

Relationship between Permanent Resident Population and Dependency Ratio of Heilongjiang Province

Yuxuan Yang, Xinyang Lyu, Zesheng Yu, Ziyao Xu

Beijing National Day School

09/06/2023

Abstract

Population decline in Heilongjiang, deteriorating its economic situation, requires more attention. Here, we investigated the correlation between the permanent resident population and dependency ratio by calculating a least-square regression line and conducting a hypothesis test for β , which proves the significance of β . In addition, we applied time series to predict the permanent resident population for the years 2020 to 2025, which yields a very small error, indicating that the prediction is accurate.

1 Introduction

The report aims to investigate how the decrease in the permanent resident population is associated with increasing the burden of the existing labor force, measured by the dependency ratio, in Heilongjiang Province. The significance of this investigation is that by gaining insight into how the decrease in the permanent resident population is associated with the increase in the burden of each labor force, we can call attention to the problem and serious effects of the decline in the permanent resident population in Heilongjiang province. Also, given the current trend, the research will predict the permanent resident population and dependency ratio from the year 2020 to 2025. A linear regression model is used to explain the relationship between the permanent resident population and the dependency ratio with data from 2010 to 2019 in Heilongjiang province. Other models may be found to be the best fit in the process. A hypothesis test for β will be conducted. The null hypothesis will be $\beta = 0$; the alternative hypothesis will be $\beta < 0$. After that, time series will be used to predict the value of the permanent resident population and dependency ratio of Heilongjiang province from the year 2020 to the year 2025.

Keywords— permanent resident population, dependency ratio, linear regression model, hypothesis test, time series

2 Background Research

After the reform and opening-up policy had been implemented in China, the overall economic strength of China, especially in Southeastern coastal cities like Shenzhen, increased sharply. However, three provinces, Heilongjiang, Jilin, and Liaoning, located in the northeast of China, are the exception to the economic growth from the 1980s to 1990s. In Heilongjiang province specifically, the reason for the stagnation or even decline of the economy is attributed to the fact that it almost solely depends on heavy industry as the backbone of Heilongjiang's economy. Environmental protection policy, high transportation expenses due to its location and harsh environment, and less demand for its unadvanced product impeded the development of heavy industry in Heilongjiang province. The decline in the economy, job opportunities, infrastructure, and public services made young people from Heilongjiang leave for highly developed cities such as Beijing and Shanghai (Fenggang Lu, 2021). According to the data of the "Seventh National Population Survey," the overall household population of the three provinces in Northeast China is 103.46 million, which is about 4.46 million less than ten years ago; the resident population is 98.51 million, which is about 11 million less than ten years ago. The primary motivation for local people to leave Heilongjiang province is to seek economic gain (Nan Zhang, 2016). LinXi University researchers suggest that older people left alone in Heilongjiang province require family care and a heavy family burden, which leads to a higher dependency ratio.

It impacts the economy, and there is a vicious cycle between the decrease in population and the decline in the economy. Previous researchers have also offered methods to solve the problem of population loss in Heilongjiang province, such as upgrading the industrial structure, increasing job opportunities, and raising the level of payment (Fenggang Lu, 2021).

3 Methods and Procedure

The raw data on the permanent resident population and dependency ratio of Heilongjiang province from 2010 to 2019 were collected from the National Bureau of Statistics, which offers authoritative data to be used. The raw data is shown below

Year	Permanent Resident Population (million)	Dependency Ratio (%)
2010	38.33	25.4
2011	37.82	24.6
2012	37.24	26.5
2013	36.66	26.5
2014	36.08	26.6
2015	35.29	27.3
2016	34.63	28.2
2017	33.99	28.3
2018	33.27	29.5
2019	32.55	31.2

Table 1. Permanent Resident Population and Dependency Ratio of Heilongjiang Province from 2010 to 2019

Then normalize the raw data:

Year	Z-score of Permanent Resident Population	Z-score of Dependency Ratio
2010	1.395786831	-1.027972002
2011	1.136365809	-1.437115088
2012	0.84133798	-0.46540026
2013	0.546310152	-0.46540026
2014	0.251282323	-0.414257374
2015	-0.150565926	-0.056257174
2016	-0.486287249	0.404028797
2017	-0.811835198	0.455171683
2018	-1.183163327	1.068886311
2019	-1.544318083	1.938315368

Table 2. Z-score of Permanent Resident Population and Z-score of Dependency Ratio

Then convert the data into a scatterplot:



Figure 1. Scatterplot of Z-score of Dependency Ratio and Z-score of Permanent Resident Population with Linear Regression Line

Figure 1. shows that the equation for the least square regression line is $y = -0.958x - 0.0005$, and the coefficient of determination, R^2 , for the regression line is 0.919, which means that about

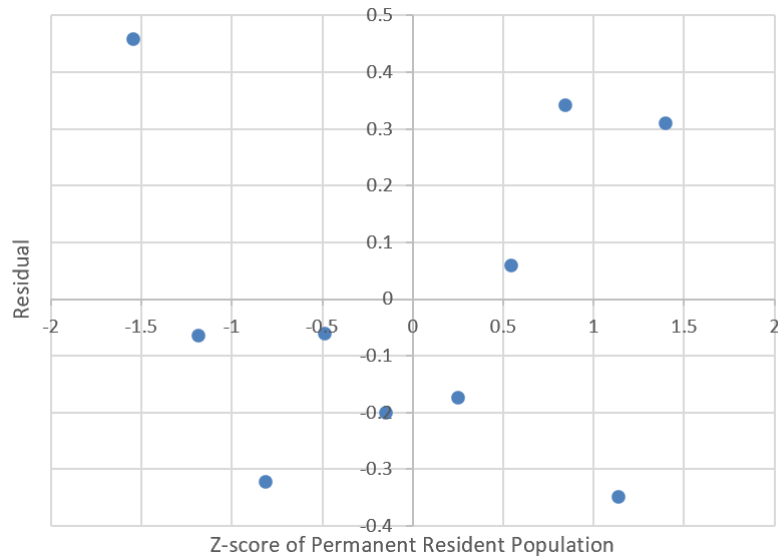


Figure 2. Residual Plot for Linear Regression Model

Figure 2. shows that the residuals are approximately evenly distributed around the x-axis and do not show an apparent curved pattern. Therefore, it suggests a linear relationship between the permanent resident population and the dependency ratio. In addition, the residual plot also shows that the magnitude of the residual fluctuation does not vary significantly as the permanent resident population increases. Therefore, the condition of homogeneity of variance is met.

Thirdly, the independence within the data is checked. Using the Durbin-Watson test yields a value of 2.015, as shown in Table 3. The closer the value of the Durbin-Watson Test is to 2, the smaller the autocorrelation is. The autocorrelation is negligible if the value is between 1.7 and 2.3. Since the value of Durbin-Watson is 2.015, the model has little autocorrelation, and the data can be regarded as independent.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.959 ^a	.919	.909	.30193	2.015

a. Predictors: (Constant), PRP

b. Dependent Variable: DR

Table 3. Model Summary

Then, it is necessary to check if the residuals satisfy normal distribution. A P-P plot and a Q-Q plot for residuals are conducted. Since both plots show linear trends, a normal distribution of residuals can be guaranteed.

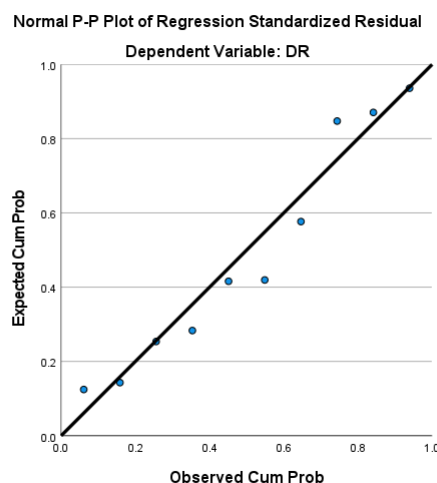


Figure 3. P-P Plot of Residuals

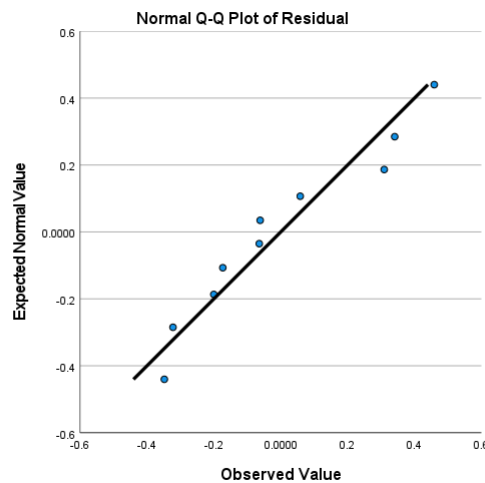


Figure 4. Q-Q Plot of Residuals

Shapiro-Wilk test is also used, a normality test appropriate and reliable for data with a sample size of less than 50, to further prove the normality of the distribution of residuals. The null hypothesis for this test is that the population from which the sample came does not differ significantly from the normal distribution. The alternative hypothesis for this test is that the population from which the sample came differs significantly from the normal distribution. Since the p-value is greater than 0.05, the null hypothesis is not rejected, meaning that the distribution of the residuals does not differ significantly from the normal distribution.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residual	.185	10	.200 [*]	.920	10	.357

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 4. Test of Normality of the Distribution of Residuals

After checking all the conditions, we conduct the hypothesis test for β . As shown in Table 4., the p-value of the coefficient of the permanent resident population is smaller than 0.001, meaning that the null hypothesis, $\beta = 0$, can be rejected.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.000	.095		-.005	.996
	PRP	-.958	.101	-.959	-9.525	<.001

a. Dependent Variable: DR

Table 5. The P-values of Coefficients

After modeling the relationship between the z-score of the permanent resident population and the z-score of the dependency ratio between 2010 and 2019 in Heilongjiang Province, the goal is to use current data to predict the permanent resident population abnormality test appropriate and reliable data). Time series is employed to achieve that goal. The ARIMA model is shown below:

Model Description

Model Type			
Model ID	PRP	Model_1	ARIMA(0,1,0)

Table 6. ARIMA Model Description

The significance of the model is less than 0.01, indicating that our model is appropriate and has a high degree of fit.

ARIMA Model Parameters

				Estimate	SE	t	Sig.
PRP-Model_1	PRP	No Transformation	Constant	-.327	.015	-21.616	<.001
			Difference	1			

Table 7. ARIMA Model Parameters

The predicted values of the permanent resident population for the year 2020 to 2025 in Heilongjiang province are shown below, both in table and graph:

Forecast

Model		2020	2021	2022	2023	2024	2025
PRP-Model_1	Forecast	-1.87	-2.20	-2.52	-2.85	-3.18	-3.50
	UCL	-1.77	-2.05	-2.34	-2.64	-2.94	-3.25
	LCL	-1.98	-2.35	-2.71	-3.06	-3.41	-3.76

For each model, forecasts start after the last non-missing in the range of the requested estimation period, and end at the last period for which non-missing values of all the predictors are available or at the end date of the requested forecast period, whichever is earlier.

Table 8. Predicted Z-score of Permanent Resident Population for the year 2020 to 2025

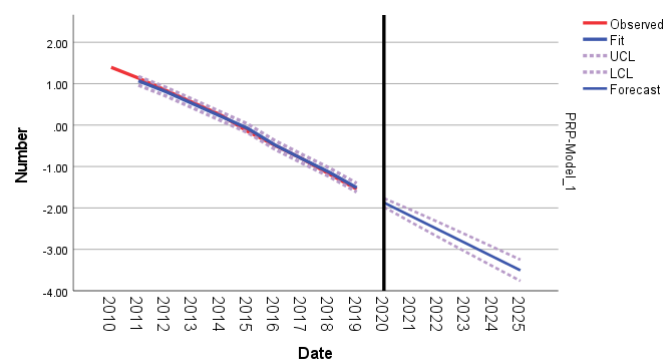


Figure 5. Observed and Predicted Z-score of Permanent Resident Population

Then, the predicted z-scores of the permanent resident population are plugged into the linear regression model to get the predicted z-score of the dependency ratio from 2020 to 2025, and then the z-scores of both variables are converted into values:

Year	Predicted Z-score of PRP	Predicted PRP (million)	Predicted Z-score of DR	Predicted DR (%)
2020	-1.87	31.90973663	1.79096	30.91187514
2021	-2.2	31.26098427	2.1071	31.53002563
2022	-2.52	30.63189107	2.41366	32.12944429
2023	-2.85	29.98313871	2.7298	32.74759479
2024	-3.18	29.33438635	3.04594	33.36574528
2025	-3.5	28.70529316	3.3525	33.96516394

Table 9. Predicted Z-scores and Predicted Values of Permanent Resident Population and Dependency Ratio

The real values of the permanent resident population from 2020 to 2022 of Heilongjiang province are available in the National Bureau of Statistics. The data provided and the predicted values calculated are used to solve for the percent error of the prediction:

Year	Predicted PRP (million)	Real PRP (million)	Percent Error
2020	31.90973663	31.71	-0.63%
2021	31.26098427	31.25	-0.04%
2022	30.63189107	30.99	1.17%

Table 10. Percent Error of Permanent Resident Population from 2020 to 2022

From the percent error shown in Table 10., it can be seen that since the percent errors are small, the prediction is accurate.

4 Result

The result of the study is that the slope of the linear relationship between the permanent resident population and the dependency ratio is significant ($\beta = -0.9535$). Therefore, this linear model can predict the dependency ratio with the permanent resident population from 2010 to 2019 in Heilongjiang province. Using time series, the permanent resident population and the dependency ratio from 2020 to 2025 in Heilongjiang province are predicted, yielding a minor error.

5 Discussion

Firstly, although the p-value of the coefficient of the independent variable, permanent resident population, is small enough to make β significant, it is still possible that there is another model that fits better when modeling the relationship between permanent resident population and dependency ratio of Heilongjiang province, given that the residual plot of the linear regression model shows a slightly curved pattern. Although not mentioned, a polynomial regression model for the two variables and the test to see if the coefficients of the polynomial regression model are significant are also conducted. However, all the p-values for the coefficients in the polynomial regression model are more significant than 0.05. In addition, in the polynomial regression model, two suspectable points, the year 2010 and the year 2011, deviate from the general pattern of a parabola. Until now, anything special in these two years that can lead to abnormal values has not been found. If these two points can be adequately explained, they can be excluded from the model, and a more fitted model may be generated.

Secondly, there are some potential topics that future studies can focus on. Although a significant linear relationship has been proved between the permanent resident population and the dependency ratio in Heilongjiang province, a causal relationship between the two variables cannot be concluded. Therefore, future studies can focus on whether the change in the permanent resident population causes the change in the dependency ratio or if a third variable is interacting between these two variables. The second topic that researchers can focus on is whether people overestimate the negative effect of the decline in the permanent resident population in Heilongjiang. Indeed, the decrease in the permanent resident population negatively predicts the dependency ratio in Heilongjiang province, meaning that the decrease in the permanent resident population increases the dependency burden of every labor force. However, given this increase in the dependency ratio, Heilongjiang province is still in the population profiteering stage, which may contrast with many people's thoughts. Another research question arises: How long can Heilongjiang Province maintain a population profiteering stage, given the current trend and decreased dependency ratio? Lastly, the permanent resident population change can be attributed to both changes in the natural and mechanical population growth rates. In this study, we did not include to what extent each contributor affects the dependency ratio in Heilongjiang province, which can be a research topic when finding the exact cause of the increase in the dependency ratio. Thus relevant departments of government can make more targeted policies.

6 Reference

Fenggang, L. (2021). Does population loss affect the economic growth of Northeast China? – Based on the calculation data of registered population loss in Northeast China. *Population and development*.

Nan, Zhang. (2016). The impact of population loss on regional economic development in Northeast China and its countermeasures. *Journal of Economic Research*, (17), pp. 63–64.