

Tony Tran
Professor Pablo Crespo
Data Bootcamp, Section 2
December 22nd, 2022

Explaining and Predicting Vietnamese Inter-Province Net Migration

1. Introduction

Internal migration is a topic that is heavily researched and followed as the inflow and outflow of a country's population can cause large disruptions to the housing, political, social, and ecological structures of a local area (Davanzo). By forecasting net migration, policy makers can anticipate migrations and formulate policies accordingly.

The objective of this paper is to identify the significant economic and non-economic characteristics that influence net migration and create a model of how these significant characteristics of different provinces and centrally controlled municipalities in Vietnam influence the magnitude of net migration.

This paper tries to explain two different measures of net migration in Vietnam: the net migration rate and absolute net migration. The net migration rate measures the difference between the percentage of a province's population that moved out of the province and the percentage of a province's population that moved in. The absolute net migration takes the net migration rate and multiplies it by the population of the location to arrive at the number of people that moved in or out of the province in a year.

The fitted model could potentially be used to predict future internal migration. The predictions can be used for policymakers to understand the correlations between various economic and noneconomic characteristics of an area to its net migration.

2. Literature Review

2.1 Why Do People Migrate?

Although there are a multitude of reasons why people decide to internally migrate, in brief, people migrate to maximize utility: people migrate when they consider that the benefits of migration outweigh the costs of migration (Davanzo).

The main reasons why people migrate internally are to change job, for marriage, to follow family members, to escape distressed locations, or even to satisfy some longing or impulse of theirs (Voss)(DaVanzo)(Ngoc Hung Pham). However, the main reason

why people internally migrate is to go where opportunity is, and most of the time, economic opportunities (Rogers)(Davano).

Sample data from NLS Young Men for the period 1971-73 found that 85% of young men moved for economic reasons and 47% of migration is caused by the decision to change jobs (Bartel). A factsheet published by the United Nations Population Fund on the population and housing census in 2019 in Vietnam showed that 36.8% of migrants' main reason for migrating was to find or start a new job and this was the most common reason in 2019.

In the decision to move, employment, occupation, and salary are the major motives and migration serves to close the employment and wage differentials between different locations (DaVanzo). The decision to move is motivated by the changing economic conditions of different areas and when novel economic opportunities present themselves, positive net migration is expected (Dang)(Rogers). Ultimately, migration is for people to be in the location where their production capabilities are maximized, and such migration is critical for an efficient labor market.

2.2 Previous Migration Regression Models and Research on Migration in Vietnam

The Vietnam modern wave of migration and urbanization began with the implementation of the Doi Moi policy in 1986 which marked the shift of the Vietnamese economy from a centrally planned economy to a “socialist-oriented market economy” (The ASEAN Post Team). Additionally, the 1986 economic reform began promoting foreign direct investment (FDI), promoting international trade, and instituting land law reform (Du). The implementation of the Doi Moi policy initiated a period of rapid industrialization, and growing FDI investments and trade. After the implementation of the Doi Moi policy, employment share in the Vietnamese industry and mining sector grew from 11.6% at the time of implementation to 22.7% in 2013 (Phuong Nguyen-Hoang)(Danh). The rapid industrialization has caused urbanization with net migration to large cities like Ho Chi Minh City to reach a maximum of over 20% annually and current population growth stabilized at 4%, 3% above the population growth (Vietnam General Statistics Office)(Macrotrend, Ho Chi Minh...)(Macrotrend, Vietnam Population...).

A paper by Smith and Scarpaci which studies the migration to the two largest cities in Vietnam, Hanoi and Ho Chi Minh City, found that the process of migration is an

integral process of industrialization and industrialization is a major reason that constitutes urbanization. While in-migration and industrialized levels are positively correlated, we should take a closer look at which aspects of industrialization and development are significant which will be discussed further in the variables section.

In 1997, Anh Dang found that regions that are more developed places tend to attract an inflow of people and less developed provinces on average have outflows of people. He also found that migrants, especially females, moved to more urbanized and industrialized areas. In addition, the regression run by Dang also shows that besides economic factors, non-economic factors such as living standards, distance, education, and recreation play a significant role in either explaining in-migration or out-migration. The limitation of Dang's paper is that the regression was run using data from 1984-1989, which is the period before the Doi Moi policy and the start of the Doi Moi policy. However, the economic and industrial landscape of Vietnamese provinces has undergone significant changes since 1989, and the results and validity of his regression might not be true currently.

Another study by Ian Coxhead has found that migration in Vietnam is more frequent among younger people, more educated people, people from female-led households, and non-ethnic minorities. The age factor is significant since the benefits from migration from young people are greater than that of old people. This is simply because they have on average more time to live and benefit from the relocation. Young people are also likely to move because their costs of moving are less than older people; as people age in a place, they have more assets tied to that place which increases their cost to move.

A regression on inter-provincial migration in Vietnam has also been done by Phuong Nguyen Hang in 2010 using the General Statistics Office of Vietnam 2004-2005 survey. He found that Vietnamese migrants care about moving costs, expected income differentials, the disparity in the quality of public services offered, and demographic composition.

Many of the previous regressions ran on migration in Vietnam have used census data to run their regression. For this paper, the characteristics and statistics of provinces will be used to conduct the regression.

3. Data

The dataset used for this regression is cross-sectional data which compiled different datasets published by the General Statistics Office of Vietnam for different characteristics of Vietnamese provinces and centrally controlled municipalities in 2018. 2 provinces were excluded due to missing data. The regressed dataset consists of 56 provinces and 5 centrally controlled municipalities for a total of 61 observations. The following variables are included in the dataset:

Dependent variables:

- Net migration rate (%): This variable is calculated by taking the percentage of the province population who migrated in to that province in the year 2018 minus the percentage of the province population who moved out of that province in 2018. Net migration rate reflects the change in the population of an area due to migration in a year.
- Absolute migration: The net migration rate times the population. We will run a different set of regressions for absolute migration.

Independent variables:

Economic Variables:

- Income per capita (Hundred thousand Vietnamese Dong)
- Industrial Production Index
- Number of foreign direct investment projects
- Total registered foreign direct investment capital (in hundred million
- Percent of adults over 15 who are employed
- Percent of trained adults over 15 who are employed
- Number of new enterprises
- 3 years average income growth

Non-Economic Variables:

- Sex ratio (males per 100 females)
- Number of university teachers / Number of university teachers per 100,000 people
- Population density (people / km²)
- Population
- Area (km²)

- Percentage of households with electricity
- Municipalities (0 for provinces and 1 for municipalities)

4. Model & Possible Limitations

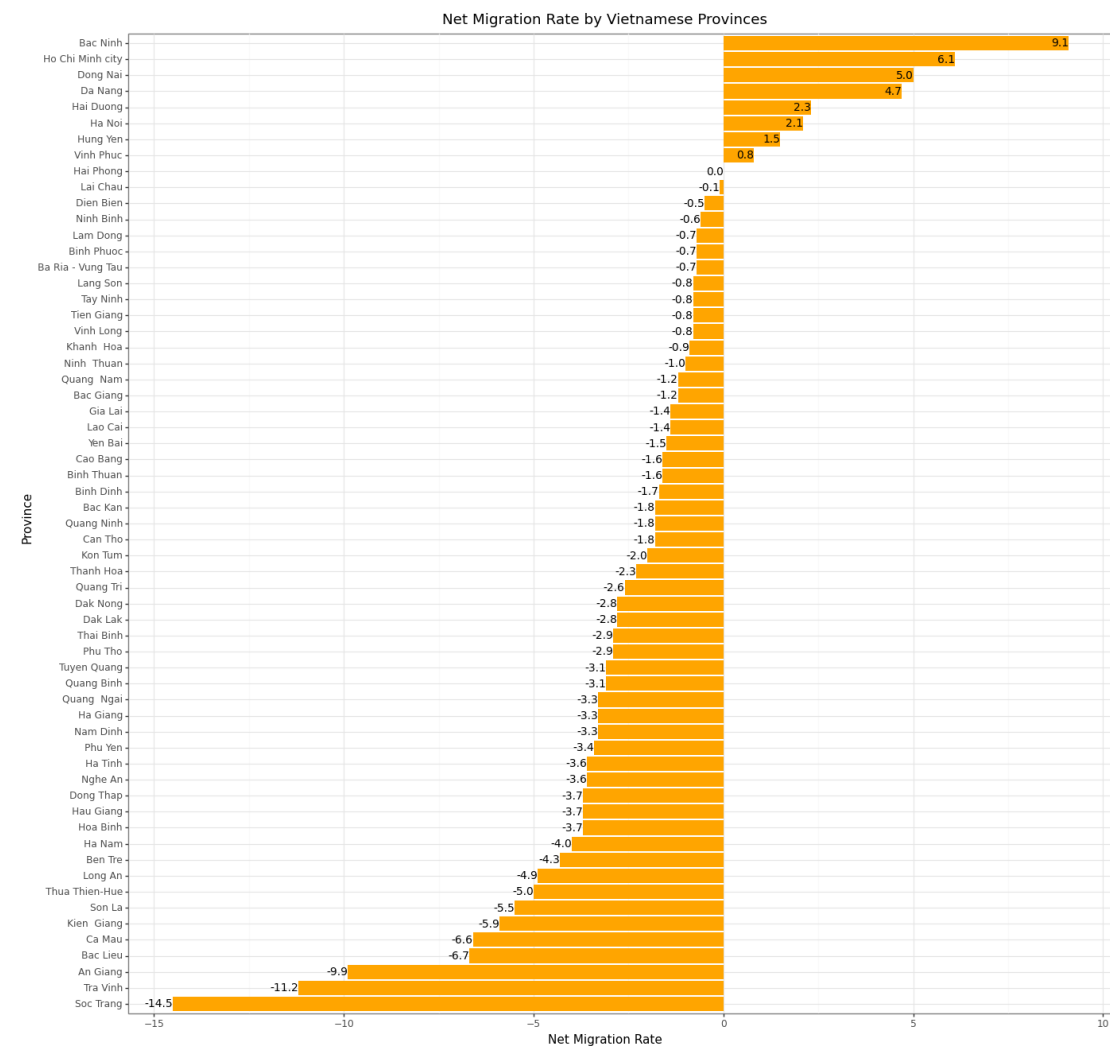
The model that will be used for understanding the relationship between the characteristics and the net migration of a province is either an ordinary least squares regression or a regression through the origin. Some limitations of the regression include:

- 1) Small sample size: Since there are only 58 provinces and 5 centrally controlled municipalities in Vietnam in 2018, the maximum number of observations in the regression is only 63. However, in the regression, two provinces are not included due to too many data fields missing.
- 2) Reverse causality: A few variables included in the regression intended to explain net migration are likely to be partially explained by net migration. For example, foreign direct investments might attract people to migrate to a province. However, provinces with high migration, indicating a strong inflow of human capital, might lead to higher foreign direct investment.
- 3) Weak relationships: Some of the variables such as the number of university teachers or the number of university teachers per 100,000 people are used in this regression to try to capture the effect of education on migration. However, the number of university teachers is only a small contribution to explaining the education level of a province and does not capture wholly the effect of education on migration.
- 4) Aggregate data of places do not capture the characteristics of those who move: Age, education levels, and marital status all contribute to whether people move. However, data from different provinces are aggregate data which includes characteristics of those who will move and those who do not move. Therefore, this dataset does not explain who is more likely to move, only what characteristics a province has or lacks that cause people to be more likely to move in or out.

After understanding which characteristics play an instrumental role in determining the migration rate, we would use those characteristics to perform a random forest classifier. This is in hopes of getting a better prediction than the ordinary least squares regression method.

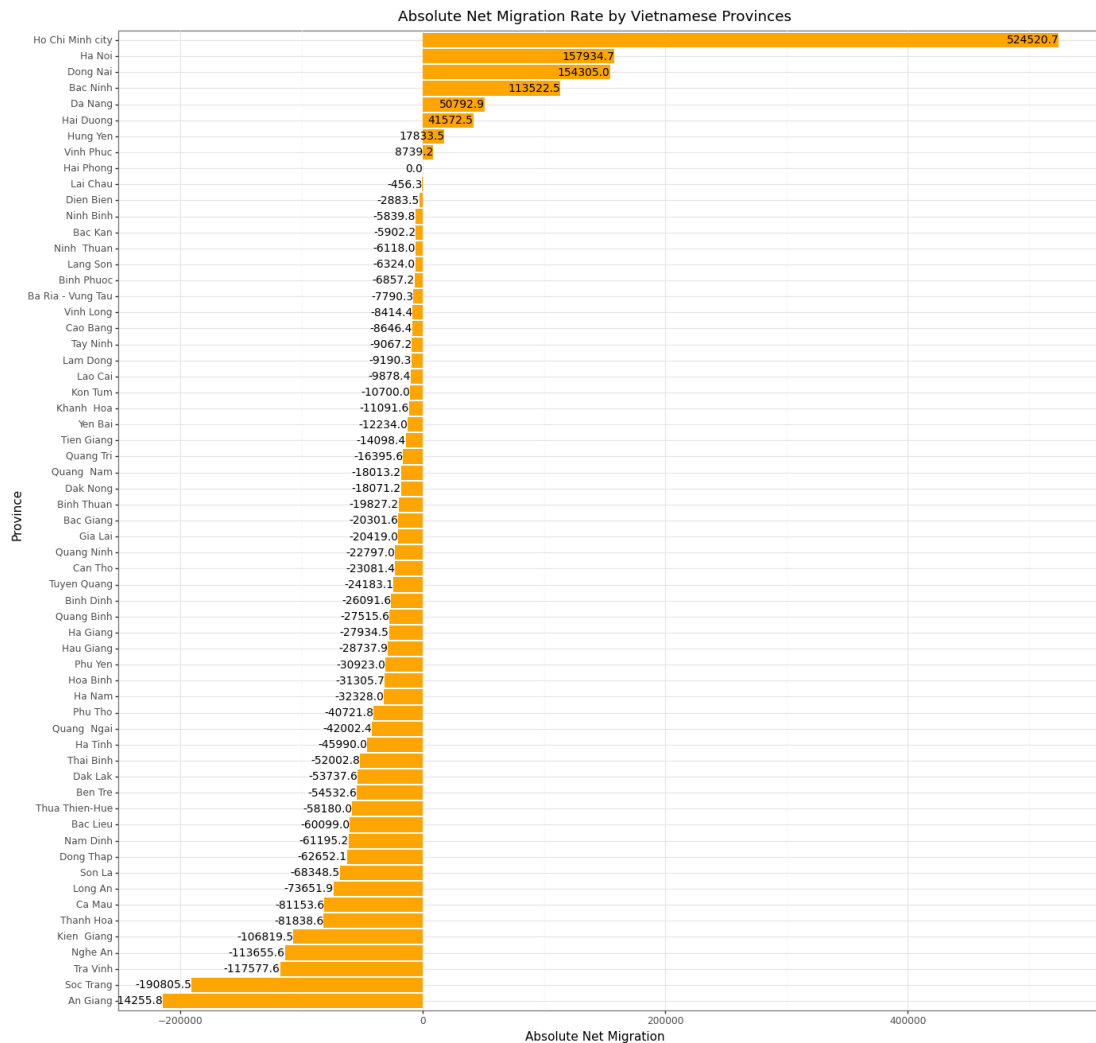
5. Data Visualization

Net Migration Rate:



Observation: Most provinces have a negative net migration rate. Therefore, the average net migration rate is not 0. Since the average net migration rate is not 0, we might have to use an intercept in the linear regressions. The hypothesized reason why the intercept is not 0 is that provinces with a more populous population tend to have a positive net migration because usually, both their economic and non-economic conditions are better than those with a smaller population. Because the province with a large population receives in absolute terms a large number of migrants, in percentage terms the number is not that high. So, a few large provinces receive lots of immigration but are relatively small compared to their total population. This causes the average net migration rate to not be equal to 0.

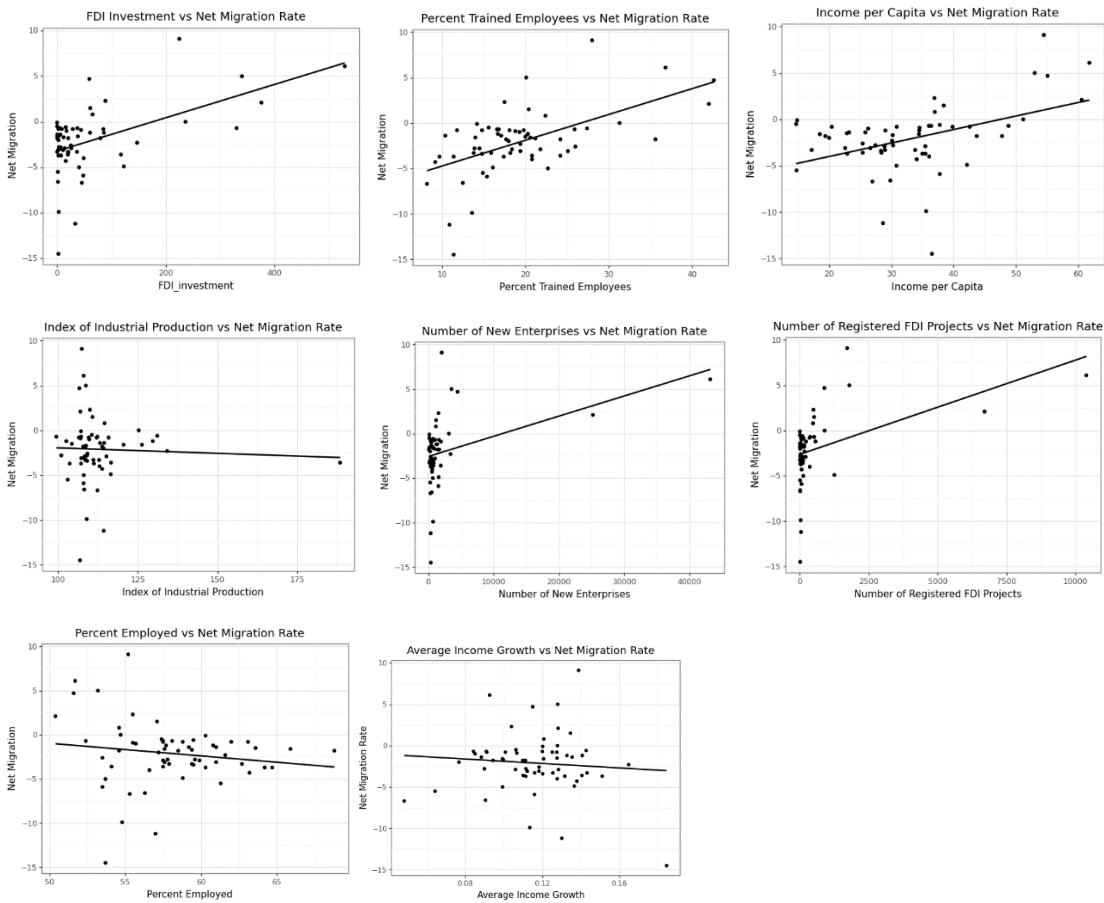
Absolute migration:



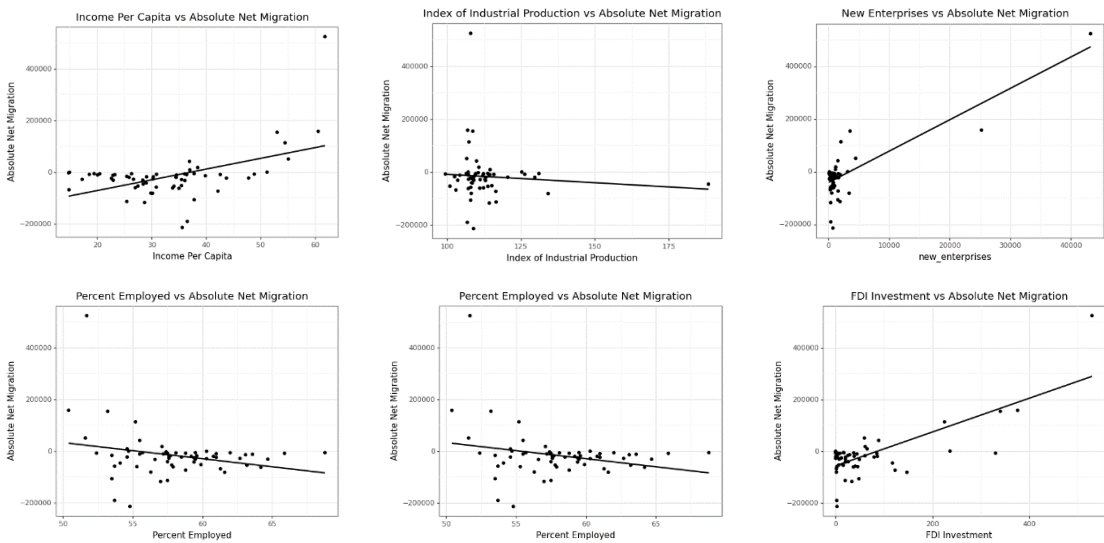
Observation: When looking in absolute terms, we can see that the top few provinces receive in absolute terms a large number of immigrants. The few top provinces all have large cities in them with a large population and usually better economic conditions (from experience living in Vietnam). This shows that people are moving from smaller to bigger provinces which is consistent with previous findings that Vietnam is industrializing and a key part of industrialization is urbanization. The effects of the rural population moving to urban areas can be seen in

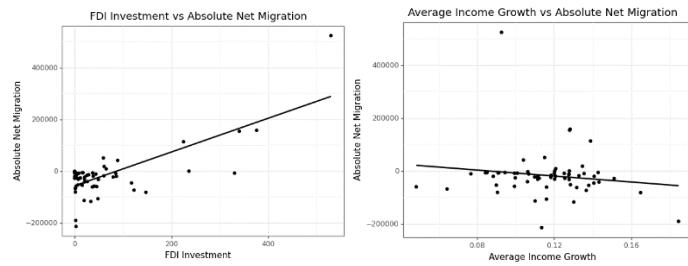
Correlation of Net Migration with Economic Variables:

1) Net Migration Rate



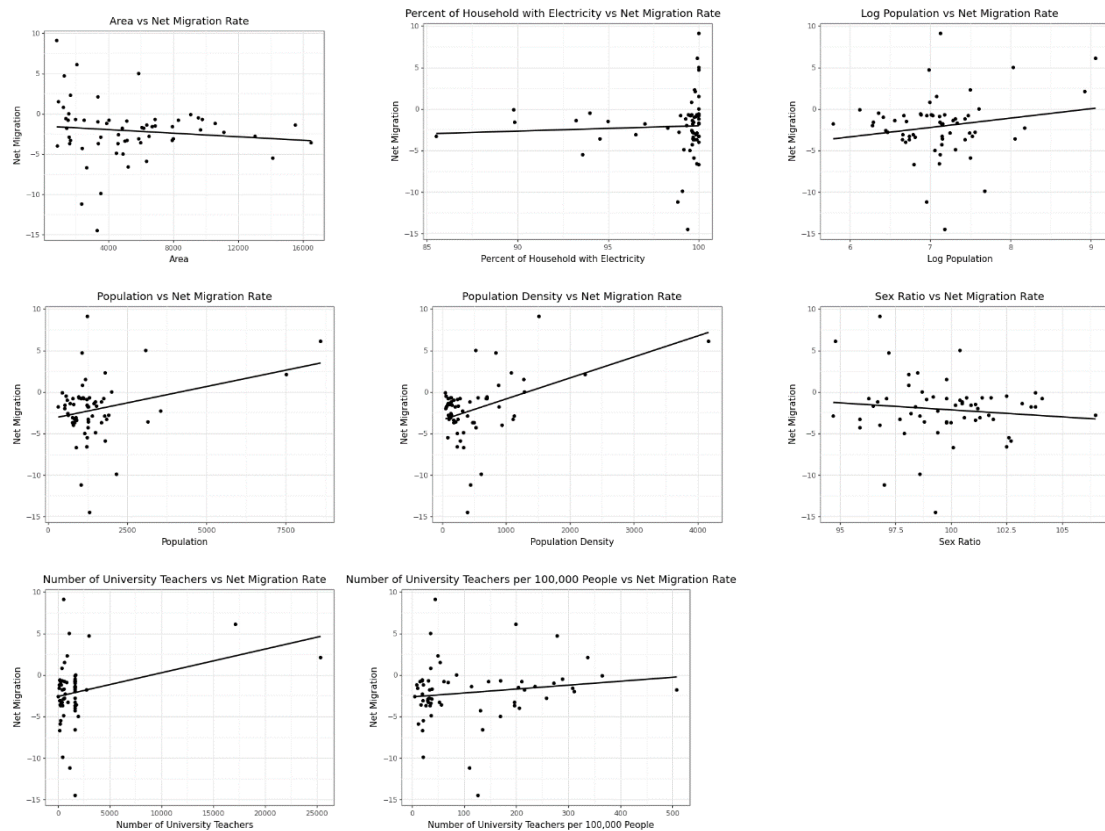
2) Absolute Migration



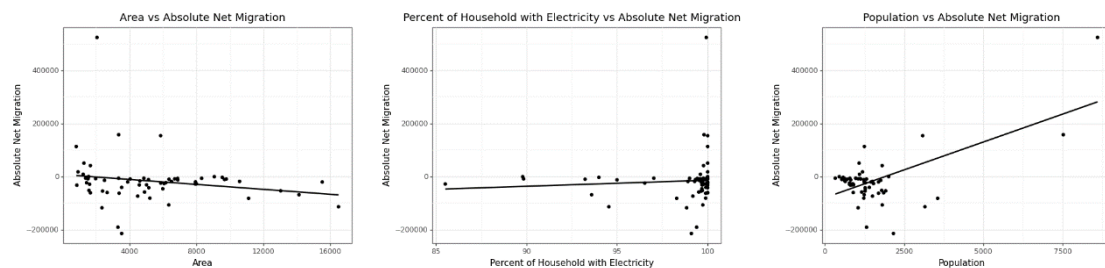


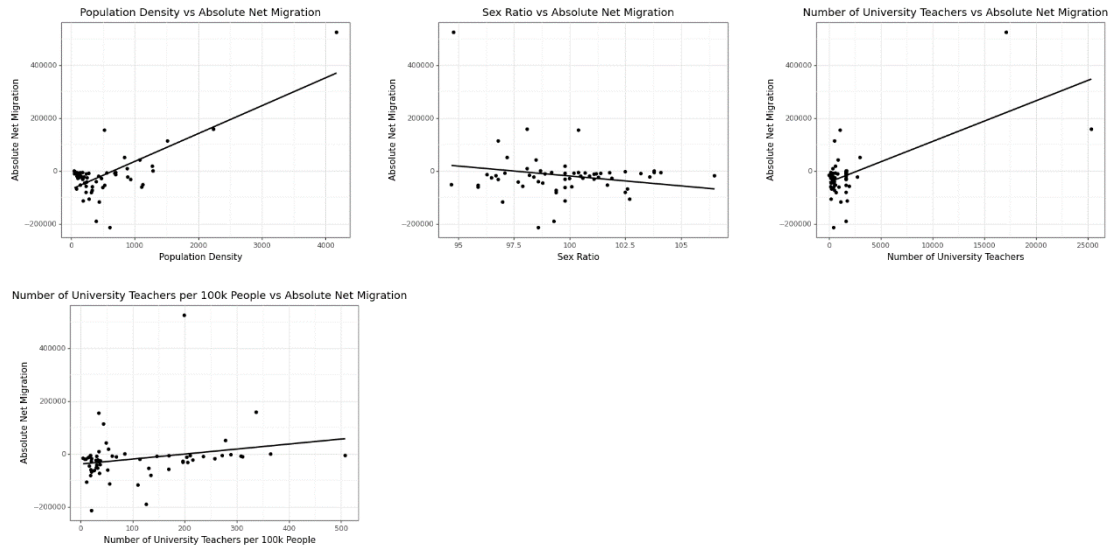
Correlation of Net Migration with Non-Economic Variables:

1) Net Migration Rate:



Absolute Migration:





6. Results

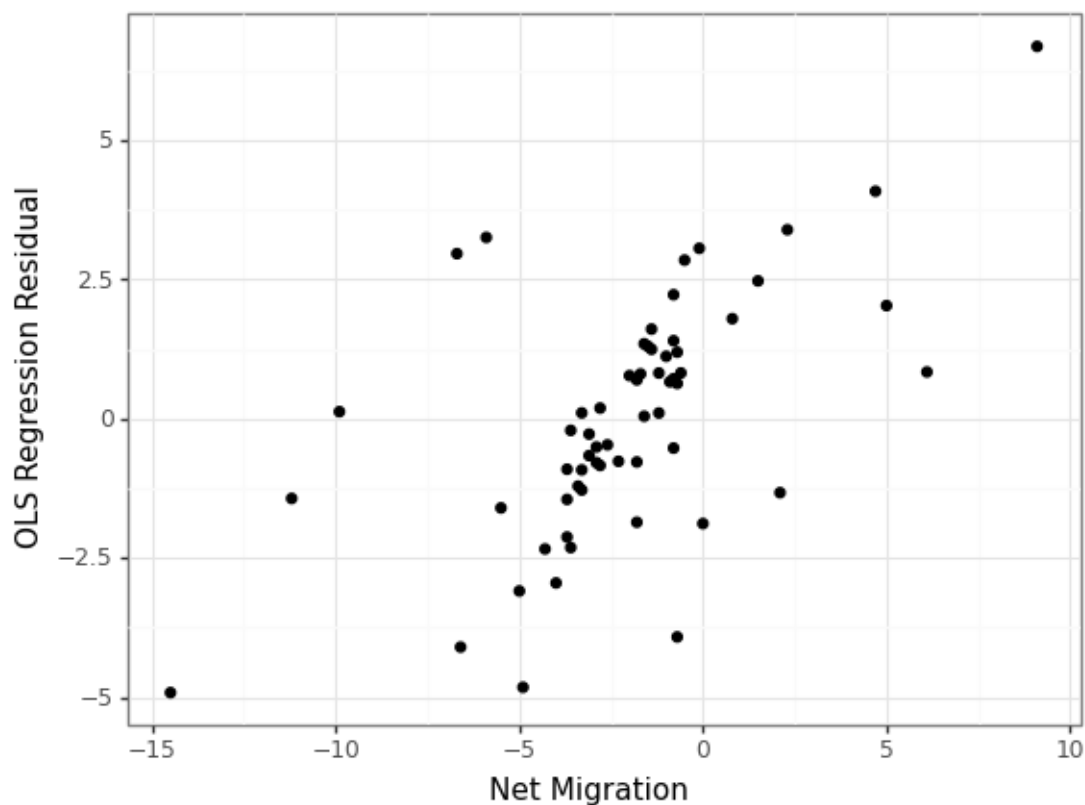
a. Net Migration Rate Regressions Result:

	Dependent variable: net_migration				
	(1)	(2)	(3)	(4)	(5)
FDI_investment				0.016*** (0.005)	0.012*** (0.004)
income_per_capita	-0.044*** (0.014)		0.184*** (0.045)	0.070 (0.057)	0.092** (0.042)
khmer_more_70k		-9.640*** (1.412)			-7.657*** (1.052)
lnpop			-1.152*** (0.219)	-0.762*** (0.245)	-0.742*** (0.178)
Observations	61	61	61	61	61
R ²	0.132	0.437	0.408	0.486	0.733
Adjusted R ²	0.117	0.428	0.388	0.459	0.715
Residual Std. Error	3.921 (df=60)	3.157 (df=60)	3.265 (df=59)	3.069 (df=58)	2.229 (df=57)
F Statistic	9.109*** (df=1; 60)	46.621*** (df=1; 60)	20.356*** (df=2; 59)	18.267*** (df=3; 58)	39.205*** (df=4; 57)
Note:	*p<0.1; **p<0.05; ***p<0.01				

b. Absolute Migration Regressions Result:

Dependent variable: net_mig_abs						
	(1)	(2)	(3)	(4)	(5)	(6)
FDI_investment		653.513*** (78.243)		323.182*** (91.030)	434.481*** (89.121)	870.367*** (159.401)
Intercept	-154515.291*** (33415.588)	-56525.182*** (9332.228)	-58452.067*** (8885.179)	-59436.132*** (8115.501)	271981.085*** (95205.075)	
fdiper100k						-91.820*** (23.287)
income_per_capita	4159.055*** (960.561)					1118.493*** (357.756)
khmer_more_70k			-122384.504*** (24728.563)	-113885.329*** (22699.862)	-102198.192*** (21021.796)	-93832.126*** (14311.477)
lnpop					-48762.420*** (13965.207)	
new_enterprises						13.272*** (1.742)
population						-63.768*** (8.134)
population_density			103.434*** (10.629)	64.727*** (14.595)	79.668*** (14.013)	
Observations	61	61	61	61	61	61
R ²	0.241	0.542	0.680	0.738	0.785	0.906
Adjusted R ²	0.228	0.534	0.669	0.724	0.769	0.895
Residual Std. Error	80804.102 (df=59)	62788.794 (df=59)	52939.800 (df=58)	48325.682 (df=57)	44182.332 (df=56)	30032.481 (df=55)
F Statistic	18.747*** (df=1; 59)	69.762*** (df=1; 59)	61.564*** (df=2; 58)	53.456*** (df=3; 57)	51.012*** (df=4; 56)	87.872*** (df=6; 55)
Note:						*p<0.1; **p<0.05; ***p<0.01

c. Net Migration OLS Regression Residuals



7. Discussion

The models and variables that are found to significantly explain migration included

most of the economic variables while most of the non-economic variables in the dataset that were hypothesized as having a link to net migration were not significant in predicting the net migration values in Vietnam. This does not mean that people do not move for non-economic reasons, however, this could signify the limitation of the dataset used for the regression.

After fitting 4.1 model 4 and plotting the residuals from the variables included in the dataset, we found that for most of the provinces, the residuals were relatively close and cantered at zero; however, there are a few areas that the regression model poorly predicted: these provinces have strong forces that lead to large negative net migration that the model didn't include. After further inspection of the poorly predicted provinces, it was observed that these provinces have large amounts of Khmer people. Therefore, a dummy variable of 0 for provinces with a Khmer population of less than 70,000 and 1 for provinces with a Khmer population of more than 70,000 was added to regression 5. After considering for the Khmer population, the residuals became relatively homogeneous. Further examination of migration data and trends for provinces with high Khmer populations in the years 2019 and 2020 also showed similar results. However, the significance of the Khmer population might be because of omitted variables rather than cultural reasons for Khmer people to move away from their homes. One hypothesis could be that currently, a high proportion of the Khmer population holds an agriculture job. As the country industrialized, people shift from agriculture to industrial jobs which could explain the negative net migration rate of high Khmer population areas. This is also consistent with previous findings that non-ethnic minorities are more likely to migrate.

Although many variables were not significant in predicting net migration as first thought, however, almost all economic variables were significant in predicting net migration. This is consistent with previous migration theories and research that the main motive for migration was due to differences in economic opportunities between regions and migration was a way of exploiting job opportunities and wage arbitrage.

Out of the economic variables, the economic variable FDI investment was the only variable that was positively significant at the 1% level in all models, which signifies that FDI could be the most significant factor in explaining in-migration. Looking at 4.2 models 1 and 2, when switching income per capita for FDI investment, the R-squared

increased significantly from 0.241 to 0.542. This shows that people are more interested in the new opportunities that a region has to offer rather than the current economic conditions. This makes sense as when foreign companies invest in Vietnam, they shift the demand for labor up and people will move in to take advantage of the higher price. However, for places already with high income per capita, there might not be as much demand for workers as the equilibrium price is already high.

We know that FDI investments are significant in attracting people. As FDI pour in, people from other provinces are required to migrate to the province with FDI to increase the supply of labor needed to meet the needs of the investment. So which province is providing the labor necessary for these FDI investments? In the regression, a variable that was negatively consistently significant in all models is the Khmer population. It is therefore hypothesized that provinces with a high Khmer population move to provinces where FDI investments are high.

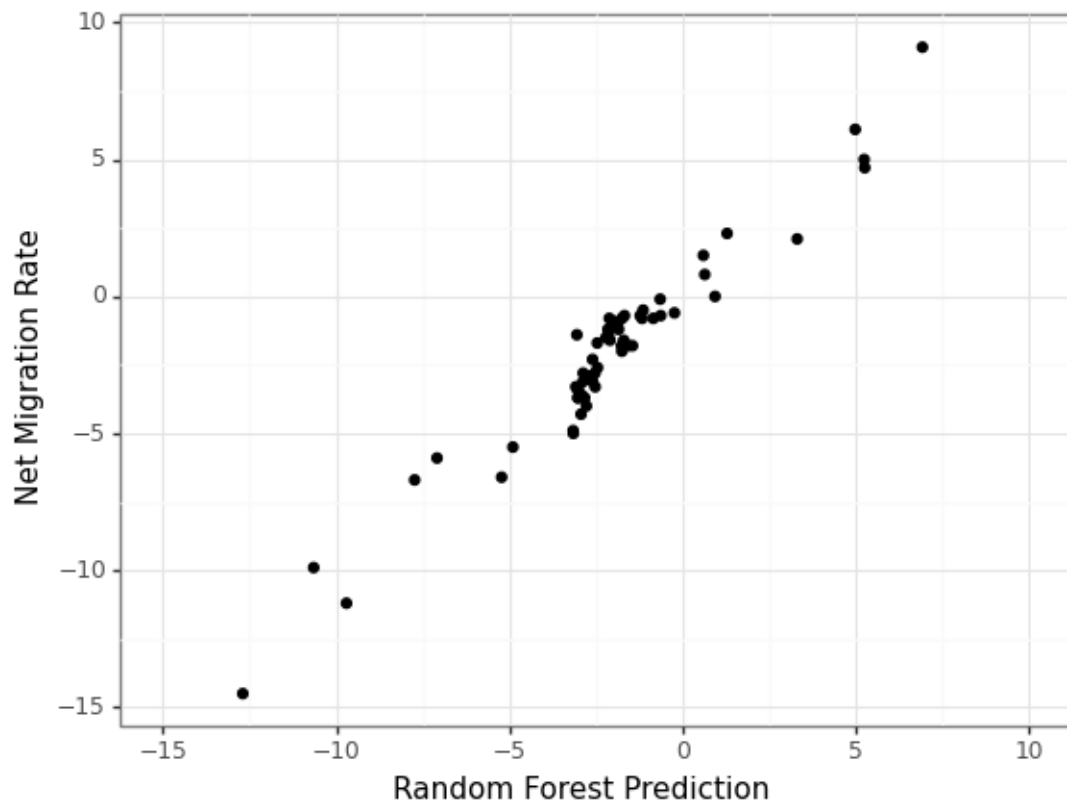
The interaction between the out-migration of the Khmer population to places where there is a high concentration of foreign direct investment shows the decision to migrate is largely driven by shifts in the demand curve in the labor market that indicate new opportunities for economic growth. The response of migration will raise the supply of labor offsetting partially the increase in price brought by the FDI investment. This would mean that the supply of labor in the province that experiences an out-migration would decrease and the equilibrium wage in that province will be higher.

Another factor aside from economic factors that affected migration was population. In particular, both models found that population had a negative correlation with migration when controlled for economic factors.

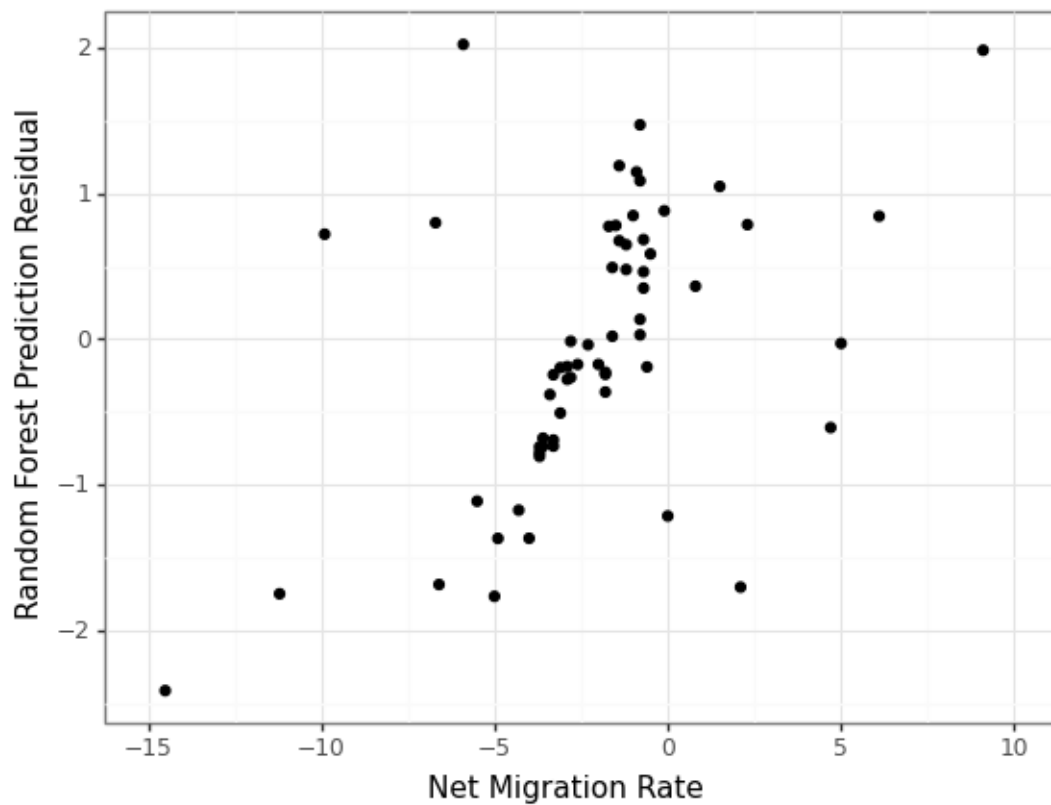
6. Fitting a Random Forest Classifier

In this section, we are attempting to take the features from the model from net migration and absolute migration with the highest R-square to fit a random forest classifier in hopes of getting a better prediction than the ordinary least squares regression. We will use the features in model 5 to predict the net migration rate. After using cross-validation and finding the optimal depth (Appendix 1), we then apply the random forest classifier to predict the net migration rate and absolute migration. The comparison between the predicted values and actual values along with the residual graphs is shown below:

Predicted vs Actual Values Graph:



Residual Plot of Random Forest Classifier:



We can see that the random forest classifier did a much better job at predicting net migration than the linear regression. The residuals of the random forest only range from -2 to 2% while the range of the residuals of OLS ranges from -5 to more than 5%. So, the random forest classifier did help us better predict the net migration rate. It also helped us avoid data snooping and overfitting with the help of cross-validation.

8. Conclusion

The decision to migrate happens when a person believes that the net present value of the benefits of moving is greater than the costs of moving. Despite the fact that a benefits cost analysis may involve multiple factors and considerations that can differ for everyone, economic benefits are still the most important consideration when deciding whether to migrate. Out of the economic changes, the economic change that impacts migration the most is the shift in labor demand and supply. Although non-economic reasons are not fully captured in this model, they are also important and should also be considered.

Works Cited

- Dang, Anh, Sidney Goldstein, and James McNally. "Internal Migration and Development in Vietnam." *The International Migration Review* (1997): 312-337.
- Bartel, Ann P. "The Migration Decision: What Role Does Job Mobility Play?" *The American Economic Review* (1979): 775-786.
- Davanzo, Julie. "MICROECONOMIC APPROACHES TO STUDYING MIGRATION DECISIONS." *The National Institute of Child Health and Human Development, National Institute of Health* (1980): 1-56.
- Du, Phuoc Tan. *TRANSFORMATION OF SOCIO-ECONOMIC STRUCTURE OF HO CHI MINH CITY UNDER THE DOI-MOI POLICY AND THE ACCOMPANYING GLOBALIZATION PROCESS*. Nagoya: Meijo Asian Research Journal, n.d.
- Ian Coxhead, Nguyen Viet Cuong, Linh Hoang Vu. "Migration in Vietnam: New Evidence from Recent Surveys." *VIETNAM DEVELOPMENT ECONOMICS DISCUSSION PAPER* (2015): 2-33.
- Macrotrends. *Ho Chi Minh, Vietnam Metro Area Population 1950-2022*. 2022. 2 December 2022.<<https://www.macrotrends.net/cities/22457/hanoi/population#:~:text=The%20current%20metro%20area%20population,a%2004.21%25%20increase%20from%202020>>.
- Macrotrends *Vietnam Population 1950-2022*. 2022. 2 December 2022.<<https://www.macrotrends.net/cities/22457/hanoi/population#:~:text=The%20current%20metro%20area%20population,a%2004.21%25%20increase%20from%202020>>.
- Ngoc Hung Pham, Manh Dung Tran, Anh Duc Le, and Thuy Linh Le. "The Impact of Migration on Vietnam Household Living Standards." *Journal of Risk and Financial Management* (2022): 201.
- Phuong Nguyen-Hoang, John McPeak. "Leaving or Staying: Inter-Provincial Migration in Vietnam." *Asian and Pacific Migration Journal* (2010): 473-500.
- Rogers, Andrei. "A Regression Analysis of Interregional Migration in California." *The Review of Economics and Statistics* 49.2 (1967): 262-267.
- Scarpaci, David W. Smith & Joseph L. "URBANIZATION IN TRANSITIONAL SOCIETIES: AN OVERVIEW OF VIETNAM AND HANOI." *Urban Geography* (2000): 745-757.
- Voss, Guangqing Chi, and Paul. "Migration Decision-making: A Hierarchical Regression Approach ." *The Journal of Regional Analysis & Policy* (2005): 11-22.
- The ASEAN Post Team. *Vietnam's Reforms Paying Off*. 5 June 2020. 19 December 2022.
- United Nations Population Funds. *MIGRATION AND URBANIZATION IN VIET NAM*. Factsheet. Ha Noi, 2019.
- Vietnam General Statistics Office. "In-migration rate, out-migration rate and net-

migration rate by province." 2005-2021.

Appendix:

Appendix 1: Optimal Depth for Net Migration Prediction

