MATHEMATICAL MODELING AND COMPUTING, Vol. 12, No. 2, pp. 425–436 (2025)

Forecasting CPI in Malaysia: Comparing linear regression, nonlinear regression, and nonlinear programming methods

Zulkifli F.[1], Abidin R. Z.[1], Deni M. S.[2], Zulkifli N. Z.[1,3]

[1] *College of Computing, Informatics, and Mathematics, University Teknologi MARA, Perak Branch, Tapah Campus, 35400 Tapah Road, Perak, Malaysia*

[2] *College of Computing, Informatics, and Mathematics, University Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*

[3] *Stratos Pinnacle Sdn Bhd. Signature 2, VO6-07-02 Lingkaran SV, Sunway Velocity Cheras, 55100 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia*

(Received 15 December 2024; Revised 25 April 2025; Accepted 8 May 2025)

This study explores factors influencing the Consumer Price Index (CPI) through an analysis of economic indicators and predictive models. It begins with normality testing and correlation analysis to identify significant variables, followed by model fitting using Linear Regression Model (LRM), Nonlinear Regression Model (NRM), and Nonlinear Programming (NLP). The results show strong positive correlations between CPI and variables like the Coincident Index, Labour, and Volume. Model comparisons indicate that NRM is the most effective predictor of CPI, with slightly lower Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) values than LRM and NLP. While NLP uses fewer variables, it may simplify model interpretation and reduce computational complexity. This research highlights the importance of accurate predictive models in CPI forecasting for evidence-based policymaking. A limitation is the small dataset, suggesting future studies could explore alternative models, use larger datasets, or conduct simulations to enhance CPI prediction accuracy.

Keywords: *consumer price index; linear regression model; nonlinear regression model; nonlinear programming methods.*

2010 MSC: 91B42, 62M20, 62P20, 90C30, 91B84 **DOI:** 10.23939/mmc2025.02.425

1. Introduction

The Consumer Price Index (CPI) is a vital economic indicator that reflects the average change in prices paid by consumers for a basket of goods and services over time. Accurate forecasting of CPI movements is crucial for policymakers, businesses, and households as it informs monetary policy decisions, budget planning, and investment strategies [1, 2]. In the context of Malaysia, a dynamic economy influenced by global trends and domestic factors, precise anticipation of CPI fluctuations is essential for ensuring economic stability and facilitating informed decision-making [3].

Although traditional regression methods have been widely used for CPI forecasting, they often struggle to capture the intricate dynamics of economic relationships. Linear regression, the most common approach, assumes a linear relationship between predictors and CPI, which may not fully account for the complex non-linearity present in economic data [4]. Consequently, there is an increasing need to explore alternative modeling techniques that can accommodate the non-linear and dynamic nature of economic variables.

Recent developments in econometric modeling have led to the emergence of nonlinear regression and linear programming approaches as promising alternatives for forecasting [5, 6]. Nonlinear regression

This research was supported by the PhD Graduate MyRA Grant, referenced under RMC File No. 600-RMC/GPM LPHD 5/3 (096/2022).

techniques offer flexibility in modeling non-linear relationships between predictors and CPI, potentially improving forecast accuracy [7]. On the other hand, linear and nonlinear programming (NLP) provides a systematic framework for optimizing regression coefficients while incorporating domain-specific constraints and preferences [8].

Previous studies have highlighted the limitations of conventional regression methodologies in predicting CPI movements and have emphasized the need for alternative approaches to address these shortcomings. For instance, Sek [9] and Sek et al. [10] demonstrated the effectiveness of nonlinear regression model (NRM) in capturing the complexities of nonlinear relationships that exist within inflationary trends, leading to more accurate forecasts. However, the integration of NLP, which incorporates optimization-focused methodologies into the regression framework, to enhance the reliability and applicability of CPI forecasts remains largely neglected in current literature. The distinctions between nonlinear regression and NLP methodologies will be thoroughly examined in the methodology section of this study.

In this study, our aim is to build upon these insights by conducting a comparative analysis of linear regression, nonlinear regression, and NLP approaches for CPI forecasting in Malaysia. By evaluating the performance of these methods using historical CPI data, we seek to assess their respective strengths and weaknesses in capturing CPI trends and fluctuations. Specifically, we will examine the predictive accuracy, interpretability, and practical relevance of each approach, with the goal of identifying the most effective modelling technique for CPI forecasting in the Malaysian context.

2. Literature review

Forecasting plays a crucial role in economic analysis and policymaking, influencing decisions in monetary policy, fiscal planning, and investment strategies [11, 12]. Extensive research has been conducted on regression methods to predict CPI movements, with the aim of achieving accurate forecasts in the face of economic complexities.

Regression techniques are commonly employed in CPI forecasting, often utilizing linear models to establish relationships between economic indicators and CPI fluctuations [13, 14]. For instance, Karadağ [14] conducted a study to analyze the relationship between the industrial production index, Brent crude oil prices, and CPI in the Turkish economy. The study utilized quarterly data from 2010: Q1 to 2020: Q4. Conversely, Liang [15] utilized a multiple linear regression model to predict CPI in the context of an epidemic. The model included factors such as money supply and total social retail goods. However, these studies often overlook the inherent non-linearity and interactions present in economic data.

Critics have highlighted the limitations of traditional regression approaches in capturing economic complexity [16]. Assumptions of linearity between predictors and CPI may result in suboptimal forecasts, particularly during times of economic uncertainty or structural shifts [17]. Moreover, traditional models may struggle to incorporate complex predictor interactions, potentially missing significant data patterns [18].

There is a growing interest in NLP techniques for economic forecasting, which offers the promise of addressing the limitations of regression methods [19]. NLP, such as optimization-based regression, allows for the incorporation of domain-specific constraints and preferences. Initial studies have shown promising results. Mathelinea et al. [8] applied nonlinear optimization to optimize regression coefficients for forecasting the optimum order quantity while minimizing total inventory costs to fulfill their research objectives. Their study indicates that public hospitals in Malaysia could benefit from employing forecasting techniques to determine the optimal inventory quantity for drugs. Furthermore, they suggest implementing NLP with budget constraints to achieve the optimal total annual inventory cost.

The absence of research that integrates NLP techniques into CPI forecasting presents a substantial barrier to the advancement of the literature review in this domain. The lack of such studies restricts the range and depth of available insights, limiting the ability to thoroughly analyze and evaluate the effectiveness of NLP methodologies in forecasting CPI trends. As a result, the exploration of alternative approaches and the identification of best practices in CPI forecasting are impeded, thus hindering the progress of knowledge, and understanding in this critical field of study.

3. Methodology

3.1. Dataset description

The dataset used in this study consists of monthly economic indicators and CPI data from January 2018 to February 2024. The dataset includes variables related to CPI, as well as other relevant economic indicators sourced from reputable institutions such as national statistical agencies, central banks, and international organizations.

The scope of this study does not warrant the partitioning of the data into training and testing sets. This decision is influenced by the relatively small sample size of 74 observations. Dividing the data could yield subsets that lack sufficient samples for meaningful analysis or model training. Thus, opting to utilize the complete dataset enables the exploitation of all available data for model development and evaluation, potentially enhancing the model's overall performance and reliability.

The details of the variables utilized in this study are outlined as follows:

a) Consumer Price Index (CPI): The primary dependent variable of interest, representing the average change in prices paid by consumers for a basket of goods and services over time. CPI data is collected and reported monthly.
b) Producer Price Index (PPI): A key independent variable that reflects changes in the prices received by domestic producers for their goods and services. PPI data is collected and reported monthly and serves as a proxy for upstream inflationary pressures.
c) Youth Unemployment Rate: The number of individuals aged 15–24 who are currently seeking employment or available to work but are not employed. This variable provides insights into labour market dynamics and potential demographic factors influencing CPI trends.
d) Coincident Index: A comprehensive measure of current economic performance, capturing various indicators such as industrial production, employment levels, income, and sales. The Coincident Index serves as a proxy for overall economic activity and may influence CPI trends.
e) Labour Force Participation Rate: The number of employed and unemployed individuals actively participating in the labour force, expressed in thousands. This variable provides insights into the overall size and composition of the labour force, which can impact CPI dynamics through its influence on consumer spending and income levels.
f) Quantity of Items Sold Index: An index representing the quantity of goods and services sold, with a base year of 2015 – 100. This variable reflects changes in consumer demand and economic activity, which can affect CPI trends.

3.2. Traditional regression model

Linear regression, a conventional method, is widely applied to model the association between a dependent variable (Y) and one or more independent variables (X) and becomes traditional regression. It assumes a linear relationship between the independent variables and the dependent variable. In linear regression, the model is expressed as equation (1),

$$Y = \beta_0 + \beta_1 X + \varepsilon, \tag{1}$$

where Y is the dependent variable (e.g., CPI), X is the independent variable (e.g., PPI, Youth Unemployment Rate, Coincident Index, Labour Force Participation Rate, Quantity of Items Sold Index), β_0 is the intercept, representing the value of Y when X is zero, β_1 is the slope coefficient, representing the change in Y for a one-unit change in X, ε is the error term, representing the difference between the observed value of Y and the value predicted by the model.

The parameters β_0 and β_1 are estimated using the method of ordinary least squares (OLS), which minimizes the sum of squared differences between the observed and predicted values of Y. The formulas for estimating these parameters are as follows:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}, \tag{2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}, \tag{3}$$

where $\hat{\beta}_1$ is the estimated slope coefficient, $\hat{\beta}_0$ is the estimated intercept, X_i and Y_i are the observed values of the independent and dependent variables, respectively, $\bar{X}$ and $\bar{Y}$ are the sample means of the independent and dependent variables, respectively, n is the number of observations.

3.3. Nonlinear regression model (NRM)

Nonlinear regression is a flexible method for modeling relationships between variables that are not linear. It allows for more complex functional forms by incorporating nonlinear terms or transformations of the independent variables. The general form of a NRM is given by

$$Y = f(X, \beta) + \varepsilon, \tag{4}$$

where Y is the dependent variable (CPI), X is the vector of independent variables, β is the vector of parameters to be estimated, $f(\cdot)$ is the nonlinear function relating X to Y, ε is the error term.

The parameters β are estimated using nonlinear least squares (NLS), which minimizes the sum of squared differences between the observed and predicted values of Y.

The general formula for estimating parameters in nonlinear regression is

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \left(Y_i - f(X_i, \beta)\right)^2, \tag{5}$$

where $\hat{\beta}$ is the vector of estimated parameters, $f(X_i, \beta)$ is the predicted value of the dependent variable Y_i based on the independent variable X_i and the parameter vector β, The objective is to minimize the sum of squared differences between the observed and predicted values of Y_i.

The estimation of parameters in nonlinear regression typically requires iterative optimization algorithms, such as the Gauss–Newton method or the Levenberg–Marquardt algorithm, to find the values of β that minimize the objective function.

In summary of the general regression procedure, the steps for estimating NRM are as follows [20]:

1: Select a functional form for the regression equation, ensuring that the first partial derivatives of the dependent variable with respect to the unknown parameters exist and are continuous across the relevant domain.

2: Choose initial values for each unknown parameter. Ideally, these initial estimates should closely approximate the true values based on the first-order terms of Taylor expansion. However, in practice, they may deviate significantly. Initial estimates can be informed by prior experience with similar problems, examination of scatter plots of observations, means of observed variables, and theoretical insights. Sometimes, simple initial guesses such as zeros or ones suffice for the converging iterations.

3: Compute the first partial derivatives of the dependent variable and evaluate the calculated values (Y) using the parameter estimates determined in Step 2. Perform this computation for each observation of Y and X_i.

4: Utilize ordinary linear regression methods to identify improvements in the initial estimates derived in Step 2, which most effectively account for the unexplained residuals ($Y_{\text{observed}} - Y_{\text{calculated}}$).

5: Calculate the refined parameter estimates by incorporating adjustment factors derived from Step 4 into the initial estimates chosen in Step 2.

6: Iterate through this procedure by treating the refined estimates from Step 5 as new initial guesses in Step 2. Continue iterating until the adjustment factors obtained in Step 4 approach zero sufficiently to no longer influence the significant digits of the parameter estimates obtained in Step 5.

3.4. Nonlinear programming (NLP) procedure

Edwards [20] outlined two challenges in implementing of nonlinear regression. Firstly, it is crucial to have initial parameter estimates for nonlinear regression problems that are sufficiently close to the true values. This ensures that the first-order terms of the Taylor expansion dominate over higher-order terms. Failure to achieve this proximity may result in iterations diverging, sometimes necessitating a trial-and-error approach, and relying on fortuitous outcomes to establish a converging iteration sequence.

Secondly, a distinct set of machine instructions, comprising algebraic expressions of the first derivatives, must be developed for each specific functional form utilized in regression or programming. While this may lead to multiple sets of machine instructions and tapes, it also enables varied improvements in research outcomes for problems where nonlinear analysis is deemed superior to linear analysis.

As an alternative, the NLP procedure has been suggested to address issues related to iterations. In general, the NLP procedure aims to identify a local maximum (or minimum) of any differentiable objective function, while considering linear constraints. The approach for finding global maxima for functions with two or more local maxima would depend on the specific forms of the objective functions utilized.

This study will use customized NLP to optimize inflation forecasting in Malaysia. The NLP framework will be tailored to optimize CPI by incorporating relevant economic indicators that influence consumer prices. The primary objective of using NLP is to minimize forecast errors and improve the reliability of CPI predictions. To achieve NLP optimization, the Generalized Reduced Gradient (GRG) nonlinear method will be employed.

The utilization of the GRG nonlinear method follows a systematic approach designed to effectively address the challenges of NLP. In general, the process begins with formulating the problem, precisely defining the objective function to be minimized or maximized, along with any constraints. Next, an initial feasible solution is determined, often using heuristic methodologies, or building upon previous solutions. Once the initial solution is established, the problem is transformed into standard form by replacing inequalities within the constraints with equalities and introducing slack variables to facilitate this conversion.

The following provides a step-by-step procedure for applying the GRG nonlinear method to estimate parameters in a regression model for optimizing CPI forecasting:

1: Formulate the objective function to minimize the sum of squared differences between the observed CPI values and the predicted values generated by the regression model. The objective function can be expressed as

$$\text{Minimize} \sum_{i=1}^{n} (Y_i - f(X_i, \beta))^2 \tag{6}$$

2: Specify constraints by identifying the relationship between variables in the regression model. For example, if it is known that certain independent variables have a positive or negative relationship with CPI, constraints can be specified to enforce these relationships during parameter estimation. The Pearson correlation coefficient can fulfill this purpose, while normality will be assessed using the Shapiro-Wilk Test, Histogram plot, and QQ-Plot.

3: Select initial estimates for the parameters β that satisfy the defined constraints and serve as the starting point for parameter estimation. These initial estimates can be determined using heuristic methods or by leveraging previous solutions from similar regression models.

4: Transform the problem into standard form by converting any inequality constraints into equality constraints and introducing slack variables as needed. Ensure that the optimization problem is formulated in a manner suitable for solving using the GRG algorithm.

5: Utilize the GRG algorithm to iteratively optimize the objective function and estimate the parameters β in the regression model. At each iteration, the algorithm adjusts the parameter estimates to minimize the sum of squared differences between the observed and predicted CPI values.

6: Evaluate the estimated parameters β to assess their accuracy and reliability in capturing the relationship between the independent variables and CPI. Validate the regression model's predictive performance by comparing the predicted CPI values with historical CPI data. Analyze any discrepancies between predicted and observed CPI values.

7: Fine-tune the estimated parameters β based on insights gained from the evaluation and validation phase to improve the accuracy of CPI forecasts. Modify the optimization strategy as needed to address any shortcomings identified during the evaluation process. This may involve updating initial parameter estimates or adjusting constraints.

3.5. Comparative analysis of CPI forecasting methods: Linear regression, nonlinear regression, and NLP approaches

To evaluate the forecasting performance of Linear Regression, Nonlinear Regression, and NLP approaches for predicting the CPI in Malaysia, we will conduct a comprehensive comparative analysis. This analysis will include quantitative metrics and graphical methods.

We will begin by selecting key metrics to evaluate the forecasting performance. These metrics include the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). By utilizing these metrics, we can gain insights into the accuracy and directionality of forecast errors. This will enable a thorough comparison of the different methods.

The formulas for calculating these metrics are as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|, \tag{6}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}, \tag{7}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \times 100, \tag{8}$$

where Y_i is the observed value and $\hat{Y}_i$ is the predicted value for the i-th observation.

These metrics will be calculated for each method using historical CPI data and the corresponding predicted values.

In addition to quantitative metrics, we will employ graphical methods to visually compare the forecast results. Scatter plots or time series plots will be used to illustrate the alignment between predicted and observed CPI values for each method. These visualizations will provide insights into the accuracy and precision of the forecasts, aiding in the comparative analysis. All analytical procedures will be conducted solely using the R software platform.

4. Results and discussion

4.1. Normality test and correlation analysis

The normality test is essential for conducting correlation analysis, and the results of the correlation analysis are used to establish constraints in NLP.

Table 1. Shapiro–Wilk test results for normality.

Variables	Statistic	p-value
CPI	0.886637	7.23E-06
PPI	0.914382	9.79E-05
Coincident Index	0.919574	0.000167
Unemployed	0.986563	0.62382
Labour	0.94533	0.003019
Volume	0.927101	0.000371

The results of the Shapiro–Wilk test presented in Table 1 demonstrate that the CPI, PPI, Coincident Index, Labour, and Volume variables show test statistics that are close to 1, indicating distributions that are approximately normal. However, the associated p-values for these variables are all significantly small (less than 0.05), providing strong evidence against the null hypothesis of normality. Therefore, it can be concluded that these variables do not conform to a normal distribution.

On the other hand, the Unemployed variable exhibits a Shapiro–Wilk test statistic close to 1, suggesting a distribution that is also approximately normal. However, its p-value is relatively large (greater than 0.05), implying weaker evidence against the null hypothesis of normality. Consequently, it is determined that the Unemployed variable can be considered to have an approximately normal distribution.

The histograms depicted in Figure 1 illustrate the distributions of the variables CPI, PPI, Coincident Index, Unemployed, Labour, and Volume. Each histogram displays a bell-shaped curve, indicating that the data for each variable is approximately normally distributed. Furthermore, the absence of significant outliers or skewness suggests that the data conforms well to the expectations of a normal distribution.

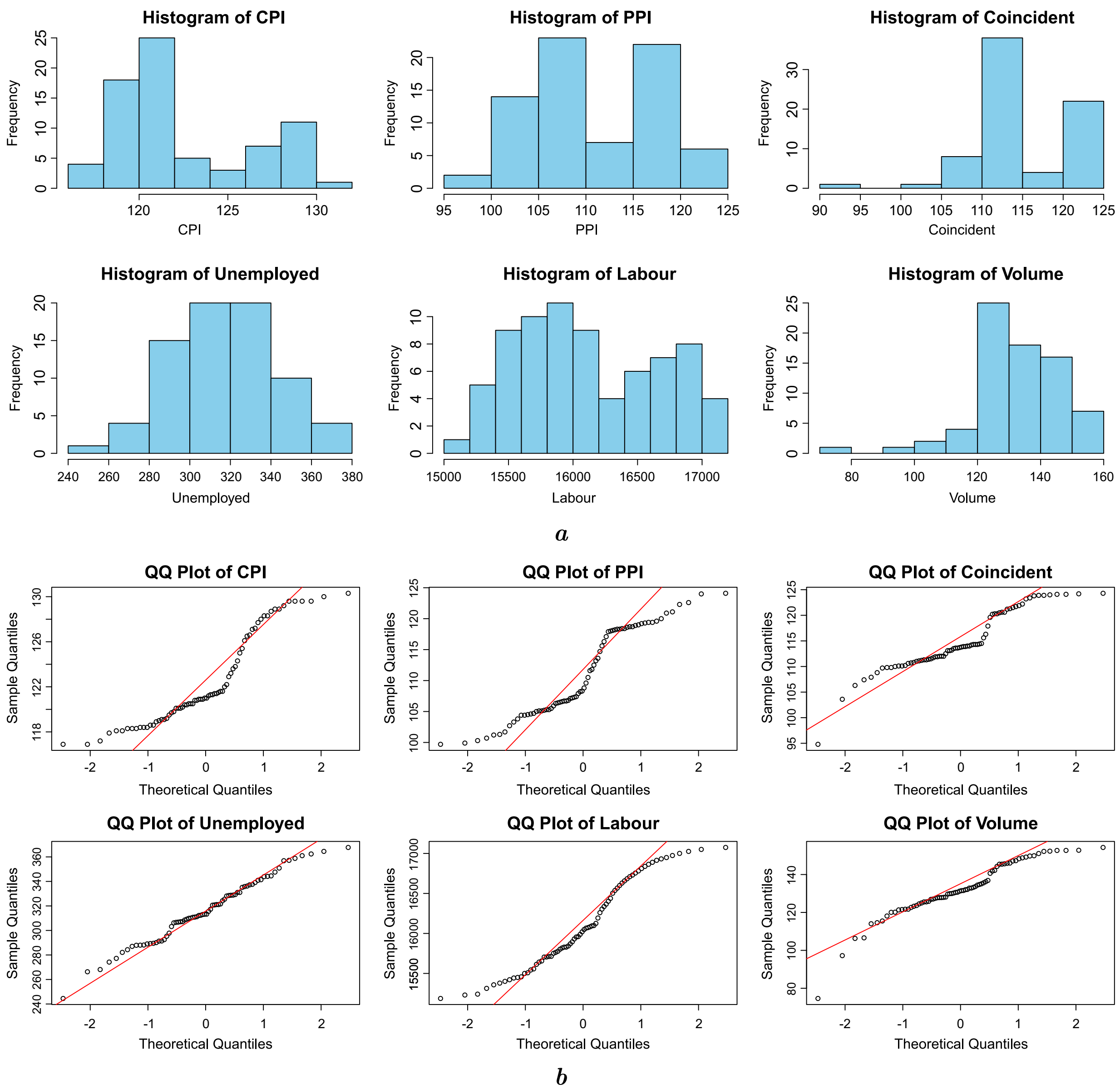

Fig. 1. (*a*) Histogram and (*b*) QQ plot of the variables.

In the QQ plots, the points for each variable closely adhere to the diagonal line, indicating that the quantiles of the observed data align well with those of a theoretical normal distribution. This alignment further substantiates the conclusion drawn from the histograms that the data demonstrates good normality.

The correlation results presented in Table 2 provide valuable insights into the relationships between the dependent variable, CPI, and the independent variables. The PPI exhibits a strong positive correlation of 0.797 with CPI, indicating a robust linear relationship between these two factors. This correlation is statistically significant, as evidenced by the p-value of 2.04E-17. Similarly, the Coincident Index displays a notably strong positive correlation of 0.848 with CPI, suggesting a significant linear association. The low p-value of 1.59E-21 further reinforces the statistical significance of this correlation.

Table 2. Correlation coefficients and statistical significance.

Variables	Correlation Coefficient	p-value
PPI	7.97E-01	2.04E-17
Coincident Index	8.48E-01	1.59E-21
Unemployed	0.28430608	0.014091
Labour	9.72E-01	2.91E-47
Volume	7.55E-01	8.28E-15

Moreover, the Labour variable demonstrates an exceptionally strong positive correlation of 0.972 with CPI, indicating a remarkably robust linear relationship. The remarkably low p-value of 2.91E-47 further confirms the statistical significance of this correlation. On the other hand, the Unemployed variable shows a more moderate positive correlation of 0.284 with CPI, although it remains statistically significant at the 0.05 significance level (p-value = 0.014091). Lastly, the Volume variable indicates a strong positive correlation of 0.755 with CPI, suggesting a notable linear association. The statistically significant p-value of 8.28E-15 emphasizes the reliability of this correlation.

Given that all correlation coefficients between CPI and the independent variables are positive, it is necessary to ensure that the optimization parameters in NLP are constrained to be positive. This constraint aligns with the observed positive relationships in the data.

4.2. Model fitting

The results of model fitting provide insights into the predictive performance and significant variables influencing the CPI, laying the foundation for a comprehensive discussion on the economic factors impacting consumer price dynamics.

4.2.1. Linear regression model (LRM)

Table 3. Regression coefficients and statistical significance for the LRM.

Variables	Estimate Coefficients	Standard Error	t-statistic	p-value
(Intercept)	2.5426272	4.6478767	0.547	0.5861
PPI	-0.0369543	0.0315132	-1.173	0.245
Coincident Index	0.1949702	0.0854999	2.28	0.0257
Unemployed	-0.0104817	0.0054167	-1.935	0.0571
Labour	0.0069621	0.0004275	16.287	<2e-16
Volume	-0.0548844	0.0254694	-2.155	0.0347

The presented results represent the output of a model fit, particularly for a linear regression analysis. In Table 3, the coefficients estimated for each independent variable are provided alongside their respective standard errors, t-statistics, and associated p-values. The intercept coefficient is estimated at 2.5426 with a standard error of 4.6479, yielding a t-statistic of 0.547 and a p-value of 0.5861, suggesting no significant effect. For the independent variable PPI, the estimated coefficient is -0.0369 with a standard error of 0.0315, resulting in a t-statistic of -1.173 and a p-value of 0.245, indicating no statistically significant relationship.

Conversely, the Coincident Index exhibits a positive relationship with a coefficient estimate of 0.195, a standard error of 0.0855, and a statistically significant t-statistic of 2.280, corresponding to a p-value of 0.0257. Similarly, the Labour variable demonstrates a highly significant positive relationship, with a coefficient estimate of 0.00696, a standard error of 0.00043, and an extremely significant t-statistic of 16.287, with a p-value below 2e-16.

Next, the Unemployed variable shows a negative relationship with a coefficient estimate of -0.01048, a standard error of 0.00542, and a marginally significant t-statistic of -1.935, with a p-value of 0.0571. Lastly, the Volume variable displays a negative relationship with a coefficient estimate of -0.05488, a standard error of 0.02547, and a statistically significant t-statistic of -2.155, corresponding to a p-value of 0.0347. The most suitable linear regression equation incorporates the statistically significant coefficients for Coincident Index, Labour, and Volume.

4.2.2. Nonlinear regression model (NRM)

The results from the NRM, as shown in Table 4, highlight significant coefficients, particularly for the variables Coincident Index, Labour, and Volume, with respective p-values of 0.001768, $<$ 2e-16, and 0.000728. These findings suggest that variations in the variables significantly influence the dependent variable. Specifically, a unit increase in the Coincident Index corresponds to a 0.2227 increase in the dependent variable, while a unit increase in Labour results in a 0.0070 increase. Conversely, a unit increase in Volume leads to a decrease of 0.0646 in the dependent variable. These insights underscore

Table 4. Regression coefficients and statistical significance for the NRM.

Variables	Estimate Coefficients	Standard Error	t-statistic	p-value
PPI	-0.0455926	0.0271326	-1.68	0.097409
Coincident Index	0.2227278	0.0684655	3.253	0.001768
Unemployed	-0.0099073	0.0052869	-1.874	0.065176
Labour	0.0070495	0.0003945	17.871	$< 2e{-16}$
Volume	-0.0645512	0.0182488	-3.537	0.000728

the importance of these variables in predicting the outcome and advocate for their inclusion in the NRM to enhance predictive accuracy.

4.2.3. Nonlinear programming (NLP)

The optimal parameter estimates for the variables in the NLP model, as shown in Table 5, indicate that only the Coincident Index and Labour variables have non-zero estimates, with values of 0.042023807 and 0.007299049, respectively. This suggests that these variables significantly contribute to the model's prediction, while the other variables, PPI, Unemployed, and Volume, have estimated parameters close to zero, indicating minimal impact on the model's output.

Table 5. Optimal parameter estimates for variables in NLP.

Variables	Optimal Parameters Estimate
PPI	0.00
Coincident Index	0.042023807
Unemployed	0.00
Labour	0.007299049
Volume	0.00

The discrepancy in coefficient estimates for the Coincident Index between the NLP method and the LRM and NRM could be due to the differing methodologies employed by these models in handling variable contributions. The NLP method may impose specific constraints or optimize the objective function in a way that differs from the approaches used by LRM and NRM, resulting in variations in the estimated coefficients. Specifically, the NLP approach aims to optimize a predefined objective function and might treat the Coincident Index differently, affecting its impact on the model's predictions. Additionally, the reduced number of variables in the NLP model could influence how sensitively the model responds to each variable's contribution. This variation underscores the importance of understanding the distinct ways in which each modeling approach integrates and processes predictor variables.

4.3. Model comparison

The models have experienced another round of model fitting, this time using only the significant variables. As a result, the following model equations have been derived:

1. LRM: $\hat{\text{CPI}} = 5.9795544 + 0.2008670 \cdot \text{Coincident Index} + 0.0061880 \cdot \text{Labour}$.
2. NLM: $\hat{\text{CPI}} = 0.2527268 \cdot \text{Coincident Index} + 0.0063817 \cdot \text{Labour}$.
3. NLP: $\hat{\text{CPI}} = 0.042024165 \cdot \text{Coincident Index} + 0.007299046 \cdot \text{Labour}$.

Table 6. Comparison of model performance metrics.

Model	MAE	RMSE	MAPE (%)
LRM	0.7000469	0.8704636	0.5705954
NRM	0.6929086	0.8828365	0.566058
NLP	0.6981735	0.9638452	0.5742656

Following an examination of the performance metrics detailed in Table 6, the NRM emerges as the most effective predictor of the CPI, showcasing marginally lower MAE and RMSE values compared to both the LRM and the NLP methods. Additionally, the NRM exhibits a slightly reduced MAPE, indicative of enhanced accuracy in forecasting the CPI. Therefore, the NRM stands out as the preferred model in terms of predictive capability among the three approaches evaluated. Conversely, while the NLP method demonstrates slightly superior performance in MAE and RMSE compared to the LRM, the disparities in performance metrics between the two models remain relatively minor. Despite the NLP method displaying a slightly higher MAPE than the LRM, it maintains competitive performance. Thus, the selection between the LRM and NLP may hinge on additional considerations such as model intricacy and computational efficiency.

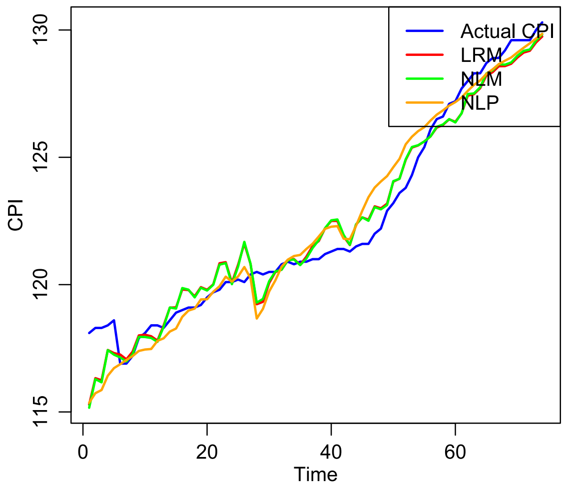

Fig. 2. Plot the comparison of actual CPI and predicted CPI values.

In addition, the reduced number of variables in the NLP method compared to the NRM may suggest a simpler model structure. While fewer variables can simplify model interpretation and computational demands, it is essential to assess whether this reduction impacts the model's forecasting accuracy. In real-world forecasting scenarios, the choice between models often involves striking a balance between model simplicity and predictive performance. Therefore, evaluating the NLP method's forecasting accuracy relative to its reduced variable set becomes crucial in determining its suitability for practical application.

Figure 2 presents a visual comparison of the actual CPI values against the predicted CPI values generated by different forecasting models. The plot reveals the fluctuations and trends in the actual CPI values over time, providing a baseline for assessing the accuracy of the predictions. Overlaying the actual CPI trend are multiple lines representing the predicted CPI values produced by various forecasting models, such as LRM, NRM, and NLP. By scrutinizing the alignment between the predicted and actual CPI values, we can gauge the accuracy and reliability of each model. Discrepancies between the predicted and actual trends may signify areas where the models excel or require refinement. Moreover, the plot extends into the future, allowing us to visualize how each model forecasts CPI trends beyond the available historical data. This comprehensive comparison aids in evaluating the forecasting performance of each model and informs decision-making for future forecasting endeavours.

5. Conclusion

In conclusion, this study has conducted a comprehensive examination of the factors influencing the CPI, utilizing various analytical methods to elucidate its dynamics. Through rigorous normality testing, correlation analysis, and model fitting, significant predictors of CPI fluctuations, including the Coincident Index, Labour, and Volume, have been identified. These findings have profound implications for economic forecasting and policymaking, offering valuable insights into the interplay between economic indicators and inflationary trends.

Despite the robustness of the analysis, it is important to acknowledge certain limitations. The study's reliance on a relatively small dataset may constrain the generalizability of the findings, and further improvements in model accuracy could be achieved by incorporating additional variables or utilizing larger datasets [21]. Moreover, while efforts were made to ensure the reliability of the models, there may be unaccounted-for factors that could influence CPI dynamics, warranting further investigation [22].

Future research could focus on addressing these limitations by exploring alternative modeling techniques, leveraging larger datasets, or conducting simulations to enhance the robustness of the findings. Additionally, investigating the impact of external factors, such as geopolitical events or technological advancements, could provide a more nuanced understanding of CPI fluctuations and improve the accuracy of forecasting models.

[1] Hajargasht G. Reliability of Ideal Indexes. Preprint arXiv:2210.13684 (2022).

[2] Zulkifli F., Ismail I. L., Chek M. Z. A., Jamal N. F., Ridzwan A. N. A. A., Jelas I. Md., Jelas I. M., Noor S. I. M., Ahmad A. B. Time series forecasting of future claims amount of SOCSO's Employment Injury Scheme (EIS). AIP Conference Proceedings. **1482**, 396–401 (2012).

[3] Konarasinghe K. M. U. B. Modeling Consumer Price Index of Malaysia: Application of Exponential Smoothers. Journal of New Frontiers in Education and Social Sciences. **2** (1), 16–33 (2022).

[4] Hauzenberger N., Huber F., Klieber K. Real-time inflation forecasting using non-linear dimension reduction techniques. International Journal of Forecasting. **39** (2), 901–921 (2023).

[5] Kuranga C., Pillay N. A comparative study of nonlinear regression and autoregressive techniques in hybrid with particle swarm optimization for time-series forecasting. Expert Systems with Applications. **190**, 116163 (2022).

[6] Pavlicko M., Vojteková M., Blažeková O. Forecasting of Electrical Energy Consumption in Slovakia. Mathematics. **10** (4), 577 (2022).

[7] Wang J. C., Holan S. H. Bayesian multi-regime smooth transition regression with ordered categorical variables. Computational Statistics & Data Analysis. **56** (12), 4165–4179 (2012).

[8] Mathelinea D., Chandrashekar R., Omar N. F. A. C. Inventory cost optimization through nonlinear programming with constraint and forecasting techniques. AIP Conference Proceedings. **2184** (1), 040011 (2019).

[9] Sek S. K. A new look at asymmetric effect of oil price changes on inflation: Evidence from Malaysia. Energy & Environment. **34** (5), 1524–1547 (2022).

[10] Sek S. K., Sim K. Y., Har W. M., Aric K. H. Examination on the asymmetric effects of commodity price changes on sectoral CPI inflation of Malaysia. AIP Conference Proceedings. **2500** (1), 020023 (2023).

[11] Chek M. Z. A., Ahmad A. B., Ridzwan A. N. A. A., Jelas I. Md., Jamal N. F., Ismail I. L., Zulkifli F., Noor S. I. M. Univariate time series modeling and an application to future claims amount in SOCSO's invalidity pension scheme. AIP Conference Proceedings. **1482** (1), 392–395 (2012).

[12] Marrakchi N., Bergam A., Fakhouri H., Kenza K. A hybrid model for predicting air quality combining Holt–Winters and Deep Learning Approaches: A novel method to identify ozone concentration peaks. Mathematical Modeling and Computing. **10** (4), 1154–1163 (2023).

[13] Liang J. Multivariate linear regression method based on SPSS analysis of influencing factors of CPI during epidemic situation. 2020 2nd International Conference on Economic Management and Model Engineering (ICEMME). 294–297 (2020).

[14] Karadağ H. The Relationship Between Industrial Production Index, Oil Prices and Consumer Price Index in the Turkish Economy. Journal of Economic Policy Researches. **8** (2), 211–223 (2021).

[15] Liang J. Multivariate linear regression method based on SPSS analysis of influencing factors of CPI during epidemic situation. 2020 2nd International Conference on Economic Management and Model Engineering (ICEMME). 294–297 (2020).

[16] Lmakri A., Akharif A., Mellouk A. Estimation in short-panel data models with bilinear errors. Mathematical Modeling and Computing. **10** (3), 682–692 (2023).

[17] Yusuff A. Q., Ajayi S. O., Akanbi O. A., Amusa S. O. Econometric Analysis of Consumer Price Index on Some Major Economic Indicators. International Journal of Applied Science and Mathematical Theory. **6** (3), 19–25 (2020).

[18] Steyerberg E. W. Overfitting and Optimism in Prediction Models. In: Clinical Prediction Models. Statistics for Biology and Health. Springer, Cham. 95–112 (2019).

[19] Ibeh C. V., Asuzu O. F., Olorunsogo T., Elufioye O. A., Nduubuisi N. L., Daraojimba A. I. Business analytics and decision science: A review of techniques in strategic business decision making. World Journal of Advanced Research and Reviews. **21** (02), 1761–1769 (2024).

[20] Edwards C. Non-Linear Programming and Non-Linear Regression Procedures. Journal of Farm Economics. **44** (1), 100–114 (1962).

[21] Zulkifli F., Abidin Z. R., Deni S. M. A New Method for Calculating Consumer Price Indices: Incorporating Consumer Perceptions and Attitudes with Item Response Theory. International Journal of Academic Research in Economics and Management and Sciences. **12** (1), 419–428 (2023).

[22] Zulkifli F., Abidin R. Z., Ariff M. I. M., Ahmad N. A., Arshad N. I., Ependi U., Razak M. S. A. Measuring the National Digital Identity Initiative in Malaysia: A Pilot Study with Rasch Measurement. Journal of Advanced Research in Applied Sciences and Engineering Technology. **38** (2), 153–164 (2024).

Прогнозування CPI в Малайзії: порівняння лінійної регресії, нелінійної регресії та методів нелінійного програмування

Зулькіфлі Ф.[1], Абідін Р. З.[1], Дені М. С.[2], Зулькіфлі Н. З.[1,3]

[1] *Коледж обчислювальної техніки, інформатики та математики, Технологічний університет MARA, Філія в Пераку, кампус Тапа, 35400 Тапах Роуд, Перак, Малайзія*

[2] *Коледж обчислювальної техніки, інформатики та математики, Технологічний університет MARA, 40450 Шах Алам, Селангор, Малайзія*

[3] *Стратос пінакле Сдн Бгд. Сігнатуре 2, VO6-07-02 Лінгкаран СВ, Санвей Велосіті Черас, 55100 Куала-Лумпур, Вілайя Персекутуан Куала-Лумпур, Малайзія*

У статті досліджуються фактори, що впливають на індекс споживчих цін (CPI), шляхом аналізу економічних показників і прогнозних моделей. Стаття починається з перевірки нормальності та кореляційного аналізу для визначення значущих змінних, а потім підгонки моделі за допомогою моделі лінійної регресії (LRM), моделі нелінійної регресії (NRM) і нелінійного програмування (NLP). Результати показують сильну позитивну кореляцію між CPI та такими змінними, як індекс збігу, праця та обсяг. Порівняння моделей вказує на те, що NRM є найефективнішим предиктором CPI з дещо нижчими значеннями середньої абсолютної похибки (MAE) і середньої абсолютної відсоткової похибки (MAPE), ніж LRM і NLP. Хоча NLP використовує менше змінних, це може спростити інтерпретацію моделі та зменшити обчислювальну складність. Це дослідження підкреслює важливість точних прогнозних моделей у прогнозуванні CPI для розробки політики на основі фактичних даних. Обмеженням є невеликий набір даних, що свідчить про те, що майбутні дослідження можуть вивчати альтернативні моделі, використовувати більші набори даних або проводити моделювання для підвищення точності прогнозування CPI.

Ключові слова: *індекс споживчих цін; модель лінійної регресії; модель нелінійної регресії; методи нелінійного програмування.*