Journal of Finance - Marketing Research; Special Issue 1; 2026
e-ISSN: 3030-430X
DOI: https://doi.org/10.52932/jfmr.v17i1enes

Journal of Finance - Marketing Research
http://jfm.edu.vn

REGIONAL BRAIN DRAIN IN VIETNAM: AN ANALYSIS OF THE IMPACT OF INCOME AND REGIONAL ECONOMIC CONDITIONS ON THE MIGRATION BEHAVIOR OF HIGHLY EDUCATED PERSONNEL

Le Thi Thuy Hang[1*], Nguyen Tuan Duy[1], Ta Minh Cong[1], Nguyen Thuy Duong[1], Tran Phu Dong[1], Pham Ngoc Phuong Nghi[1], Nguyen Phuoc Quy Son[1]

[1]University of Finance - Marketing, Vietnam

ARTICLE INFO

DOI:
10.52932/jfmr.v17i1enes.1342

Received:
April 10, 2026
Accepted:
July 31, 2026
Published:
August 08, 2026

Keywords:
Brain drain,
Highly educated personnel,
Regional economic conditions,
Machine Learning,
SVR, SHAP

Article classification:
Research paper

JEL codes:
J61, R23, O15, C45

ABSTRACT

Purpose – This study aims to analyze and quantify the multi-dimensional impact of income and regional economic conditions on the migration behavior of highly educated personnel (Associate Professors, Professors) across key economic regions in Vietnam, amidst the increasing phenomenon of "regional brain drain".

Design/methodology/approach – The study utilizes panel data from 6 key economic regions over the period 2010-2024. Drawing on human capital and regional economics theories, the paper employs a Machine Learning ecosystem including Support Vector Regression (SVR), SHAP explanatory method, and K-Means clustering to handle non-linear relationships and isolate migration drivers.

Findings – The empirical results reveal an "Income Paradox" among intellectuals: upon reaching financial security, the pulling power of income disparity diminishes. Instead, Urbanization and the Educational ecosystem emerge as the dominant drivers, creating breakthrough impacts in attracting talent. Furthermore, expanding residential areas without synchronized amenities acts as a barrier to migration.

Practical implications – Instead of relying solely on short-term financial incentives, policymakers and local authorities must shift their focus to upgrading "soft" social infrastructure (education, healthcare) and building local "talent incubators" to retain intellectuals and promote sustainable brain circulation.

*Corresponding author:
Email: ltt.hang@ufm.edu.vn

Originality/value – This study provides pioneering empirical evidence by applying Machine Learning (SVR and SHAP) to decode the non-linear impact "thresholds" in the migration behavior of Vietnamese intellectuals, successfully overcoming the limitations of traditional linear econometric models.

1. Introduction

Highly educated personnel, particularly highly educated personnel, are consistently regarded as the key factor determining national competitiveness and long-term innovation. For a country in the process of industrialization and modernization like Vietnam, the rational allocation of highly educated personnel across regions plays a crucial role in narrowing the development gap. However, over the past decade, Vietnam has faced an escalating phenomenon of "regional brain drain," creating severe "knowledge sinks" in peripheral areas such as the Central Highlands and the Mekong River Delta (World Bank, 2019).

Specifically, the post-COVID-19 period has acted as a "catalyst" (World Bank, 2021) that altered market structures, accelerating the wave of highly educated personnel migration converging intensely on core growth poles such as Hanoi, Ho Chi Minh City, and Da Nang. In reality, the migration decision of intellectuals is not merely a geographical relocation but essentially an investment strategy aimed at optimizing human capital. While major metropolitan areas create dynamic "knowledge ecosystems" with strong agglomeration effects (Glaeser, 2011), underdeveloped regions lack specialized infrastructure systems and high-quality social services. This inadvertently creates a downward spiral: low income and outdated infrastructure push talent away, and the very shortage of talent prevents regional labor productivity from breaking through.

Although the topics of migration and regional inequality have received profound attention (Docquier & Rapoport, 2012), there currently exists a significant gap in the methodological approaches and integrated theoretical models to explain the migration behavior specifically of the highly educated personnel (university degree and above) in Vietnam. Domestic studies primarily lean toward describing macro trends or analyzing general labor migration without isolating intellectuals as an independent subject. Furthermore, most local "red carpet" policies (talent attraction policies) remain superficial, lacking quantitative evidence measuring the simultaneous relationship between individual income variables and regional macroeconomic conditions.

Stemming from these practical demands and research gaps, this paper is conducted to identify and quantify the multi-dimensional impact of the income factor alongside specific regional economic conditions on the migration behavior of highly educated personnel. Particularly, in the context where the nation is accelerating digital transformation and developing a knowledge economy, solving the problem of elite human resource allocation becomes even more urgent(World Bank, 2019).

The novel contribution of this study is the pioneering application of the Support Vector Regression (SVR) Machine Learning algorithm combined with the SHAP (Shapley Additive Explanations) method. This analytical framework not only overcomes the limitations

of traditional linear econometric models but also allows for the decoding of non-linear impact mechanisms and the limit "thresholds" of each macro factor. By providing highly reliable empirical evidence, the study expects to establish a solid scientific basis to help policymakers build more substantive talent attraction strategies, thereby promoting sustainable brain circulation and harmonious economic development across localities (Athey, 2018).

2. Theoretical background and literature review

2.1. Theoretical background

According to the *Human Capital Theory* (Becker, *1966*), migration is not merely a geographical relocation but a form of investment in human capital aimed at maximizing income. Workers decide to migrate when the total expected net benefits (especially the incremental income margin) outweigh the barriers and costs (financial, opportunity, and psychological). When applied to the highly educated personnel, this theory explains the "brain drain" phenomenon through two core mechanisms:

Positive selection: Highly educated individuals are more prone to migrate because the absolute income differential between the destination and the origin is significantly more attractive to them than to unskilled labor.

Positive sorting: Intellectuals exhibit a strong tendency to converge on major economic hubs, where the ecosystem is willing to pay the highest "rewards to skill". Furthermore, a high level of education acts as a filter that helps workers minimize mobility costs and barriers (due to their superior ability to gather information and adapt quickly). In summary, the brain drain towards metropolitan areas is an inevitable consequence as highly educated personnel actively seek destinations that yield the highest rate of return on their educational investments Becker *(1966*).

Expected Income Hypothesis (Mochiah, 2014) diverging from traditional models, Harris and Todaro posit that the migration decision is driven by "expected income"-a variable calculated by multiplying the wage at the destination by the probability of finding employment. When applied to the brain drain phenomenon in Vietnam, this model elucidates the immense magnetic pull of major economic centers (e.g., Ho Chi Minh City, Hanoi). Because highly educated individuals possess a substantial competitive advantage, their probability of securing formal employment is almost absolute. This dramatically amplifies their expected income, creating a strong pulling force that drains human resources from rural and mountainous areas. In addition, modern extended studies introduce two critical variables governing the migration optimization problem of intellectuals: (i) Spatial and living costs: The escalation of housing prices and living costs in urban areas acts as a "natural deterrent," potentially neutralizing the benefits of expected income if they rise excessively; (ii) Social networks (Tzeng, 1995) (friends, professionals) at the destination help reduce information asymmetry and psychological costs, further increasing the probability of accessing high-quality employment.

Regional Economics and Uneven Development Theory (Myrdal, 1957), the theory of regional economics and uneven development asserts that economic growth does not occur uniformly but tends to concentrate in areas with initial advantages. According to Gunnar Myrdal, developed regions continuously attract capital, labor, and resources from less developed areas, exacerbating regional disparities. He explains this through two mechanisms: the backwash effect, where the center aggressively drains resources from the periphery, and the spread effect, where the center's growth generates economic opportunities for surrounding areas. In reality, the backwash effect usually

dominates, leading to widening regional inequality. Inheriting this concept, (Krugman, 1991) developed the New Economic Geography with the Core-Periphery model. According to this model, firms tend to cluster in central areas to exploit economies of scale and reduce transport costs, thereby forming a developed core and a dependent periphery. This serves as the direct theoretical foundation explaining the uneven development trend and the brain drain towards major growth poles in Vietnam.

Quality of Life and Circular Migration Theory (Paas & Halapuu, 2012). Associated with the research of Richard Florida, this theory suggests that highly educated personnel migrate not only for employment or income but also for the quality of life. In the modern context, migration is no longer unidirectional but can be circular; workers leave to study or work and may return when living conditions become more suitable. Florida explains a locality's attractiveness through the 3T model: Technology, Talent, and Tolerance. Technology creates a modern working environment, talent reflects the quality of human resources, and tolerance indicates the society's openness to diversity. The combination of these three factors shapes the quality of life and determines the ability to attract and retain high-quality labor. Therefore, when a locality ensures a high standard of living, the migration of intellectuals can become circular, thereby mitigating brain drain and fostering sustainable development.

2.2. Literature review

Studies on the impact of income and regional disparities (Floerkemeier et al., 2021), numerous international studies demonstrate that income and developmental disparities between regions are crucial drivers of the brain drain phenomenon. Docquier and Rapoport argue that highly skilled workers frequently migrate when income and employment opportunities elsewhere significantly surpass those in their origin. Simultaneously, domestic income levels also affect migration decisions by determining the ability to afford relocation costs. In Southeast Asia, research by Olinyk and Vasa (2021) shows that higher-income countries typically attract labor from economies with inferior conditions, clearly reflecting a "push-pull" mechanism. A report by the International Monetary Fund (IMF) also indicates that regions with superior income and employment prospects draw labor from underdeveloped areas; however, if this persists, it can exacerbate regional imbalances and deplete human capital in the sending regions.

Studies on infrastructure, education, and public services. Besides income and employment, recent literature highlights that infrastructure, education, and public services robustly influence brain drain. A study by Vega-Muñoz (2024) across 178 countries shows an inverse relationship between the quality of public services (such as healthcare, education, and social security) and the out-migration trend of highly educated personnel the better the public services, the higher the retention rate. An OECD report (2023) also emphasizes that disparities in transport infrastructure, digital connectivity, and access to public services are primary reasons compelling skilled labor to abandon less developed regions for major metropolitan areas. Overall, these studies demonstrate that brain drain is driven not solely by economic factors but also significantly by the quality of infrastructure, education, and public services.

Studies on intellectual migration and brain drain in Vietnam. Recent empirical studies in Vietnam have focused on quantifying the dynamics propelling the highly educated personnel from the periphery to major growth poles. Two prominent approaches stand out:

The Macro-condition approach: Argues that high income is merely a "necessary condition." The "sufficient condition" to retain talent is the

regional economic ecosystem (concentration of FDI, hi-tech zones, healthcare, and educational infrastructure). Peripheral regions lose their brain trust predominantly due to the absence of a modern working environment (Nha & Tinh, 2023).

The Income approach: Indicates that individuals with a university degree or higher exhibit extreme "sensitivity" to inter-regional income gaps. Expected income acts as the primary pulling force, transforming migration into an inevitable investment to maximize the rate of return on education (Tran & Nguyen, 2021).

2.3. Research gaps

Through a comprehensive review of domestic and international literature, this paper identifies three major gaps regarding the migration of the highly educated personnel:

Lack of micro-level quantitative models: Current studies in Vietnam primarily favor descriptive statistics or utilize provincial-level macro data (panel data). There remains a scarcity of econometric models measuring the probability of migration decisions (benefit trade-offs) at the individual intellectual level.

Fragmentation of variables: Previous research often isolates migration drivers, either heavily emphasizing income disparities or focusing exclusively on infrastructure and innovation ecosystems. There is currently no integrated model that simultaneously incorporates both dimensions (Personal Income and Regional Economic Conditions) to determine the "necessary and sufficient conditions" shaping the brain drain flow.

Limits in spatial representation: Existing micro-surveys generally concentrate on localized migration routes (e.g., exclusively toward Hanoi). There is a lack of comprehensive studies fully dissecting intellectual out-migration from disadvantaged regions (the Mekong River Delta, Central Highlands, Northern Midlands and Mountains) toward the growth poles.

3. Research model and methodology

3.1. Research data

This study employs panel data collected at the macroeconomic regional level from six key economic regions in Vietnam: the Red River Delta, the Northern Midlands and Mountains, the North Central and Central Coast, the Central Highlands, the Southeast, and the Mekong River Delta. The data covers the period from 2010 to 2024.

The data processing in this study is conducted through three main steps:

Step 1: Data collection: Data is gathered and synthesized from official and highly reputable reporting sources, including the General Statistics Office of Vietnam (GSO – Annual Statistical Yearbooks), the Ministry of Planning and Investment (MPI), the Ministry of Education and Training (MOET), the Ministry of Labour - Invalids and Social Affairs (MOLISA), and the United Nations Development Programme (UNDP) in Vietnam over the 2010–2024 period.

Step 2: Data processing: From the dataset collected via reliable sources, the research team utilized Microsoft Excel to calculate the variables. Subsequently, data cleaning was performed, which included applying logarithmic transformations to variables with excessively large values to ensure data consistency and standardization.

Step 3: Model execution: The processed data was then imported into STATA 17 software to analyze and perform necessary tests, aiming to determine the impact of independent and control variables on the dependent variable, thereby drawing conclusions for the study.

3.2. Description of variable groups impacting highly educated personnel

Economic variable group (Income)

The impact of income on the migration decisions of highly educated personnel is explained through the Human Capital Theory (Sjaastad, 1962). Highly skilled individuals have invested substantial resources (time, effort, finance) to cultivate their knowledge and skills. When their "human capital" is undervalued in rural areas with inadequate income, migration becomes a strategic decision. Confident in their capabilities, they believe that economic hubs will offer job positions with commensurate salaries, sufficient to compensate for all their initial educational investments and costs.

Furthermore, the migration decisions of this intellectual group are not solely based on current static wages but are guided by "expected income" (Harris & Todaro, 1970). Even when facing mobility cost barriers, high living expenses in urban areas, or fierce competition risks in economic centers, they still choose to migrate. With high income levels and reasonable employment opportunities, the long-term "expected income" generated is significantly greater than a low but completely secure wage in rural areas.

Hypothesis H1: Income (INC) has a positive but non-linear impact (threshold effect) on the highly educated personnel. Upon reaching a certain threshold of financial security, the marginal importance of income diminishes, giving way to quality-of-life factors.

Scale & development opportunity variable group (Urban population)

Urbanization in Vietnam has undergone profound migration transformations over the past three decades. Internal migration is not merely a solution for improving livelihoods but also a means to access education, healthcare, and an environment for advanced professional development. The Regional Economics and Uneven Development theories by Krugman (1991) and Myrdal (1957) explain the magnetism of destinations for the intellectual workforce.

According to the Core-Periphery Model, the larger the urban population, the stronger the "backwash effect," forcefully attracting highly educated labor from the periphery. This pull is generated by three core mechanisms: a knowledge spillover environment that stimulates innovation; labor specialization that optimizes the match between skills and jobs; and the development of a high-end service ecosystem meeting the rigorous standards of intellectuals. Overall, "core" urban regions create a comprehensive ecosystem, serving as an ideal environment for highly educated personnel to maximize their capabilities and advancement opportunities.

Hypothesis H2: Urbanization (URB) acts as the primary driver, exerting the strongest influence and creating a breakthrough impact in attracting highly educated personnel.

Living environment & infrastructure system variable group (Clean water, Electricity, Housing area, Health, Education)

Besides economic drivers like income disparity or job opportunities, the migration decisions of highly educated personnel are deeply governed by non-economic factors. "Living Environment and Infrastructure" are considered the core drivers for attracting highly skilled personnel (Florida, 2019). The fundamental difference in migration behavior is that while unskilled labor is heavily driven by wages for survival needs, the intellectual class sets stringent requirements for a comprehensive living environment to support long-term development.

The migration decision of intellectuals relies not only on job opportunities but also on the overall quality of the living environment.

First, "necessary conditions" such as physical infrastructure (electricity, clean water, and standard housing) form the foundational base enabling them to maintain a high quality of life and work effectively, especially amidst deep technological integration. Additionally, "sufficient conditions" pertaining to social infrastructure, such as healthcare and educational systems, play a decisive role when they consider long-term plans for themselves and their families. The harmonious combination of these two elements creates the magnetism of economic centers, making them attractive destinations capable of retaining highly educated personnel in the long run.

First, regarding the educational ecosystem (EDU): For intellectuals (Associate Professors, Professors), the educational environment is not merely a basic public service but a critical extension of their professional and family life. A locality with a dense concentration of high-quality educational institutions provides a dual benefit: an excellent research and teaching environment for their own careers, and superior developmental opportunities for their children. Consequently, education transcends basic infrastructure to become a primary pulling force.

Hypothesis H3: The Educational ecosystem (EDU) acts as the core "soft infrastructure," possessing a macro-level importance equivalent to Urbanization in generating talent attraction.

Second, regarding Residential area (ARE): Housing is an essential need, but for intellectuals, the absolute size of the living space is not always proportional to their satisfaction. In the context of spatial economics, localities with very high average housing areas per capita are often peripheral or suburban areas characterized by sprawling land funds but severe deficits in synchronized urban amenities (e.g., entertainment, academic networking spaces). Therefore, merely expanding residential space without integrating it into a comprehensive urban ecosystem can inadvertently create a structural barrier to attracting talent.

Hypothesis H4: Residential area (ARE) exerts a complex and multi-dimensional impact. Possessing large land funds without synchronized urban amenities can become a barrier rather than an advantage in highly educated personnel.

Third, regarding Basic physical and health infrastructure (WAT, ELE, HEA): In the modern development phase of Vietnam's key economic regions, basic utilities such as electricity and clean water coverage have reached near-absolute saturation levels. According to Herzberg's two-factor theory, these are "hygiene factors", their absence causes severe dissatisfaction, but their presence no longer creates a competitive advantage or proactive motivation for intellectuals to migrate. Similarly, basic healthcare density (doctors per 1000 people) ensures fundamental safety but does not serve as a breakthrough catalyst for intellectual relocation compared to specialized academic or urban environments.

Hypothesis H5: Basic infrastructure factors (WAT, ELE, HEA) serve as "foundational conditions" (where coverage has reached saturation); therefore, their marginal impact magnitude is weaker and does not create as substantial a difference as the Urbanization and Education groups.

3.3. Variables in the model and proposed model

Based on the theoretical foundation established in Section 2 and the research hypotheses in Section 3.1, the quantitative model of this study utilizes 1 dependent variable and 7 independent variables, as summarized in Table 1. The raw data, upon collection, was processed and standardized using MS Excel before being imported into the econometric software.

Table 1. Summary of variables in the research model

Variable Type	Symbol	Observed Variable Name	Unit	Data Source
Dependent Variable	SCH	Number of highly educated migrants in regions	Persons	GSO / MOLISA
Independent Variable	INC	Income per capita	Thousand VND	GSO
	URB	Average urban population by region	Thousands of persons	GSO
	WAT	Percentage of population with access to hygienic water sources	%	GSO / UNDP
	ELE	Percentage of households using electricity for daily life	%	GSO
	ARE	Average housing floor area per capita	person	GSO
	EDU	Number of educational institutions per 1000 people	Institutions	MOET / GSO
	HEA	Number of doctors per 1000 people	Persons	GSO

Furthermore, Appendix 1 outlines the expected signs and the nature of the impact of these variables in the research model *(see Appendix 1 online).*

The general functional equation is expressed as follows:

$$SCH_{it} = f(INC_{it}, URB_{it}, WAT_{it}, ELE_{it}, ARE_{it}, EDU_{it}, HEA_{it}) + \varepsilon_{it}$$

Wherein, function f is not restricted to a linear form but is estimated through the following algorithms:

Support Vector Regression (SVR): SVR with a Radial Basis Function (RBF) kernel is utilized to map data into a high-dimensional space. This algorithm is capable of perfectly capturing the non-linear relationships and limit "thresholds" inherent in complex migration behaviors. Notably, SVR operates with extreme stability and effectively prevents overfitting on the small panel dataset ($N = 90$) of this study.

To address the reviewer's request for a more detailed SVR training procedure, the model was specified using a Radial Basis Function (RBF) kernel, which is highly optimal for capturing the complex, non-linear nature of migration data. The hyperparameters were strategically tuned: the regularization parameter was set to C = 1.0 to prioritize model stability and mitigate overfitting within the small panel dataset, while the epsilon-tube width was set to 0.4 to filter out minor data noise and focus on major socio-economic trends. Additionally, the gamma parameter was configured to 'scale' to automatically adapt to the differing measurement scales of variables such as income and infrastructure. Model validation was rigorously conducted through a Train-Test split and Learning Curve analysis, which confirmed spatial convergence without falling into a 'High Variance' overfitting state. Finally, model performance was evaluated, yielding a Root Mean Square Error (RMSE) of 0.6338 and an R-squared (R^2) of 0.2550, demonstrating a highly reliable predictive baseline for the subsequent SHAP marginal impact analysis.

SHAP (Shapley Additive Explanations) Method: To resolve the "black-box" nature of SVR, the study integrates the SHAP method based on cooperative game theory. This method helps fairly allocate the contribution value of each variable to the model's prediction output,

thereby clearly isolating the importance and the direction of impact (positive/negative) of each infrastructural and economic factor on talent attraction.

K-Means Clustering: Based on the impact features extracted from SHAP, the K-Means clustering algorithm is employed to group the six economic regions into clusters with similar talent attraction strategies, assisting in proposing practical policy implications.

4. Research results and discussion

4.1. The current situation of regional brain drains in Vietnam

During the 2010-2024 period, the migration landscape of the highly educated personnel (university level and above) in Vietnam exhibited severe spatial polarization. As illustrated in Appendix 2 *(see Appendix 2 online)*, the North Central and Central Coast region suffered the most severe talent deficit (with over 1,200 individuals leaving), which contrasts sharply with the absolute "retention" capacity of the Southeast region.

Regarding the temporal trend, Appendix 3 *(see Appendix 3 online)* shows that the brain drain flow fluctuated strongly with a structural break occurring in the 2014-2017 period, reflecting the "magnetic pull" that drained personnel from major economic centers recovering after the crisis.

Recognizing this leakage, the macro-level challenge is to enhance the capacity for talent accumulation. Therefore, in the subsequent quantitative models, the dependent variable (SCH) is standardized to represent the "Capacity to attract experts" to the region.

4.2. Descriptive statistics and correlation analysis

Descriptive statistics of the dataset: Appendix 4 *(see Appendix 4 online)* presents the descriptive statistics of the variables in the model. The study utilizes a panel dataset collected from 6 key economic regions over a 15-year period (2010-2024). The total number of observations in the model is 90 ($N = 90$). The statistical results show that all variables have a full 90 observations, with no missing data, ensuring high reliability and representativeness for the subsequent econometric estimates.

Correlation analysis and multicollinearity: As shown in Appendix 5 (*see Appendix 5 online*), The Pearson correlation matrix issues a strong multicollinearity warning for the variable pairs INC – ARE ($r = 0.8210$) and WAT – ELE ($r = 0.8504$). This information overlap confirms that the traditional OLS linear model would distort the true nature of the regression coefficients, creating a mandatory basis for the study to switch to applying a non-linear Machine Learning method (SVR).

4.3. SVR model estimation results and shap analysis

The study applies the Support Vector Regression (SVR) model with an RBF kernel combined with the SHAP explanatory method to handle non-linear relationships, optimize small sample data, and "open the black box" of the algorithm. Table 2 details the parameter configuration of this SVR model.

Table 2. Parameter configuration of the Support Vector Regression (SVR) model

Parameters	Value
Kernel	rbf
C (Regularization)	1.0
ϵ (Epsilon)	0.4
Gamma	scale

Model performance evaluation: With the fine-tuned configuration ($C = 1.0, \epsilon = 0.4$), the model achieves an $R^2 = 0.2550$ and a low RMSE (0.6338). Notably, the learning curve analysis presented in Appendix 6 (*see Appendix*

6 online) confirms that the model does not fall into an overfitting state. The Generalization Gap between the training and validation curves is maintained stably, ensuring good overall predictive capability.

Isolating attraction drivers using SHAP: The SHAP summary plot Appendix 7 *(see Appendix 7 online)* provides a visual overview of the importance and non-linear impact direction of the 7 macro factors. The results indicate that the "Leading drivers" group (URB, EDU) acts as the strongest magnet; observations with rapid urbanization rates and high-quality education (red dots) exert a breakthrough positive impact on attracting highly educated personnel. Conversely, Residential area (ARE) has a complex impact; expanding current land funds (red dots) creates a hindering effect (negative SHAP value), reflecting the reality that large land funds are often located in suburban areas lacking synchronized amenities. Notably, the model detects an "Income Paradox" (INC) as this factor ranks in the weaker driver group, proving that once highly educated personnel reach a financial safety threshold, pure income disparity gives way to the urban and academic environment.

Regarding the methodological advantage, the SVR framework effectively overcomes the intrinsic limitations of traditional linear econometric models, such as Ordinary Least Squares (OLS) or Poisson regression. Preliminary statistical analysis of our dataset revealed severe multicollinearity among independent variables (e.g., Income and Residential Area correlation r = 0.8210; Education and Urbanization r = -0.8174), which would artificially inflate the variance and distort coefficients in an OLS model. Furthermore, linear models assume constant marginal effects and struggle to identify behavioral 'thresholds'. For instance, a traditional linear model would theoretically predict absolute talent retention in regions with superior economic and educational metrics. However, empirical spatial data reveals significant non-linear anomalies: the Red River Delta, despite being a leading educational hub, experienced a net loss of over 400 high-quality laborers. The SVR model, utilizing the RBF kernel to map independent variables into a multidimensional feature space, successfully filters out multicollinearity noise and captures these complex decision boundaries. Additionally, the SVR-SHAP approach explicitly uncovers the 'Income Paradox'. It mathematically demonstrates that once highly educated experts reach a specific income threshold, financial disparities lose their primary pull effect, yielding to the quality of the urban and educational ecosystem. This comparative capability proves SVR's superior robustness over OLS in modeling the complex realities of regional talent migration.

Clustering attraction strategies: The K-Means algorithm (K = 3) based on SHAP values Appendix 8 *(see Appendix 8 online*) indicates 3 strategic characteristics of the localities:

Cluster 0 (Living space & Urbanization): Ideal structure. Attracts through clean land funds combined with modern urban infrastructure and stable income.

Cluster 1 (Education & Society): Attracts via "soft infrastructure". Highly educated personnel accept cramped spaces to access excellent research environments.

Cluster 2 (Unbalanced development): The paradox of rapid urbanization but educational gaps, harboring a high risk of "brain drain" if soft infrastructure is not invested in.

5. Conclusion and policy implications

This study has successfully applied the Support Vector Regression (SVR) machine learning model combined with the SHAP explanatory tool to decode the migration behavior of highly educated personnel in

Vietnam. The SVR model demonstrated superior effectiveness in controlling overfitting and handling non-linear impact "thresholds" on a small dataset. The empirical results discovered and confirmed the existence of an "Income Paradox" among intellectuals: Urbanization (URB) and the educational ecosystem (EDU) are the decisive "magnets" for attracting talent. Meanwhile, Income (INC) is relegated to a secondary role due to its pronounced non-linear nature (diminishing marginal impact once experts reach a financial safety threshold). Furthermore, Residential area (ARE) exerts a complex impact, only proving effective when accompanied by synchronized amenities. Through the clustering algorithm, the study also issues a warning regarding the risk of "overheated" development in certain localities (urban expansion lacking social infrastructure), emphasizing the imperative of developing a comprehensive living ecosystem for sustainable talent accumulation (Giang et al., 2022).

Drawing from the research findings, the author proposes several strategic implications to regulate migration flows and foster sustainable brain circulation. First, for the Government and macroeconomic management agencies, it is necessary to shift the mindset from direct financial support to creating a comprehensive living and working ecosystem. Prioritizing national resources to build satellite "knowledge cities" (Florida, 2019), while resolutely allocating key social infrastructure projects (central-level hospitals, national universities), is extremely urgent to balance soft infrastructure amenities across economic regions. Second, for local authorities, attracting highly educated personnel should not stop at superficial cash-based "red carpet" policies. Localities need to focus on long-term retention policies such as providing high-quality official housing and improving the academic environment, while proactively attracting high-tech FDI projects to cultivate a professional environment commensurate with the intellectuals' qualifications. Third, for organizations and enterprises, establishing transparent career development pathways, autonomous research environments, and modern organizational cultures will serve as "soft infrastructure" compensating for local peripheral amenity deficits, thereby helping to retain elite personnel in the long run (Florida, 2019).

Despite achieving notable results, the study still possesses certain limitations, particularly concerning the data structure and sample size. First, the reliance on macro-regional data across six key economic regions over a 15-year period yields a relatively small sample size with 90 observations. Although the Support Vector Regression (SVR) algorithm operates with extreme stability and is highly effective at preventing overfitting on this specific small panel dataset, the restricted number of observations inherently limits the statistical power of the model. This small sample size may affect the absolute robustness of the findings when generalized or extrapolated to different economic contexts outside the study's scope.

Second, the use of aggregated regional-level data captures broad macroeconomic trends but fails to account for micro-level spatial disparities; specifically, the macro-regional data scale does not reflect the differential nuances among individual provinces or cities within a single region. This aggregation can mask localized pull factors and unique municipal policies, thereby limiting the generalizability of the results to specific, localized administrative units. Furthermore, due to the nature of secondary data, the study cannot fully encompass individual psychological factors, family ties, or indigenous culture in migration decisions. Therefore, future studies could expand by adopting a mixed-methods approach (Creswell & Creswell, 2017),

combining Machine Learning models with in-depth interview surveys and larger micro-level datasets to forecast talent migration flows more accurately and identify specific "expected income thresholds" within the modern labor market context.

Funding Acknowledgment

This research received no specific grant from any funding agency.)

Data Availability

Data available upon reasonable request (contact corresponding author).

AI Usage Statement

During manuscript preparation, the authors used AI Gemini for language editing. Authors reviewed, edited, and take full responsibility for the final content.

Reference

Athey, S. (2018). The impact of machine learning on economics. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 507–547). University of Chicago Press. https://doi.org/10.7208/chicago/9780226613475.003.0018

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approach* (5th ed.). SAGE Publications.

Docquier, F., & Rapoport, H. (2012). Globalization, brain drain, and development. *Journal of Economic Literature, 50*(3), 681–730. https://doi.org/10.1257/jel.50.3.681

Florida, R. (2019). *The rise of the creative class* (Revised and expanded ed.). Basic Books.

Floerkemeier, M. H., Spatafora, M. N., & Venables, A. (2021). *Regional disparities, growth, and inclusiveness*. International Monetary Fund.

Giang, N. H., Nguyen, T. T., Tay, C. C., Phuong, L. A., & Dang, T. T. (2022). Towards predictive Vietnamese human resource migration by machine learning: A case study in Northeast Asian countries. *Axioms, 11*(4), 151. https://doi.org/10.3390/axioms11040151

Glaeser, E. L. (2011). *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier*. Penguin Press.

Harris, J. R., & Todaro, M. P. (1970). Migration, unemployment and development: A two-sector analysis. *The American Economic Review, 60*(1), 126–142.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy, 99*(3), 483–499. https://doi.org/10.1086/261763

Mochiah, E. E. A., Osei, R. D., & Akoto, I. O. (2014). Deciding to urban-migrate and agricultural development: Evidence from the Millennium Challenge Account (MCC)–Millennium Development Authority (MiDA) intervention zones, Ghana. *Modern Economy, 5*(13), 1413–1428. https://doi.org/10.4236/me.2014.513131

Myrdal, G. (1957). *Economic theory and underdeveloped regions*. Duckworth.

Nha, P. H. T., & Tịnh, Đ. P. T. (2024). Tác động của các cách thức làm giảm lao động nông nghiệp dư thừa ở Đồng bằng sông Cửu Long [Impact of ways to reduce excess agricultural labor in the Mekong Delta]. *Tạp chí Khoa học Đại học Cần Thơ*, *60*(SDMD), 202-209. https://doi.org/10.22144/ctujos.2024.451

Paas, T., & Halapuu, V. (2012, June). Economic growth in the framework of Florida's 3T model: Theoretical considerations and empirical evidence. In *Enterprise Odyssey: International Conference Proceedings* (pp. 144–156). Faculty of Economics and Business, University of Zagreb.

Sjaastad, L. A. (1962). The costs and returns of human migration. *Journal of Political Economy, 70*(5, Part 2), 80–93. https://doi.org/10.1086/258726

Tran, T. T. T., & Nguyen, N. K. A. (2021). Kiến thức và hành vi tình dục an toàn của người di cư lao động tự do tại phường Hoàng Liệt, quận Hoàng Mai, Thành phố Hà Nội năm 2021 [Knowledge and safe sexual behaviors of freelance migrants in Hoang Liet Ward, Hoang Mai District, Hanoi City in 2021]. *Tạp chí Y học Việt Nam, 507*(2), 249-254. https://doi.org/10.51298/vmj.v507i2.1452

Tzeng, R. (1995). International labor migration through multinational enterprises. *International Migration Review, 29*(1), 139–154. https://doi.org/10.1177/019791839502900107

Vega-Muñoz, A., González-Gómez-del-Miño, P., & Contreras-Barraza, N. (2024). Determinants of brain drain in 178 countries from 2006 to 2022. *(Cần bổ sung tên tạp chí, volume, issue, pages và DOI hoặc URL chính thức; thông tin hiện tại chưa đủ để chuẩn hóa APA.)*

World Bank (2019). *World development report 2019: The changing nature of work*. World Bank. https://doi.org/10.1596/978-1-4648-1328-3