et al. 2024; Félix et al. 2023; Pan et al. 2016
	Innovation	<ul style="list-style-type: none"> • Invention • University • High-end human capital 	Cheng et al. 2019; Fang 2018; Powell et al. 2002
	Location/Transportation	<ul style="list-style-type: none"> • Spatial proximity • Air network • High-speed railway 	Chen et al. 2010; Colombo et al. 2019; Duan et al. 2020; Martin et al. 2005; Zheng et al. 2020
	Policy/Institution	<ul style="list-style-type: none"> • Institution • Government intervention • Public policy 	Li and Zahra 2012; Mason and Pierrakis 2013; Yang et al. 2024; Zacharakis et al. 2007
	Others	<ul style="list-style-type: none"> • Culture • Social ties 	Bringmann et al. 2018; Li and Zahra 2012

positive correlation between founding teams' experience and the acquisition of external investments. Zhou et al. (2023) suggested that while entrepreneurial experience alone is insufficient to attract VC, industrial experience and political background positively influence VC attraction.

In addition to the characteristics of both sides of investments, external factors also significantly affect the spatial distribution of VC activities. In general, urban economy, finance, innovation, location and transportation, and institution and policy all have certain influence. The urban economy and finance assume an important role in shaping the landscape of VC activities, acting as fertile ground for innovation, entrepreneurship, and investment prospects. Instances of new firm launches tend to escalate during periods of macroeconomic growth (Audretsch and Acs 1994), presenting significant investment opportunities. As VC firms aim for high returns, public listing and mergers and acquisitions are commonly viewed as successful exit strategies, particularly the public listing of the invested enterprise, which not only enables VC firms to yield profits but also provides entrepreneurs or management teams with a call option. Chen et al. (2010) observed that VC investments tend to concentrate in regions with a high success rate of previous investments over five years. Motivated by the potential for high returns connected with innovation, VC firms tend to prioritize targets with robust innovative capabilities (Hirukawa and Ueda 2011). Consider the biotechnology industry, for instance; start-ups and VC activities are predominantly situated in proximity to colleges and universities (Powell et al. 2002).

The distance and proximity between VC firms and entrepreneurs can influence communication, affecting the information asymmetry between both parties. Local bias significantly

represents the spatial investment choices of VC firms (Cumming and Dai 2010). However, advancements in transportation technology, especially high-speed transportation like civil aviation and high-speed railways, have mitigated geographical constraints and influenced the spatial distribution of VC investments (Duan et al. 2020; Zheng et al. 2020). Regional differences in regulations and policies can also impact VC firms' decisions. In cross-border investments, institutional proximity plays a crucial role (Yang et al. 2024; Zacharakis et al. 2007). Additionally, factors such as culture, and social connections can influence a VC firm's project selection (Bringmann et al. 2018; Li and Zahra 2012).

Overall, these factors play varying roles in the spatial flow of VC investments. Some studies have combined different factors to analyze their effects and explore the relative importance of different characteristics of venture capitalists and entrepreneurs (Gompers et al. 2020; Pintado et al. 2007). For example, Pintado et al. (2007) found that honesty and integrity are the most important indicators for venture capitalists, followed by sector knowledge, work experience, and management team. As for factors at the city level, they are important for regional studies and policy development, as government intervention in investor and entrepreneur characteristics is limited. However, there is a research gap regarding the relative importance of the factors at the city level. Some studies have compared the significance of different factors (Aizenman and Kendall 2012; Cheng et al. 2019; Fang 2018), but they have not identified which factors are more important. This study aims to comprehensively examine various determinants, including economic, financial, innovation, location, and policy aspects, to investigate their impact and relative importance on intercity VC investments.

Main analyzing approaches for influencing factors of VC investments. Existing studies have used various analytical methods to measure the impact of multifarious factors on the selection and distribution of VC investments, including interviews and surveys (e.g., Falik et al. 2016; Li and Yang 2022), regression analysis (e.g., Bellucci et al. 2023; Blank and Carmeli 2021), and machine learning algorithms (e.g., Bai and Zhao 2021; Zhang et al. 2023).

Interview and survey are traditional analyzing approaches. Through direct communication or questionnaire collection with key players such as venture capitalists, entrepreneurs, industry experts, and project managers, researchers can get first-hand detailed and in-depth data information. By quantifying (e.g., scoring each factor) and counting the feedback information from respondents, the determinants affecting VC investments can be identified, such as the analysis by Falik et al. (2016). The advantage of this approach is the ability to tailor the questions to the different backgrounds and experiences of the interviewees to gain insight into project information, especially the complex information that is not easy to quantify. For example, personalities such as honesty of entrepreneurs and the risk appetite of investors, which are important factors (Li and Yang 2022; Pintado et al. 2007), can be obtained through this method.

Regression analysis is a common method for explaining the influence of various factors on VC investments, including algorithms such as multiple linear regression (Cheng et al. 2019), logit regression (Falik et al. 2016), Poisson regression (Blank and Carmeli 2021), and difference-in-difference regression (Bellucci et al. 2023). This method can be applied to data from questionnaire surveys or various databases. For example, Falik et al. (2016) used logit regression to deal with survey data from Israeli entrepreneurs to explore the importance of start-up experience. Cheng et al. (2019) utilized VC data from Zero2IPO Group and applied multiple linear regression and spatial autoregressive models to investigate the significance of the urban factors.

The development of machine learning methods provides a significant means for analyzing the factors influencing venture capital. Machine learning models can automatically extract patterns and features from data, enhancing the ability to handle large datasets and complex situations. Some studies have used specific models, such as decision tree, RF, and adaptive moment estimation method, to study the factors influencing VC investment decisions (Bai and Zhao 2021; Zhang et al. 2023). These models have identified the key determinants in investment decision-making, such as the planning strategy and team management (Bai and Zhao 2021).

These methods have their own advantages and disadvantages. Interviews and surveys, for instance, can be time-consuming, costly, and limited in sample size, making them inefficient for large-scale, multi-city VC research. Regression analysis can determine the significance of each factor, but traditional MLR models often lack the direct ability to assess variable importance (Grömping 2015). Many studies employ correlation and regression coefficients to assess variable importance, but these are unsuitable when variables are cross correlated, as this may lead to unstable and misleading results. Additionally, some nonlinear factors in VC activities (Du et al. 2024) may not be effectively identified. Advanced statistical methods have been developed to measure variable importance, including variance decomposition, variable transformation, and machine learning algorithms (Bi 2012; Grömping 2015). Methods based on variance decomposition are most widely used in MLR, including the Lindeman, Merenda, and Gold method (LMG; Lindeman et al. 1980), dominance analysis (Budesu 1993), and proportional marginal variance decomposition (Feldman 2005). RF algorithm, a

nonparametric regression technique, also provides a specific approach to derive variable importance (Breiman 2001).

Therefore, to identify the role of factors affecting intercity VC investments more accurately—including their significance, direction, and importance—this study employs both LMG and RF methods. LMG identifies the significance and directionality of each factor based on linear regression, while the RF algorithm assesses the relative importance of each factor from a nonlinear perspective. The results of these two methods can also complement and verify each other.

Data and methods

Data processing

Dependent variable: VC investments. This study utilized data on VC investment deals in China to assess VC investments at the city level. The data were sourced from the CVSource database (<http://www.cvsource.com.cn/>), one of the prominent VC databases in China. It provides comprehensive information on VC investment transactions, including the locations of VC firms and invested enterprises, along with transaction timelines. Although certain investment deals involve collaborative efforts among multiple VC firms, the term “leading firm” typically refers to the entity that assumes primary responsibility for spearheading the investment round and plays a more pivotal role compared to other participants. Consequently, we identified the cities of the leading VC firms and the invested enterprises based on their respective addresses. The study encompasses 337 cities in mainland China, consisting of 4 municipalities and 333 prefecture-level cities. Additionally, to mitigate the influence of the COVID-19 pandemic in 2019, we employed data from 2018 as the study sample. To eliminate the impact of local bias, our analysis focuses exclusively on intercity VC investments and excludes intracity VC investments.

Consequently, we obtained statistics and distribution of VC investments at the city level in China in 2018, as shown in Fig. 1. It can be observed that VC investments show the obvious characteristics of the spatial agglomeration. Cities situated in the eastern region, particularly those in the economically prosperous and densely populated regions, such as the Yangtze River Delta, Pearl River Delta, and Beijing-Tianjin-Hebei regions, attracted a significant number of VC inflows. VC investments in the central and western regions of China were comparatively scarce. However, some provincial capitals in these regions, such as Chengdu, Chongqing, Wuhan, Changsha, and Xi'an, were able to attract a considerable amount of VC inflows.

Independent variables: selection and pre-processing. Drawing on the analysis of the existing literature, this study has identified the independent variables related to the five dimensions of economy, finance, innovation, location, and policy. These variables were chosen to examine the determinants that influence VC investments at the city level, as indicated in Table 2.

- (1) **Economic Factors:** Our selection encompasses the following indicators, i.e., GDP (Gross Domestic Product), GDP growth rate, and the share of the tertiary industry. These metrics offer valuable insights into the scale, expansion, and industrial structure of a city's economy.
- (2) **Financial Factors:** To grasp the financial dynamics and market vitality within a city, we considered four key indicators: the count of newly established enterprises over the past five years, the presence of VC firms, bank deposits, and the marketization index of the cities.
- (3) **Innovation Dynamics:** Recognizing the pivotal role of innovation in fostering a robust start-up ecosystem, we included indicators such as the number of universities,

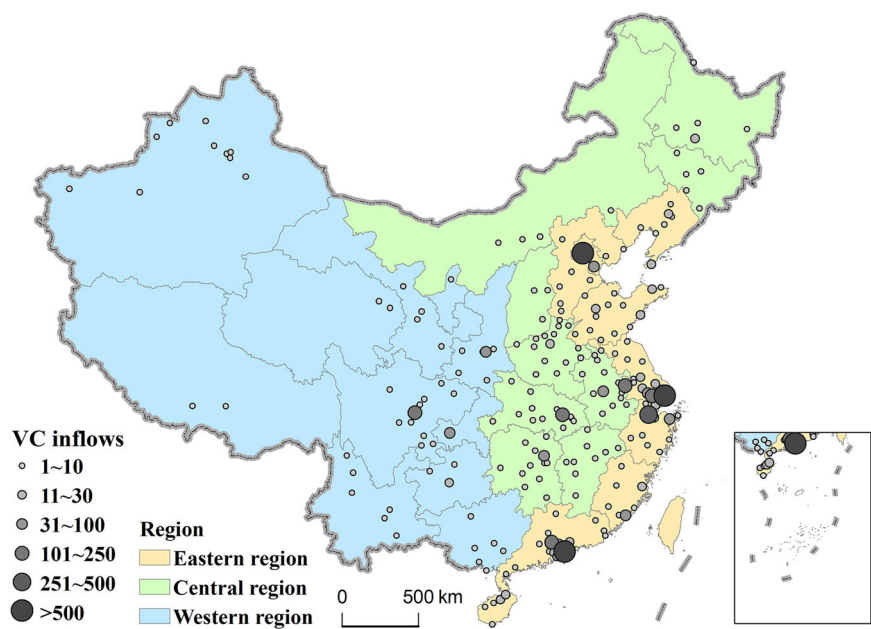


Fig. 1 Inflow of VC investments at the city level in 2018. Note: The dots indicate the number of VC investment inflows from other Chinese cities, with larger dots representing higher levels of investments.

Table 2 Description of independent variables.			
Variable		Abbreviation	Source
Economy	Gross domestic product	GDP	China City Statistical Yearbook 2018
	Growth of GDP	Growth	
Finance	Percentage of the tertiary industry to GDP	Tertiary	Qichacha database (https://www.qcc.com/)
	Number of newly listed enterprises in recent five years	IPO	
	Number of VC firms	VCF	CVSource database (http://www.cvsources.com.cn/)
	Deposits of national banking system at year end	Deposit	China City Statistical Yearbook 2018
	China marketization index	Marketization	China's Provincial Market Index Report 2018.
Innovation	Number of universities	University	China City Statistical Yearbook 2018
	Number of undergraduates in universities	Student	
	Research and development internal outlay	RaD	
	Number of granted patents	Patent	
Location	Presence of international airport	Airport	OAG (Official Aviation Guide) database
	Degree centrality of airline network	AIR	
	Degree centrality of high-speed train network	HSR	12306 China Railway
	The shortest distance between the city and China's VC centers (i.e., Beijing, Shanghai, and Shenzhen)	DIS_Center	Baidu Maps
Policy	Whether it is a provincial capital or municipality	Administration	Retrieved from government website
	Whether there is a national economic and technological development zone	ETDZ	
	Whether there is a national high-tech industrial zone	HTIZ	Baidu Index (http://index.baidu.com)
	Baidu Index, a metric that measures the frequency of searches for specific keywords	Baidu_Index	

- student population, research and development (R&D) investments, and patents. These metrics reflect the educational level, talent pool, capital infusion, and intellectual property resources of the cities.
- (4) **Location Conditions:** Acknowledging the significance of a city's access to external resources, we integrated variables such as the presence of international airports, flight connections, high-speed train services, and the proximity of each city to China's major VC centers, namely Beijing, Shanghai, and Shenzhen (Pan et al. 2016). Using schedule data from flights and high-speed trains, we constructed transportation networks based on the methodology proposed by Yang et al. (2023) and computed the degree centrality of each city within the networks.
- (5) **Policy Environment:** Our analysis considered factors such as administrative levels, economic and technological development zones, and industrial zones as crucial indicators of the policy landscape. To capture the visibility and prominence of each city, we also incorporated the Baidu Index, a metric that measures the frequency of searches for specific keywords.
- The amount of VC investments in each city can provide a comprehensive view of the city's investment landscape and

Table 3 Collinearity test and selection of variables.			
Variables	Step 1	Step 2	Step 3
GDP	13.841	7.139	6.929
Growth	1.486	1.319	1.247
Tertiary	1.923	1.813	1.811
IPO	4.559	4.461	3.803
VCF	4.234	4.218	4.217
Deposit	22.049		
Marketization	2.755	2.529	2.219
University	7.842	7.780	7.776
Student	4.678	4.609	4.397
RaD	5.118	5.116	4.532
Patent	10.989	10.086	
Airport	3.197	3.182	3.171
AIR	2.835	2.769	2.767
HSR	1.661	1.639	1.632
DIS_Center	2.398	2.289	2.278
Administration	4.477	4.191	4.176
ETDZ	1.542	1.534	1.450
HTIZ	1.849	1.827	1.808
Baidu_Index	2.039	2.019	2.003

attractiveness. Therefore, in line with many existing studies (Fang 2018; Guzella et al. 2024; Mason and Pierrakis 2013; Pan et al. 2020), most variables use the total amount, except for logical variables and those related to distance and marketization index.

To address the collinearity issue and uneven data distribution, we applied logarithmic transformations to all variables except for Airport, Administration, ETDZ, and HTIZ, which are logical variables. To avoid complications stemming from zero values, we adjusted the variables by adding 1 prior to applying the logarithmic transformation. Subsequently, we employed the variance inflation factor (VIF) to address multicollinearity among the explanatory variables. We employed the full MLR model, encompassing all explanatory variables, and progressively eliminated variables with VIF values greater than 10. This iterative process was repeated with the reduced model until no variables exhibited a VIF exceeding 10. The process and its outcomes are detailed in Table 3. Consequently, Deposit and Patent were excluded due to their significant correlation with indicators such as GDP and R&D. The remaining 18 variables demonstrated VIF values below 10 and were thus utilized as inputs for the regression model. Notably, all independent variables are reflective of values from 2017, which is the preceding year for the dependent variables.

Regression models and variable importance. Variance decomposition and machine learning algorithms were used to gauge variable importance of intercity VC investments. Variance decomposition methods are commonly applied in the MLR model, deconstructing the coefficient of determination (R^2) into non-negative contributions assigned to each explanatory variable based on an averaging procedure (Kruskal 1987). RF, as a representative machine learning algorithm, serves as a nonparametric regression technique for ascertaining variable importance (Breiman 2001). In this study, we employed both methods to determine the relative importance of the independent variables and discern the main determinants influencing the inflows of VC investments at the city level. All statistical analyses were conducted using R 4.3.0 software.

Parametric regression based on variance decomposition. Following the preceding variable selection and preprocessing steps, we excluded independent variables that exhibited high collinearity

($VIF > 10$). Subsequently, we employed a stepwise model selection approach guided by the Akaike information criterion (AIC) to further streamline the model and eliminate any lingering redundant variables (Lachniet and Patterson 2006). The stepwise regression process was executed using the “stepAIC” function within the MASS 7.3.58.4 package.

We employed the Lindeman, Merenda, and Gold (LMG) method (Lindeman et al. 1980) to evaluate the relative importance of the selected variables. The LMG approach offers a crucial advantage: its outcomes are independent of the order of predictors in the model. This method was preferred due to its clear utilization of both direct effects and effects adjusted for other variables within the model. Specifically, the LMG method deconstructs R^2 into non-negative contributions based on semi-partial coefficients (Kruskal 1987) and incorporates both direct effects and effects adjusted for other regressors in the model (Grömping 2006). In this study, the LMG values were computed using the R package “relaimpo 2.2.7.”

Referring to Gromping (2006), the R^2 of the MLR model can be represented as the ratio of the sum of squares of regression (SSR) to the sum of squares of total deviation (SST) for a given set of n variables. When a new variable M is incorporated into the model, the additional R^2 is defined as $seqR^2(M/N)$. In any given model, the arrangement of variables constitutes a permutation of the available variables x_1, \dots, x_n , denoted by the tuple of indices $r = (r_1, \dots, r_n)$. Let $S_k(r)$ signify the set of variables that have been included in the model prior to variable x_k in the order r . Consequently, the portion of R^2 assigned to the variable x_k in the order r can be expressed as $seqR^2(\{x_k\}/S_k(r))$.

$$seqR^2(M/N) = R^2(M \cup N) - R^2(N)$$
$$seqR^2(\{x_k\}/S_k(r)) = R^2(\{x_k\} \cup S_k(r)) - R^2(S_k(r))$$

Finally, for the independent variable x_k , its LMG value was calculated as follows:

$$LMG(x_k) = \frac{1}{n!} \sum_{r \text{ permutation}} seqR^2(\{x_k\}|r)$$

Random forest (RF) variable importance. RF is a commonly used machine learning algorithm that aggregates outcomes from multiple decision trees to yield unified results (Breiman 2001). It has been widely used to deal with classification and regression problems. The techniques for evaluating variable importance within a random forest have been extensively studied, surpassing those available for other machine learning methods (Grömping 2015).

RF offers advantages due to its fewer tuning hyperparameters compared to other machine learning methods. The key hyperparameters in RF include *mtry*, which represents the number of variables utilized at each split, and *ntree*, which represents the number of trees to grow (Belgiu and Drăguț 2016). Optimizing these parameters is vital for enhancing simulation accuracy. During the tuning process, the seed was fixed to ensure that the same sample size corresponded to the same samples, thereby enabling the attainment of reproducible results and facilitating comparisons across different parameters. We partitioned 80% of the data for the training set and 20% for the test set. Subsequently, we computed the variable importance utilizing the permutation-based mean square error (MSE) reduction, which is a widely adopted metric for assessing variable importance in RF regression (Grömping 2015; Sannabe 2022).

In this study, the RF model was fine-tuned in R by exploring various parameter configurations with the assistance of the “randomForest 4.7.1.1” package. The variable importance was determined using “%IncMSE,” which gauges the importance of each predictor. This metric operates by assigning a random value

to each predictor and subsequently evaluating the increase in the error of the model prediction when the value of a more important predictor is randomly replaced. A higher %IncMSE value signifies greater importance of the corresponding variable.

Results

Variable importance based on MLR. Through the stepwise regression process, the model was streamlined, resulting in the removal of 10 variables from the initial set of 18. The final model retained 8 key variables: GDP, serving as an indicator of the local economy; VCF, IPO, and Marketization, reflecting the status of the financial markets; University, representing the innovative capacity; DIS_Center, denoting the geographic location; and HTIZ and Baidu_Index, signifying the local policy environment. The final model accounted for an impressive 85.2% of the variance (adjusted $R^2 = 0.852$, residual standard error = 0.509, F-statistic = 206.6, p -value < $2.2e - 16$). A comprehensive summary of the selected variables can be found in Table 4, along with the corresponding LMG values calculated for each variable.

Notably, the LMG values for VCF and IPO, which indicate the extent of urban financial development, are the highest, both exceeding 0.2. Following closely is GDP, with an LMG value of 0.154. University holds the fourth position, boasting an LMG value of 0.147. Baidu_Index, DIS_Center, and HTIZ demonstrate LMG values ranging from 0.05 to 0.1. Marketization exhibits the lowest LMG value. All variables, except Marketization, possess a

p -value less than 0.1, underscoring their significant influence on VC investments. Furthermore, the coefficient of DIS_Center is negative, suggesting that proximity to VC centers fosters the development of VC investments. Conversely, the estimated coefficients of the remaining variables are all positive, implying that the city's attributes related to economy, finance, innovation, and policy exert a positive impact on VC investments.

Variable importance based on RF

Model training and parameter tuning. Figure 2 presents the variations in MSE and R2 based on mtry and ntree settings. The model achieves its optimal state at mtry = 4, with a minimum MSE of 0.288 and a maximum R^2 of 0.835. For ntree, the model stabilizes around a value of 100, with subsequent changes in MSE and R^2 being marginal. Given that ntree can be set to an extensive range (Belgiu and Drăguț 2016), we selected a value of 1000. The optimal model occurs at ntree = 627, with an MSE of 0.283 and an R^2 of 0.838. Finally, in the RF model, we set mtry to 4, ntree to 627, and sampsize to 80%. Figure 3 presents the simulation outcomes of the model in both the training and test sets, demonstrating a consistent and stable overall performance. The final RF model explains 83.69% of the variance in the observed data, with a mean squared residual of 0.284. The explanation accuracy of the RF model is close to, albeit slightly lower than, that of the MLR.

Variable importance. Table 5 presents the value of %IncMSE for each variable in the final RF model. The analysis indicates that the three most important variables are GDP, IPO, and VCF, which are all closely associated with the local economy and finance. A higher %IncMSE implies that the model's prediction error will significantly increase if the values of these variables are randomly altered. The variables University, Marketization, and Baidu_Index demonstrate similar %IncMSE values, ranging from 10 to 15. HTIZ and DIS_Center exhibit the lowest %IncMSE values, indicating their relatively lower importance within the model. Overall, the results of the RF model underscore the crucial role of a city's economic and financial environment in fostering VC investments, followed by the importance of its innovation potential. Proximity to financial centers appears to hold relatively less importance in this context.

Comparison between MLR and RF. The underlying rationales of the MLR and RF methods differ, leading to varying

Table 4 The MLR model results and the LMG values.					
Variables	Estimate	Std. Error	t value	P (> t)	LMG
VCF	0.408***	0.051	8.021	0.000	0.233
IPO	0.350***	0.056	6.207	0.000	0.213
GDP	0.131*	0.073	1.790	0.075	0.154
University	0.247***	0.063	3.924	0.000	0.147
Baidu_Index	0.330***	0.110	3.003	0.003	0.083
DIS_Center	−0.074*	0.045	−1.656	0.099	0.073
HTIZ	0.186**	0.076	2.461	0.014	0.064
Marketization	0.259	0.163	1.594	0.112	0.032
(Intercept)	−4.154***	1.166	−3.563	0.000	--

Variables are presented in descending order of the LMG metrics. The metrics were normalized to sum to 100%.
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

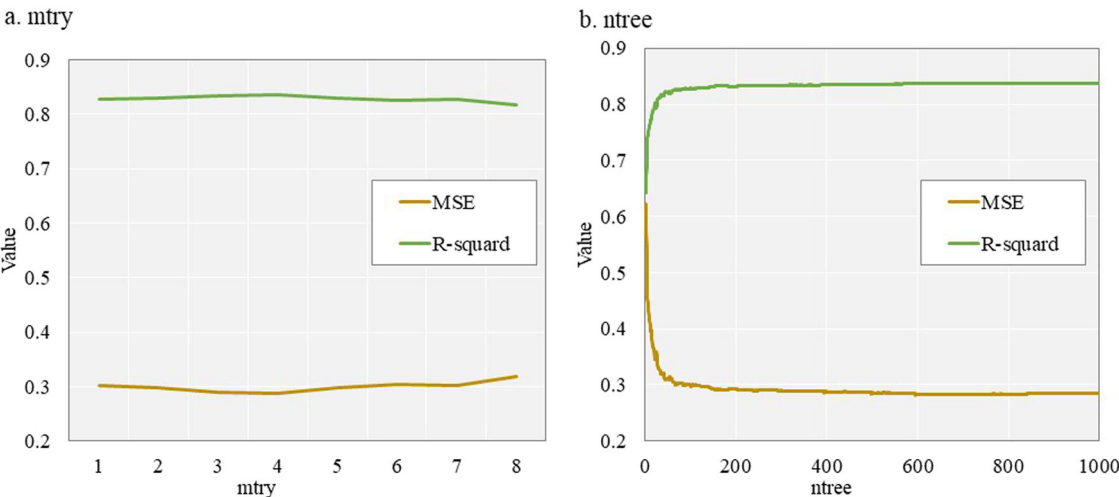


Fig. 2 The changes in simulation accuracy with mtry and ntree. Note: **a** Simulation accuracy was evaluated with different mtry values, with the optimal value being 5. **b** Simulation accuracy was evaluated with different ntree values, with the optimal value being 627.

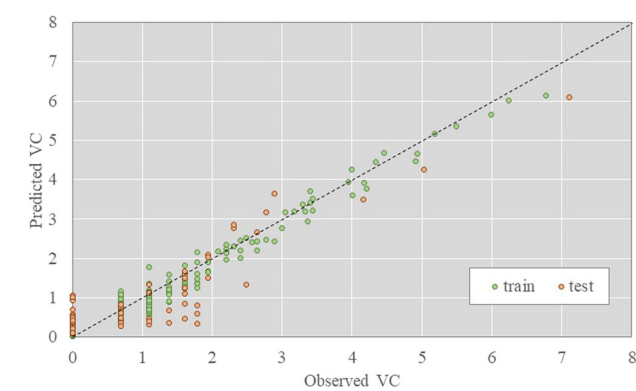


Fig. 3 Simulation of the final RF model in the training set and the test set.
Note: The green dots represent the observed and predicted values of the RF model using the training dataset, while the red dots correspond to the observed and predicted values based on the test dataset.

Table 5 Variable importance based on the RF model.	
Variables	%IncMSE
GDP	28.500
IPO	24.664
VCF	21.875
University	14.997
Marketization	13.331
Baidu_Index	11.752
HTIZ	8.589
DIS_Center	8.055

Variables are presented in descending order of the values of %IncMSE.

computational results. Comparing the findings of both models facilitates a more precise evaluation of the importance of different factors. To enable a seamless comparison, the LMG and % IncMSE in the models were standardized, and the results are depicted in Fig. 4.

It can be observed that in the results of the MLR model, VCF and IPO have a relatively high importance, both accounting for over 20%. In contrast, in the RF model, only GDP exceeds 20%. Following the MLR model, the importance of GDP and University each account for over 10%. In the RF model, IPO, VCF, University, and Marketization have a high importance, with IPO and VCF both approaching 20%. Notably, VCF, IPO, and GDP emerge as highly significant indicators in both models, emphasizing the pivotal role of local financial and economic development in VC investments. In addition, University consistently holds the fourth position in terms of relative importance, underscoring the significance of a city’s innovation potential in attracting VC inflows, as indicated by both the MLR and RF models.

In comparison to the aforementioned four indicators, other metrics demonstrate relatively lower importance in both the MLR and RF models. This includes Marketization, Baidu_Index, DIS_Center, and HTIZ. Baidu_Index and HTIZ exhibit similar levels of importance in both models. Baidu_Index’s LMG and % IncMSE scores are 8.32% and 8.92%, respectively, ranking 5th and 6th across the two models. Similarly, HTIZ holds the 7th position in both models. The importance of DIS_Center is comparable in the two models, although the ranking differs significantly. Moreover, Marketization displays a notable contrast in LMG and %IncMSE scores between the two models at 3.15% and 10.12%, respectively. In the MLR model, the impact of Marketization appears to be relatively less important.

To understand the cause of differences in the importance of variables between the models, bivariate relations between VC investments and the selected eight variables are plotted in Fig. 5.

The primary concern is the economic and financial conditions of the cities. As Fig. 4 indicates, the top three most important variables in terms of relative importance are VCF, IPO, and GDP. Not only can the development of the local economy foster the growth of new enterprises, increasing investment opportunities, but it can also bolster investors’ confidence in the smooth growth of businesses in the future, ultimately impacting the inflow of VC investments positively. In terms of finance, VCF signifies the local concentration of investment firms, promoting the inflow of external capital to a certain extent. Given that investment firms might encounter obstacles due to cultural and institutional disparities when investing in various locations, collaborating with local investment companies emerges as an effective means of resolution. Such a joint approach not only facilitates investment firms’ entry into new markets (Hochberg et al. 2010) but also aids in the accompanying regulatory process. IPO directly signifies a thriving business ecosystem in the city. The success of an enterprise also ensures the successful exit of VC investments and the realization of returns. Therefore, these three variables hold significant importance.

The slight difference is that GDP holds the highest importance in the RF model, whereas VCF and IPO exhibit the highest importance in the MLR model. As illustrated in Fig. 5, in comparison to GDP, VCF and IPO demonstrate a closer linear relationship with VC investments. Some studies have highlighted the high accuracy of RF and its capability to model intricate interactions between variables (Liu et al. 2021). If nonlinear relationships are significant, RF would exhibit greater accuracy than MLR. This observation partially accounts for why GDP appears more prominent in the RF model. Nevertheless, the consistent findings of both models underscore the crucial significance of local economics and finance.

Both University and Baidu_Index exhibit a considerable positive influence on VC. The relative importance of these two variables appears comparable in both models. Universities serve as critical fountains of new knowledge and technology, offering enterprises access to a large pool of skilled professionals and technicians. The bond between Stanford and Silicon Valley serves as a pertinent example in this context. Innovation, which is emblematic of growth potential, stands as a focal point for VC firms. Baidu_Index reflects the level of popularity a city enjoys within the network. A city that garners more attention within the network is more likely to attract the interest of VC firms.

VC centers are often characterized by a high concentration of investment firms, accompanied by a wealth of financial and business services. Proximity to these cities facilitates access to related resources. This is why DIS_Center exhibits a negative correlation with VC investments. This finding corresponds with existing research emphasizing the importance of proximity in VC investments. Scholars argue that closer proximity enables VC firms to obtain precise and effective information for projects (Hsu 2004), subsequently leading to enhanced investment performance (Chen et al. 2010).

The relationship between Marketization and VC is relatively complex and is characterized by a nonlinear relationship that is U-shaped in nature. This is a possible reason why it receives more attention in RF models. Nonetheless, Marketization does not emerge as one of the most critical variables in either the MLR or RF models.

Moreover, HTIZ signifies the presence of high-tech industrial zones within a city. Evidently, cities equipped with such industrial zones are more appealing to VC investments. However, its significance remains relatively low in comparison to local financial, economic, and innovative development.

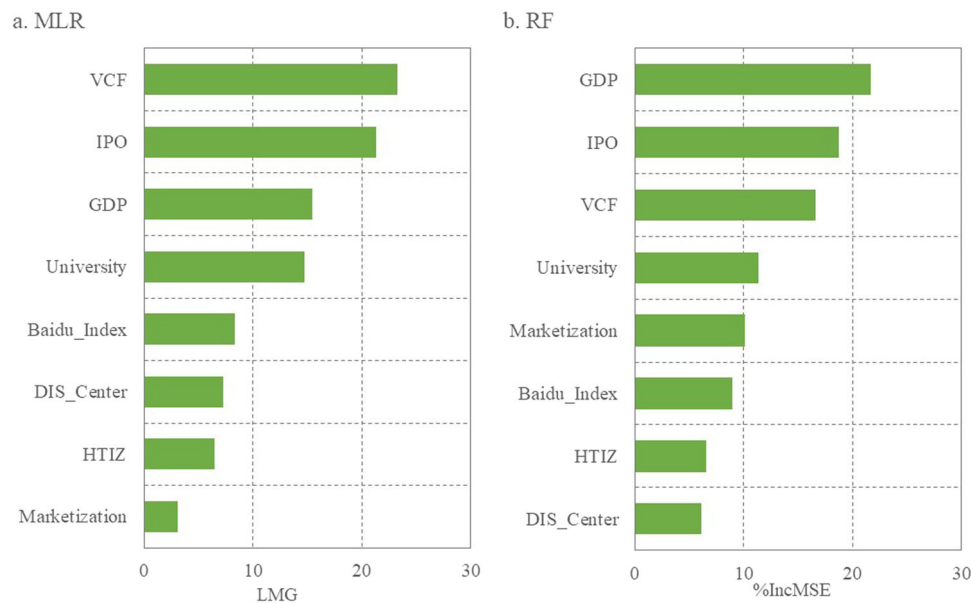


Fig. 4 Comparison of variable importance between the MLR and RF models. Note: **a** LMG metrics of variables based on the MLR model. **b** %IncMSE metrics of variables based on the RF model. The LMG and %IncMSE metrics are normalized to sum to 100%.

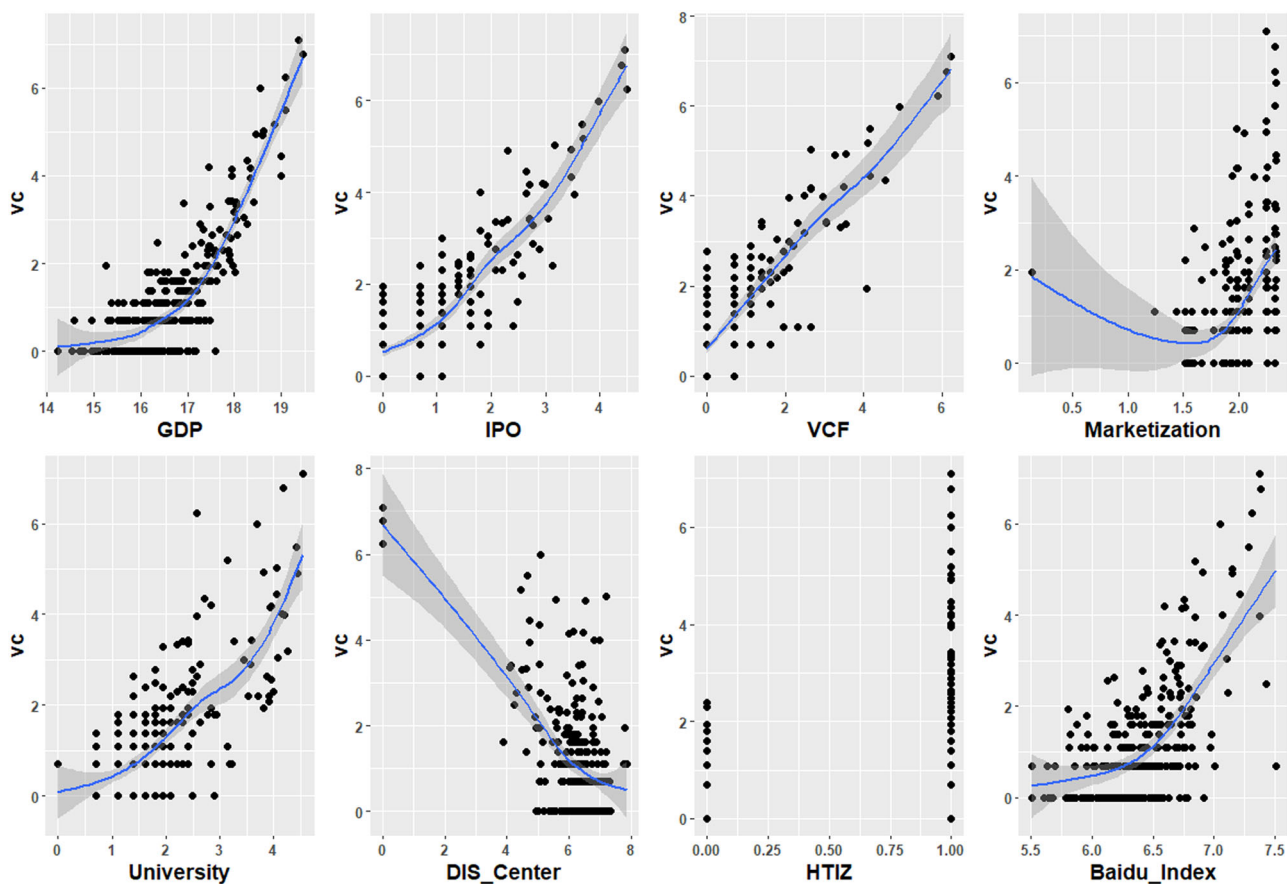


Fig. 5 Bivariate relations between VC and the selected variables. Note: Scatter plots, fitted curves, and confidence intervals illustrating the relationships between the 8 independent variables and VC. The variables, arranged from left to right and top to bottom, are GDP, IPO, VCF, Marketization, University, DIS_Center, HTIZ, and Baidu_Index.



Fig. 6 Spatial heterogeneity of the variable importance based on the MLR and RF models. Note: **a** Variable importance of intercity VC investments in eastern China. **b** Variable importance of intercity VC investments in central China. **c** Variable importance of intercity VC investments in western China. The importance of the variables is the normalization value based on the LMG and %IncMSE metrics in each region.

Spatial heterogeneity. VC investments can be influenced by numerous factors, although the impact of each factor may differ across regions. To explore the spatial heterogeneity of determinants, we conducted a detailed analysis of the importance of variables within the eastern, central, and western regions of China based on the aforementioned models. The results are presented in Fig. 6.

In the eastern region, VCF, IPO, GDP, and University emerge with relatively high importance. The importance of these four indicators significantly surpasses that of others in both the MLR and RF models. This observation can be partially attributed to the spatial distribution of VC investments in China, where the VC industry in the eastern region was established early and has achieved relatively mature development. Notably, in both models, VCF holds the highest importance, closely followed by IPO. This underscores the pivotal role of the local financial market in the eastern region for the advancement of VC investments.

In the central region, the importance of various variables exhibited considerable variation between the MLR and RF models. In the MLR model, VCF, GDP, IPO, and University emerge as relatively important. However, the RF model assigns the highest importance score to GDP, followed by Baidu_Index, University, Marketization,

VCF, and HTIZ. This contrasts with the results observed in the eastern region and is potentially linked to the less mature development of the economy and VC industry in the central region. Regardless, the outcomes of both models emphasize that the economy holds relatively greater importance in influencing VC investments in cities located in the central region.

In the western region, the importance of each variable assumes a distinct order in comparison to the eastern and central regions. According to the MLR model, VCF, IPO, and University exhibit the highest importance, with similar scores, followed by GDP and Baidu_Index. The RF model assigns the highest value to University, which significantly surpasses other independent variables. IPO, Baidu_Index, and HTIZ rank second, third, and fourth, respectively, with comparable importance values. While the distribution of importance for each variable differs notably between the two models, the prominence of University stands out consistently. This underscores the significance of education and innovation potential and their pivotal role in the development of VC investments in the Western region.

Overall, the importance of each variable differs across regions, with the eastern region emphasizing the city's financial market, while the

central and western regions prioritize economic development and innovation capacity, respectively. Considering the national spatial distribution of VC activities, the importance of variables in the eastern region aligns with the national-level analysis results. The overall economic and financial development in the western region is comparatively constrained, making innovation capacity a decisive factor in attracting VC investments in this region.

Conclusion and discussion

In the context of the global flow of VC investments, identifying the pivotal factors at the city level contributes to a deeper understanding of the spatial pattern of VC and offers practical implications for the design of intervention programs. This study employs a dual approach involving MLR and RF to examine the influence of city-level factors—encompassing economy, finance, innovation, location, and policy—on VC investments.

The following findings have been drawn from the research. First, GDP (representing the local economy), VCF and IPO (representing the financial market), University (representing innovation potential), DIS_Center (representing geographical location), and HTIZ and Baidu_Index (representing policy and network popularity) exert significant influences on VC investments. Notably, DIS_Center has a negative impact, while the other variables demonstrate positive effects. Second, upon comparing the results of the MLR and RF models, it becomes evident that a city's economy and finance play the most crucial roles among these attributes, followed by innovation potential. Third, the relative importance of each variable varies across regions. Analysis of the three regions in China reveals that the financial market holds the utmost importance in the economically developed and densely populated eastern region. In the central region, the economic development of cities significantly influences VC investments. In the Western region, the number of universities, which reflects innovation potential and talent cultivation, emerges as the most influential factor.

Analyzing the determinants of VC investments yields valuable policy implications. First, the positive impact of a city's economic and financial development on the inflow of external VC investments underscores their significant roles among the city's attributes. Specifically, the notable influence of VCF and IPO in this study suggests that relevant administrative departments can support the growth of local investment companies and facilitate enterprises' public offerings. Second, innovation emerges as another crucial factor that influences VC investments and endorsing the development of universities contributes to enhancing a city's appeal to external VC investments. This aligns with findings from the distribution of VC investments in the biotechnology industry by Powell et al. (2002). Additionally, the negative effect of distance from VC centers underscores the importance of a city's location and transportation. Research by existing studies (Zheng et al. 2020) also highlights the impact of transportation on VC investments. Therefore, policymakers should consider developing transportation infrastructure. Finally, it is crucial to emphasize spatial heterogeneity among variables; thus, policymakers should tailor decisions according to local conditions. For instance, the eastern region may benefit from enhancing financial markets, while the central and western regions should prioritize local economies and innovation, respectively.

Our analysis contributes in the following ways. This study reveals the influence and relative importance of various city-level factors on intercity VC investments. Compared to existing studies that focus on individual factors (e.g., Dai and Nahata 2016; Zheng et al. 2020) or significance analysis of multiple factors (Cheng et al. 2019; Félix et al. 2023), our study measures the relative importance and spatial heterogeneity of each factor in addition to their significance. This enriches the understanding of intercity VC

flows and provides valuable reference for policymakers and stakeholders to optimize the spatial distribution of investments and promote regional coordinated development. In addition, integrating MLR and RF methods to assess the importance of influencing factors offers a methodological reference for studying the factors impacting financial activities, including VC investments.

This study has limitations and areas for further investigation. The primary constraint lies in the reliance on cross-sectional data due to data acquisition limitations. However, the significance of different variables may evolve over time. The next phase of research involves augmenting the study with enriched time-series data, which should allow for a comprehensive examination and comparison of the evolving influence and importance of various variables from a temporal perspective. While this study concentrates on city characteristics, it is imperative to acknowledge the pivotal role of attributes associated with VC firms and entrepreneurs. Investigating the interaction between these factors and actor attributes represents a valuable avenue for further research. Furthermore, incorporating the impact of unforeseen events, such as financial crises or the COVID-19 pandemic, will contribute to a deeper understanding of risk management and market stability.

Data availability

The venture capital data analyzed in this study is not publicly available to maintain the agreement with the CVSource database that provided the data for research. Detailed information regarding the procedure for requesting access to the data, supporting the findings of this study, is available from the corresponding author upon reasonable request.

Received: 19 March 2024; Accepted: 30 August 2024;

Published online: 14 September 2024

References

- Aizenman J, Kendall J (2012) The internationalization of venture capital. *J Econ Stud* 39(5):488–511. <https://doi.org/10.1108/01443581211259446>
- Audretsch DB, Acs ZJ (1994) New-firm startups, technology, and macroeconomic fluctuations. *Small Bus Econ* 6(6):439–449. <https://doi.org/10.1007/BF01064858>
- Bai S, Zhao Y (2021) Startup investment decision support: application of venture capital scorecards using machine learning approaches. *Systems* 9(3):55. <https://doi.org/10.3390/systems9030055>
- Belgiu M, Drăguț L (2016) Random forest in remote sensing: a review of applications and future directions. *ISPRS J Photogramm Remote Sens* 114:24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Bellucci A, Borisov A, Gucciardi G, Zazzaro A (2023) The reallocation effects of COVID-19: evidence from venture capital investments around the world. *J Bank Financ* 147:106443. <https://doi.org/10.1016/j.jbankfin.2022.106443>
- Bi J (2012) A review of statistical methods for determination of relative importance of correlated predictors and identification of drivers of consumer liking. *J Sens Stud* 27(2):87–101. <https://doi.org/10.1111/j.1745-459X.2012.00370.x>
- Blank TH, Carmeli A (2021) Does founding team composition influence external investment? The role of founding team prior experience and founder CEO. *J Technol Transf* 46(6):1869–1888. <https://doi.org/10.1007/s10961-020-09832-3>
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
- Bringmann K, Vanoutrive T, Verhelsel A (2018) Venture capital: the effect of local and global social ties on firm performance. *Pap Region Sci* 97(3):737–756. <https://doi.org/10.1111/pirs.12261>
- Budescu DV (1993) Dominance analysis: a new approach to the problem of relative importance of predictors in multiple regression. *Psychol Bull* 114(3):542–551. <https://doi.org/10.1037/0033-2909.114.3.542>
- Chen H, Gompers P, Kovner A, Lerner J (2010) Buy local? The geography of venture capital. *J Urban Econ* 67(1):90–102. <https://doi.org/10.1016/j.jue.2009.09.013>
- Cheng C, Hua Y, Tan D (2019) Spatial dynamics and determinants of sustainable finance: evidence from venture capital investment in China. *J Clean Prod* 232:1148–1157. <https://doi.org/10.1016/j.jclepro.2019.05.360>

- Colombo M, D'Adda D, Quas A (2019) The geography of venture capital and entrepreneurial ventures' demand for external equity. *Res Policy* 48(5):1150–1170. <https://doi.org/10.1016/j.respol.2018.12.004>
- Cumming D, Dai N (2010) Local bias in venture capital investments. *J Empir Financ* 17(3):362–380. <https://doi.org/10.1016/j.jempfin.2009.11.001>
- Dai N, Nahata R (2016) Cultural differences and cross-border venture capital syndication. *J Int Bus Stud* 47(2):140–169. <https://doi.org/10.1057/jibs.2015.32>
- Dimov D, Murray G (2008) Determinants of the incidence and scale of seed capital investments by venture capital firms. *Small Bus Econ* 30(2):127–152. <https://doi.org/10.1007/s11187-006-9008-z>
- Du D, Wang J, Li J, Huang J (2024) Evolution of China's intercity venture capital network: preferential attachment vs. path dependence. *Cities* 148. <https://doi.org/10.1016/j.cities.2024.104874>
- Duan L, Sun W, Zheng S (2020) Transportation network and venture capital mobility: an analysis of air travel and high-speed rail in China. *J Transport Geogr* 88. <https://doi.org/10.1016/j.jtrangeo.2020.102852>
- Falik Y, Lahti T, Keinonen H (2016) Does startup experience matter? Venture capital selection criteria among Israeli entrepreneurs. *Ventur Cap* 18(2):149–174. <https://doi.org/10.1080/13691066.2016.1164109>
- Fang JW (2018) An analysis of the differentiation rules and influencing factors of venture capital in Beijing-Tianjin-Hebei urban agglomeration. *J Geogr Sci* 28(4):514–528. <https://doi.org/10.1007/s11442-018-1487-8>
- Feldman BE (2005) Relative importance and value. *SSRN Electron J*. <https://doi.org/10.2139/ssrn.2255827>
- Félix EGS, Nunes JC, Pires CP (2023) The impact of concentration among venture capitalists: revisiting the determinants of venture capital. *Ventur Cap* 25(4):457–486. <https://doi.org/10.1080/13691066.2022.2147876>
- Florida R, Smith Jr DF (1993) Venture capital formation, investment, and regional industrialization. *Ann Assoc Am Geogr* 83(3):434–451. <https://doi.org/10.1111/j.1467-8306.1993.tb01944.x>
- Gompers PA, Gornall W, Kaplan SN, Strebulaev IA (2020) How do venture capitalists make decisions? *J Financ Econ* 135(1):169–190. <https://doi.org/10.1016/j.jfineco.2019.06.011>
- Grömping U (2006) Relative importance for linear regression in R: The Package relaimpo. *J Stat Software*, 17(1). <https://doi.org/10.18637/jss.v017.i01>
- Grömping U (2015) Variable importance in regression models. *WIREs Comput Stat* 7(2):137–152. <https://doi.org/10.1002/wics.1346>
- Guzella M, Buchbinder F, Santana V (2024) Venture capital investment in Latin America: the role of experience, distances, and network features. *Emerg Markets Rev* 60. <https://doi.org/10.1016/j.ememar.2024.101145>
- Hirukawa M, Ueda M (2011) Venture capital and innovation: which is first? *Pac Econ Rev* 16(4):421–465. <https://doi.org/10.1111/j.1468-0106.2011.00557.x>
- Hochberg YV, Ljungqvist A, Lu Y (2010) Networking as a barrier to entry and the competitive supply of venture capital. *J Financ* 65(3):829–859. <https://doi.org/10.1111/j.1540-6261.2010.01554.x>
- Hsu DH (2004) What do entrepreneurs pay for venture capital affiliation? *J Financ* 59(4):1805–1844. <https://doi.org/10.1111/j.1540-6261.2004.00680.x>
- Kruskal W (1987) Relative importance by averaging over orderings. *Am Stat* 41(1):6–10. <https://doi.org/10.1080/00031305.1987.10475432>
- Lachniet MS, Patterson WP (2006) Use of correlation and stepwise regression to evaluate physical controls on the stable isotope values of Panamanian rain and surface waters. *J Hydrol* 324(1–4):115–140. <https://doi.org/10.1016/j.jhydrol.2005.09.018>
- Li S, Yang H (2022) Research on the relationship between venture capitalists' trust in the entrepreneur and their investment behaviors. *Entrep Res J* 12(2):161–184. <https://doi.org/10.1515/erj-2020-0151>
- Li Y, Zahra SA (2012) Formal institutions, culture, and venture capital activity: a cross-country analysis. *J Bus Ventur* 27(1):95–111. <https://doi.org/10.1016/j.jbusvent.2010.06.003>
- Lindeman RH, Merenda PF, Gold RZ (1980) Introduction to bivariate and multivariate analysis (Vol. 4). Glenview, IL: Scott, Foresman
- Liu L, Jiang H, Zhang Y (2023) The impact of venture capital on Chinese SMEs' sustainable development: a focus on early-stage and professional characteristics. *Humanit Soc Sci Commun* 10(1):381. <https://doi.org/10.1057/s41599-023-01893-7>
- Liu M, Hu S, Ge Y, Heuvelink GBM, Ren Z, Huang X (2021) Using multiple linear regression and random forests to identify spatial poverty determinants in rural China. *Spat Stat* 42:100461. <https://doi.org/10.1016/j.spasta.2020.100461>
- Macmillan IC, Siegel R, Narasimha PNS (1985) Criteria used by venture capitalists to evaluate new venture proposals. *J Bus Ventur* 1(1):119–128. [https://doi.org/10.1016/0883-9026\(85\)90011-4](https://doi.org/10.1016/0883-9026(85)90011-4)
- Martin R, Berndt C, Klagge B, Sunley P (2005) Spatial proximity effects and regional equity gaps in the venture capital market: evidence from Germany and the United Kingdom. *Environ Plan A-Econ Space* 37(7):1207–1231. <https://doi.org/10.1068/a3714>
- Mason C, Pierrakis Y (2013) Venture capital, the regions and public policy: the United Kingdom since the post-2000 technology crash. *Region Stud* 47(7):1156–1171. <https://doi.org/10.1080/00343404.2011.588203>
- Ning Y, Wang W, Yu B (2015) The driving forces of venture capital investments. *Small Bus Econ* 44(2):315–344. <https://doi.org/10.1007/s11187-014-9591-3>
- Pan F, Hall S, Zhang H (2020) The spatial dynamics of financial activities in Beijing: agglomeration economies and urban planning. *Urban Geogr* 41(6):849–864. <https://doi.org/10.1080/02723638.2019.1700071>
- Pan F, Zhao SXB, Wójcik D (2016) The rise of venture capital centres in China: a spatial and network analysis. *Geoforum* 75:148–158. <https://doi.org/10.1016/j.geoforum.2016.07.013>
- Pintado TR, De lema DGP, Van auken H (2007) Venture capital in Spain by stage of development. *J Small Bus Manag* 45(1):68–88. <https://doi.org/10.1111/j.1540-627X.2007.00199.x>
- Powell WW, Koput KW, Bowie JJ, Smith-Doerr L (2002) The spatial clustering of science and capital: accounting for biotech firm-venture capital relationships. *Region Stud* 36(3):291–305. <https://doi.org/10.1080/00343400220122089>
- Ressin M (2022) Start-ups as drivers of economic growth. *Res Econ* 76(4):345–354. <https://doi.org/10.1016/j.rie.2022.08.003>
- Sannabe A (2022) How to improve SME performance using iterative random forest in the empirical analysis of institutional complementarity. *Humanit Soc Sci Commun* 9(1):114. <https://doi.org/10.1057/s41599-022-01123-6>
- Wang L, Wang S (2012) Economic freedom and cross-border venture capital performance. *J Empir Financ* 19(1):26–50. <https://doi.org/10.1016/j.jempfin.2011.10.002>
- Wray F (2012) Rethinking the venture capital industry: relational geographies and impacts of venture capitalists in two UK regions. *J Econ Geogr* 12(1):297–319. <https://doi.org/10.1093/jeg/lbq054>
- Wu K, Wang Y, Zhang H, Liu Y, Ye Y, Yue X (2022) The pattern, evolution, and mechanism of venture capital flows in the Guangdong-Hong Kong-Macao Greater Bay Area, China. *J Geogr Sci* 32(10):2085–2104. <https://doi.org/10.1007/s11442-022-2038-x>
- Yang H, Du D, Wang J, Wang X, Zhang F (2023) Reshaping China's urban networks and their determinants: high-speed rail vs. air networks. *Transp Policy* 143:83–92. <https://doi.org/10.1016/j.tranpol.2023.09.007>
- Yang Y, Wang X, Tang L (2024) Institutional learning, cultural differences and the motivation of syndication among cross-border venture capital firms in China. *Eur J Int Manag* 22(1):104–123. <https://doi.org/10.1504/EJIM.2024.135216>
- Zacharakis AL, McMullen JS, Shepherd DA (2007) Venture capitalists' decision policies across three countries: an institutional theory perspective. *J Int Bus Stud* 38(5):691–708. <https://doi.org/10.1057/palgrave.jibs.8400291>
- Zhang R, Tian Z, McCarthy KJ, Wang X, Zhang K (2023) Application of machine learning techniques to predict entrepreneurial firm valuation. *J Forecast* 42(2):402–417. <https://doi.org/10.1002/for.2912>
- Zheng S, Duan L, Sun W (2020) Global air network and cross-border venture capital mobility. *Habitat Int* 106:102105. <https://doi.org/10.1016/j.habitatint.2019.102105>
- Zhou Y, Park S, Zhang JZ, Ferreira JJ (2023) How do innovative internet tech startups attract venture capital financing? *J Manag Organ*, 1–22. <https://doi.org/10.1017/jmo.2023.39>

Acknowledgements

This work is supported by National Natural Science Foundation of China (Grant No. 42225106 and No. 42401206) and China Postdoctoral Science Foundation funded project (Grant No. 2022M723120).

Author contributions

Jiaoe Wang and Delin Du conceptualized the study, and designed, structured, and drafted the paper. Delin Du and Jianjun Li collected data and conducted analyses. Delin Du, Jiaoe Wang, and Jianjun Li contributed to editing the manuscript.

Competing interests

The authors declare no competing interests.

Ethical approval

This study is not related to human participants performed by any of the authors.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Correspondence and requests for materials should be addressed to Jiaoe Wang.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com