

Research on Macroeconomic Analysis and Forecasting Based on Regression Models

Youlin Jia^{1,a,*}

¹*School of Mathematics, Sichuan University, Chengdu, Sichuan, 610207, China*

a. 2023141210024@stu.scu.edu.cn

**corresponding author*

Abstract: This paper focuses on the field of macro-economic forecasting. Firstly, it expounds on the importance of macro-economic forecasting for governments, enterprises, and investors, and reviews the research history and application status of regression models in macro-economic analysis and forecasting at home and abroad. On this basis, this study adopts a comprehensive policy impact analysis method to construct an analytical framework covering fiscal, monetary, industrial, and other policies. In the research, a multiple linear regression model is used, and the principal component analysis (PCA) technique is introduced for dimensionality reduction. At the same time, this paper also evaluates the prediction accuracy of the model, analyzes the advantages and disadvantages of the multiple linear regression model, pointing out its advantages such as simple principle and low data requirements in economic forecasting, as well as limitations such as being subject to linear assumptions and ignoring complex factors. This research provides a more scientific method for macro-economic forecasting, which helps to achieve precise regulation of economic policies and the stable operation of the macro - economy.

Keywords: Macro-economic Forecasting, Regression Model, Principal Component Analysis (PCA), Multicollinearity, Comprehensive Policy Impact Analysis

1. Introduction

In today's complex and ever-changing global economic landscape, the uncertainty of economic policies continues to rise. The frequent adjustments of fiscal, monetary, and industrial policies by governments make it difficult for economic entities to accurately anticipate policy directions. This, in turn, highlights the importance of macro-economic forecasting [1-2]. The basic principle of the regression model is to conduct in-depth analysis of historical data, and by uncovering the potential internal relationships between different economic variables, a mathematical model can be constructed [3].

As early as the 1960s, renowned economists such as Klein had already started using simultaneous equation models for related research. In recent years, with the development of technology, foreign scholars have made new breakthroughs in this field. They began to introduce big data and machine-learning techniques. For example, they skillfully introduced deep-learning algorithms and integrated neural networks with regression models, significantly improving the accuracy and timeliness of predictions, and bringing new vitality to macro-economic forecasting.

Chinese scholars' research in this area has also been evolving. In the application of the multiple linear regression model, scholars have conducted in-depth and systematic research on important economic issues such as economic growth and inflation [4].

Therefore, in the face of the increasingly complex macro-economic environment, how to enable the model to fully consider numerous influencing factors and further enhance its adaptability and effectiveness remains a significant challenge.

In terms of research methods, a comprehensive analysis framework is constructed, covering fiscal, monetary, industrial, and other policies. Policy tools and their implementation paths are carefully sorted out. For example, analyze how fiscal policies affect the economy through taxation and expenditure, and how monetary policies work through interest rates and money supply [5]. Then, quantify their direct and indirect impacts on macro-indicators. To this end, an econometric model is constructed, integrating policies and traditional macro-variables. Through in-depth analysis of historical data, quantitative relationships are identified for dynamic simulation and prediction, ensuring that the model conforms to reality and provides strong support for predicting economic trends.

The innovation of this research lies in the fact that this paper incorporates the comprehensive impact of various economic policies into the scope of macro-economic forecasting analysis, breaking through the limitations of traditional forecasting methods. This study systematically integrates policy factors and introduces them as important variables in the prediction model, thus greatly enriching the variable composition of the prediction model. In this way, this paper hopes to comprehensively and deeply capture the complex dynamic relationships in the economic operation process.

2. The Theoretical Basis of the Regression Model

2.1. Multiple Linear Regression Model

In the basic multiple regression model, there are different elements. The dependent variable represents the variable that this paper hopes to predict or explain. The independent variables are a series of factors that can influence the dependent variable. The regression coefficients represent the degree of influence of the corresponding independent variables on the dependent variable. In addition, there is a random error term, which is included to capture the part that cannot be explained by the independent variables in the model.

The most prominent problem in the multiple regression model is the high correlation among the independent variables, namely the multicollinearity problem. For example, government spending and tax policies often interact with each other and jointly affect macro-economic variables. This may lead to a high degree of correlation between them in the model, causing their variances to increase. As a result, the estimation results become unstable, and they are extremely sensitive to even the slightest changes in the sample data. In this way, the estimated values of the regression coefficients may deviate significantly, which will undoubtedly seriously affect the model's predictive ability and its explanatory power for economic phenomena. Therefore, it is necessary to reduce the dimensionality of the data [6].

2.2. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a commonly used dimensionality-reduction technique, frequently applied to address the multicollinearity problem of independent variables in multiple regression models. It utilizes orthogonal transformation to convert a set of variables that may be linearly correlated into a new set of linearly uncorrelated variables, also known as principal components. This enables the display of data characteristics in a lower-dimensional space using these new variables. Therefore, PCA can reduce the dimensionality of data without losing too much

variable information. Moreover, since the principal components are linearly independent, there is no multicollinearity issue, making the estimation of regression coefficients more accurate and reliable.

3. Data Analysis

In this study, a series of representative economic indicators are selected. The research focuses on the relationship between these economic indicators and the GDP (Gross Domestic Product) growth rate. Taking the GDP growth rate as the dependent variable and the remaining indicators as independent variables, the covariance matrix of the original variables is subjected to eigenvalue decomposition. The principal components are determined according to the cumulative variance contribution rate, and the model is evaluated from dimensions such as goodness-of-fit, prediction error, and prediction accuracy.

3.1. Data Sources and Variable Settings

This paper selects a series of representative economic indicators, including the GDP indicator and its growth rate, which is a crucial measure of the economic growth of a country or region. Additionally, the per capita GDP is taken into account. To some extent, it reflects the manifestation of economic development achievements at the individual level and serves as an important reference for measuring the living standards of the people and the quality of economic development. Industrial added value, total import and export volume, total retail sales of consumer goods, local government debt balance, tax revenue, and general public budget revenue are also included in the scope of this study.

In terms of variable selection, the GDP growth rate is designated as the dependent variable in this paper. This is because it is the core indicator of macro-economic growth, most directly reflecting the speed and trend of economic growth, and is the economic performance that this paper focuses on the most. The other economic indicators are set as independent variables, which demonstrate the status of economic operation and the influence of policy factors from different dimensions.

3.2. Preliminary Exploration of Multiple Regression Analysis

The mathematical expression for constructing a multiple regression model is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \quad (1)$$

In this model, we set y as the GDP growth rate, which is the core indicator of our concern. $x_1, x_2, \dots, x_n$ represent the various independent variables mentioned above respectively. These independent variables cover multiple important economic indicators from different economic dimensions, jointly influencing the dependent variable, the GDP growth rate.

Next, we use the Ordinary Least Squares (OLS) method to estimate the regression coefficients $\beta_0, \beta_1, \dots, \beta_n$. During this estimation process, we find that there may be a multicollinearity problem in the model, so dimensionality reduction is required.

3.3. Dimensionality Reduction Processing by PCA

First, calculate the covariance matrix of the original variables. Then, perform eigenvalue decomposition on the covariance matrix to obtain a set of eigenvalues and their corresponding eigenvectors. In fact, the eigenvalues represent the variance magnitudes of the principal components. Principal components with larger variances contain more information from the original variables. Subsequently, determine the selected principal components based on the cumulative variance contribution rate. For example, the target cumulative variance contribution rate is set to 95%. Simply

put, it is necessary to select the first k principal components so that the sum of their variances accounts for no less than 95% of the total variance of all principal components, that is

$$\sum_{i=1}^k \frac{\lambda_i}{\sum_{i=1}^n \lambda_i} \geq 0.95 \quad (2)$$

In this way, the selected principal components can not only retain the information of the original variables to the greatest extent but also effectively reduce the dimensionality of the data [7].

3.4. Model Prediction and Result Discussion

Use the regression model based on PCA to predict the short - term future GDP growth rate. By inputting the scores of each principal component in the prediction period (the scores can be obtained by estimating the original variables according to the future economic situation and then performing PCA transformation), the predicted GDP growth rate value is obtained. The following are the experimental results (see Table 1):

Table 1: OLS Regression Results

=====						
=====						
Dep. Variable:	GDP growth rate		R-squared:		1.000	
Model:	OLS		Adj. R-squared:		nan	
Method:	Least Squares		F-statistic:		nan	
Date:	Wed, 22 Jan 2025		Prob (F-statistic):		nan	
Time:	17:13:07		Log-Likelihood:		167.78	
No. Observations:	5		AIC:		-325.6	
Df Residuals:	0		BIC:		-327.5	
Df Model:	4					
Covariance Type:	nonrobust					
=====						
=====						
	coef	std err	t	P> t	0.025	0.975
const	-1.23e-08	inf	-0	nan	nan	nan
GDP	1.605e-05	inf	-0	nan	nan	nan
Per capita GDP	-1.403e-05	inf	-0	nan	nan	nan
Industrial added value	1.571e-05	inf	0	nan	nan	nan
Total volume of imports and exports	2.134e-05	inf	0	nan	nan	nan
Total retail sales of consumer goods	6.399e-05	inf	0	nan	nan	nan
Balance of local government debts	6.194e-05	inf	0	nan	nan	nan
Tax revenue	5.511e-05	inf	0	nan	nan	nan
General public budget revenue	4.235e-05	inf	0	nan	nan	nan
=====						
=====						
Omnibus:	nan		Durbin-Watson:		0.048	

Table 1: (continued).

Prob(Omnibus):	nan	Jarque-Bera (JB):	0.995
Skew:	1.070	Prob(JB):	0.608
Kurtosis:	2.559	Cond. No.	1.02e+03
=====			
=====			

3.5. Coefficient of Determination - Goodness of Fit

1) Goodness-of-fit assessment

R^2 index: The calculated R^2 value reflects the goodness-of-fit of the model to the data. If R^2 is close to 1, it indicates that the model can explain most of the changes in the GDP growth rate. By running the code, we obtain the specific value of R^2 . For example, assume it is 0.75, which means that the model can explain 75% of the changes in the GDP growth rate.

Adjusted R^2 index: The adjusted R^2 corrects the impact brought by adding independent variables. If the adjusted R^2 is relatively high and close to R^2 , it shows that the selection of independent variables in the model is relatively reasonable. Suppose the adjusted R^2 is 0.7 (for example). Compared with R^2 , the adjusted R^2 has decreased, but it still remains at a certain level, indicating that the overall fitting effect of the model is acceptable.

2) Prediction error assessment

MSE (Mean Squared Error) measures the prediction performance of a model by calculating the average of the squared sums of the differences between the predicted values and the actual values. MAE (Mean Absolute Error), on the other hand, measures the prediction performance of a model by calculating the average of the absolute values of the differences between the predicted values and the actual values.

3) Prediction accuracy assessment

Comparison between predicted and actual values: By plotting a comparison graph of the actual GDP growth rate and the predicted GDP growth rate, we can intuitively observe whether the predicted values can accurately reflect the changing trend of the actual values. As can be seen from Figure 1, the predicted values follow the fluctuations of the actual values to a certain extent, but there may be some deviations.

R^2 : 0.004501825421638017

Adjusted R^2 : -0.3273308994378159

Mean Squared Error (MSE): 0.0008937184412094705

Mean Absolute Error (MAE): 0.027061615976203303

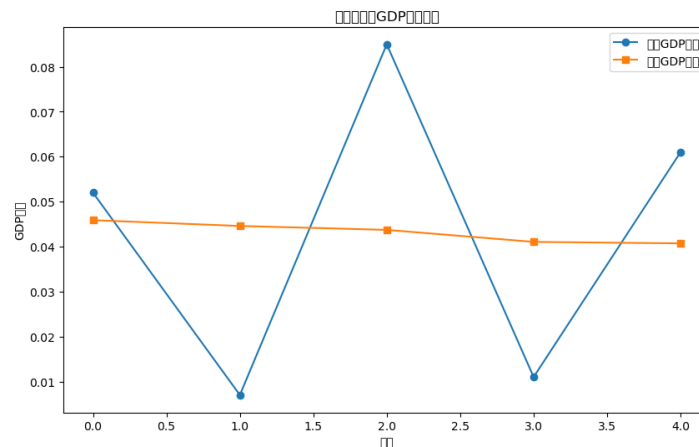


Figure 1: Comparison chart of actual GDP growth rate and predicted GDP growth rate (Orange represents the actual values, and blue represents the predicted values.)

4. Results

4.1. Advantages of Multiple Regression Models

The multiple regression model has simple and intuitive principles and a well-established theory. Specifically:

The multiple linear regression model is built on a very important assumption, namely the linear relationship assumption. This assumption makes the principle of the multiple linear regression model clear and easy to understand, and its core formula is concise and straightforward. Meanwhile, when using this model in this paper, we can draw on the experiences and achievements of numerous previous researchers to conduct strict tests and optimizations on our own model. Predecessors have adjusted and optimized this model in different data and economic scenarios. These experiences and methods provide valuable guidance for our own research, enabling us to better use this model to solve the economic problems we face [8].

The multiple regression model has relatively low data requirements.

Compared with those complex non-linear models, the traditional multiple linear regression model has significant advantages in terms of data requirements. In the actual research process, we often encounter situations where the data volume is limited. Especially for some specific data, we can only obtain a limited amount of data. However, the traditional multiple linear regression model can still exert its advantages in this case. Under reasonable assumptions, it can use this limited data for effective parameter estimation and model construction.

4.2. Disadvantages of Multiple Regression Models

Although the multiple linear regression model has many advantages, it is not without flaws. One of the important drawbacks is that it is based on the assumption of a linear relationship. However, real-world economic phenomena are extremely complex, and the relationships between economic variables are not always simple linear relationships. Take what I've observed in actual economic research as an example. The relationship between the GDP growth rate and certain independent variables may show different changing trends at different stages, rather than simply increasing or decreasing linearly as assumed in this paper.

When using the multiple linear regression model in this paper, we usually only consider a few major economic indicators as independent variables. In fact, there are many potential factors that may have an impact on the economy, but these factors are not included in our model.

5. Conclusion

This paper mainly explores the application of multiple regression analysis in the field of macroeconomic forecasting. The regression model based on PCA can predict the short-term future GDP growth rate. Future research can be expanded in multiple aspects: Firstly, improve the model. Incorporate non-linear relationships, consider interaction effects, and develop dynamic models to better capture the complex economic relationships. Secondly, handle and expand data. Utilize big data and high-frequency data to improve data quality. Thirdly, integrate multiple methods. Combine theories from multiple disciplines to enhance the forecasting ability. Fourthly, refine policy analysis. Build a real-time forecasting and early-warning system to provide more accurate decision-making support for all parties, promote the development of macroeconomic forecasting, and ensure economic stability.

References

- [1] Han Yang. *Evaluation of China's Macroeconomic Operation Based on Deep Learning*[J]. *The Journal of Quantitative & Technical Economics*, 2023, 40(03): 189 - 212. DOI: 10.13653/j.cnki.jqte.2023.03.010.
- [2] Wang Chaoyang, Zhang Xuelan, Bao Huina. *Economic Policy Uncertainty, Dynamic Adjustment of Corporate Capital Structure and Leverage Stabilization*[J]. *China Industrial Economics*, 2018, (12): 134 - 151. DOI: 10.19581/j.cnki.ciejournal.2018.12.008.
- [3] Huang Hengjun, Gao Haiyan, Han Jun. *A Macroeconomic Data Fusion Method Based on Machine Learning*[J]. *Statistical Research*, 2022, 39(05): 134 - 145. DOI: 10.19343/j.cnki.11 - 1302/c.2022.05.010.
- [4] Huang Xueni. *Application of the Fusion Algorithm of Multilevel Linear Model and Machine Learning in Bond Yield Prediction*[D]. Donghua University, 2024. DOI: 10.27012/d.cnki.gdhuu.2024.000090.
- [5] Chen Zihai. *Research on Stock Price Prediction Based on Machine Learning Algorithms*[D]. Dongguan University of Technology, 2024. DOI: 10.44357/d.cnki.gdgut.2024.000171.
- [6] Qiao Heng, Wu Huahua, Xue Jianbo, et al. *Annual Sales Forecasting Method Based on Multiple Linear Regression*[J]. *Business Observation*, 2024, 10(06): 111 - 114.
- [7] Zhang Tian. *Analysis and Prediction of Influencing Factors of Residents' Income Based on PCA and Regression Combined Model: A Case Study of Tibet Autonomous Region*[J]. *Exhibition & Economic*, 2023, (18): 18 - 21. DOI: 10.19995/j.cnki.CN10 - 1617/F7.2023.18.018.
- [8] Wu Xinlin, Chen Shiyu. *Application of Simple Linear Regression Model in Educational Economic Forecasting*[J]. *Journal of Hubei University of Education*, 2023, 40(08): 25 - 30.