



# An Optimized Ensemble Model for Inflation Prediction in Egypt

Ahmed M. Elshewey\*<sup>1</sup>

<sup>1</sup> Computer Science Department, Faculty of Computers and Information, Suez University, Suez  
43512, Egypt  
Emails: elshewy86@gmail.com

## Abstract

Inflation, an omnipresent economic phenomenon, is marked by a continual upsurge in the overall price levels of commodities and services within an economy. Accurately predicting inflation within a data-abundant setting poses a formidable challenge and remains a dynamic area of research encompassing several unresolved methodological inquiries. Among these, a significant query pertains to the identification and extraction of data offering the highest predictive capability for a targeted variable, particularly in scenarios characterized by numerous closely interconnected predictors, as encountered in the context of inflation prediction. Recently, the application of machine learning (ML) models has gained traction in predicting inflation parameters. The predictive accuracy of such models hinges significantly on the selection of an appropriate framework. Ensemble models, designed to amalgamate multiple base models, have emerged as a compelling strategy to yield superior predictive outcomes. In this study, we introduce a novel weighted average ensemble model tailored for the prognostication of inflation prediction. The proposed approach leverages three foundational base models: Linear Regression (LR), Polynomial Regression (PR), and Moving Average (MA) regression. The critical aspect of this ensemble lies in optimizing the weights assigned to each base model, thereby accentuating their strengths. To achieve this, we employ the Waterwheel Plant Optimization Algorithm (WWPA), a proficient optimization algorithm, to discern the optimal weight distribution for the base models. Comparative evaluations are conducted, pitting the proposed model against three other base models. Empirical findings conclusively demonstrate the superiority of the proposed weighted average ensemble model, underscoring its capacity to predict inflation with exceptional efficiency.

**Keywords:** Inflation; machine learning; linear regression; polynomial regression; moving average regression; ensemble model.

## 1. Introduction

The prevailing COVID-19 pandemic and the Russia-Ukraine conflict have been major drivers of the remarkable escalation in worldwide inflation levels. Navigating through this situation, policymakers are encountering substantial hurdles in pinpointing efficacious approaches to temper the surging prices. Despite these formidable challenges, precise inflation forecasts retain their pivotal significance in molding suitable policies aimed at alleviating this predicament. The advent of artificial intelligence methodologies has ushered in novel avenues for delving into inflation forecasting. These techniques play a pivotal role in connecting abstract theoretical frameworks with tangible, actionable remedies [1].

Generating dependable inflation forecasts remains an ongoing hurdle for policymakers, carrying immense significance for economic actors and their investment choices. Inflation introduces an element of unpredictability into investment determinations, leading to a truncation of the investment timeline, particularly in emerging markets. This underscores the relevance of crafting precise forecasts in economies of this nature [2]. A core aspect of this policy framework lies in its forward-looking

nature: considering the inherent time lag in monetary policy, policy decisions hinge upon foresight into the trajectories of inflation and other pertinent economic markers throughout the policy horizon. Consequently, for a central bank dedicated to inflation targeting, the caliber of forecasts for pivotal economic variables takes on supreme significance in formulating well-judged policy choices. Of these variables, inflation forecasts emerge as particularly pivotal.

At present, the BCCR employs a blend of univariate models, Bayesian models [3], and a semi-structural econometric model [4] to predict inflation. The precise prediction of critical macroeconomic variables carries immense significance for governing bodies and decision-makers, enabling them to amplify the effectiveness and impact of their economic policies. This, in turn, plays a positive role in fostering economic confidence by minimizing uncertainties related to pivotal economic indicators. In economies such as Turkey, marked by persistently high and volatile general price levels, the ramifications of escalated inflation rates are more pronounced compared to advanced economies. This divergence stems from the fact that such inflation not only truncates the timelines of investment projects but also substantially erodes savings and real incomes. Consequently, in emerging economies, the task of accurately forecasting inflation becomes not just a challenge, but a matter of utmost importance in mitigating its adverse consequences.

On a global scale, the responsibility of fortifying economic stability frequently rests with central banks. The intricacies of predicting macroeconomic indicators stem from the complex interplay of numerous highly correlated variables that impact the variable under scrutiny for prediction. Consequently, devising a forecasting model becomes a pivotal pursuit in research, especially when confronted with a broad array of potentially interconnected variables [5]. To tackle this challenge, factor models have been employed to distill information from extensive datasets, from which essential factors are extracted. This application of factor models enables the modeling of only a subset of factors rather than encompassing all variables within a forecasting equation. Nevertheless, the question of whether using the complete set of independent variables or a refined selection via pre-selection algorithms would result in more informative extracted factors remains uncertain. Several studies argue that factor models based on a larger number of variables might not necessarily yield enhanced forecasting performance compared to models relying on a narrower range of variables [6, 7]. Furthermore, the challenge of determining the optimal number of factors to incorporate into a forecasting equation remains unanswered [8].

In data-rich environments, an alternative strategy for predicting economic variables involves harnessing variable selection techniques to streamline the number of covariates. These techniques effectively pinpoint the most influential independent features associated with the target variable. This approach aids modelers in circumventing the challenge of degrees of freedom, a frequently encountered obstacle when grappling with high-dimensional data. Variable selection techniques have been observed to fulfill two primary roles: acting as substitutes for factor models, such as Least Absolute Shrinkage and Selection Operator (LASSO) and Least-Angle Regression (LARS), directly modeling the variable of interest [9, 10], or serving as pre-selection methods within factor models [11, 12] to pinpoint the specific variables of concern. Despite numerous studies showcasing the remarkable performance of machine learning (ML) models in the realm of time series analysis [13], the utilization of these models remains relatively constrained in the arena of macroeconomic forecasting [14].

When discussing prediction problems, the landscape is evolving with the ascent of machine learning techniques as prominent prediction tools. This shift is propelled by the expanding accessibility of extensive datasets, augmented computational capacities, and heightened availability of specialized software. These techniques have garnered widespread popularity, particularly in addressing classification and regression tasks concerning discrete variables. Examples include predicting loan delinquencies or consumer buying patterns, where machine learning often surpasses conventional methods. Typically, these applications involve utilizing cross-sectional data to achieve classification objectives. Yet, it's important to note that the versatility of machine learning extends beyond classification tasks. These methods can be tailored to predict continuous time-series data as well. Instances like predicting inflation or GDP growth showcase the adaptability of machine learning techniques in tackling economic variables characterized by ongoing, gradual changes.

This adaptability underscores the utility of machine learning in economic forecasting across a spectrum of scenarios. In this study, a unique ensemble model using a weighted average approach is introduced for the purpose of forecasting inflation. This novel approach is designed to harness the predictive capabilities of three fundamental base models: Linear Regression (LR), Polynomial Regression (PR), and Moving Average (MA) regression. The distinctive feature of this ensemble lies

in the meticulous optimization of the weights assigned to each base model, which enhances their individual contributions. To accomplish this optimization, the Waterwheel Plant Optimization Algorithm (WWPA) is employed. This algorithm, known for its proficiency, is utilized to determine the optimal distribution of weights among the base models.

## **2. Related Work**

Araujo et al. [2] delve into the implementation of machine learning (ML) techniques to enhance the accuracy of inflation forecasting in Brazil. It conducts an extensive out-of-sample forecasting exercise that spans various time horizons, employs a comprehensive dataset comprising 501 series, and encompasses a range of 50 distinct forecasting methods. This ensemble not only integrates innovative ML techniques introduced in this study but also encompasses traditional econometric models and forecast combination methodologies. Additionally, the study introduces analytical tools designed to illuminate the pivotal variables that contribute to inflation prediction, effectively providing insights into the inner workings of the ML "black box." The outcomes of the study reveal that while there is no universally superior model, ML methods frequently surpass conventional econometric models in terms of mean-squared error in many instances.

Moreover, the findings underscore the presence of nonlinearities within inflation dynamics, underscoring their significance for achieving accurate inflation forecasts. Among the standout forecasting options, forecast combinations, tree-based techniques (such as random forest and xgboost), breakeven inflation, and survey-based expectations often emerge. Collectively, these findings constitute a valuable enrichment of the macroeconomic forecasting domain, with a specific emphasis on the context of Brazilian inflation. Aras et al. [5] emphasize the extended utilization of machine learning (ML) models, either collaboratively with or independently of factor models. This can be achieved through the replacement of the functional form of the forecast equation in a factor model with ML models or by the direct integration of ML models using diverse feature selection techniques.

Notably, this research introduces novel tree-based models into the analysis, expanding upon recent insights from the existing literature. Furthermore, a distinctive facet of this study is the introduction of a pioneering approach that integrates feature selection techniques with Shapley values. This amalgamation aims to yield succinct explanations for inflation predictions. The comprehensive set of experiments conducted in the context of an emerging economy, Turkey, characterized by heightened volatility and uncertainty, demonstrates that tree-based ensemble models offer distinctive advantages. These models exhibit superior accuracy while concurrently providing interpretable predictions. These outcomes align well with the unique characteristics of the inflation landscape within the country. Breiman et al. [15] originally introduced the random forests (RF) algorithm, which has gained prominence in fields like medicine and biology. However, its adoption within economics has been relatively restrained.

Biau et al. [16] applied the RF algorithm to select variables for a GDP forecasting model in the Euro area. Their dataset encompassed 172 indicators, and their findings indicated that RF demonstrated favorable performance in comparison to autoregressive forecasts and predictions from the Eurozone Economic Outlook. This approach closely parallels the aim of our study, although it employs a single machine learning method as an intermediate step rather than the primary forecasting tool, as contrasted with an autoregressive benchmark. David et al. [17] investigated the utilization of random forests for variable selection in forecasting models relevant to economic phenomena was examined. However, our methodology diverges by employing random forests not as a tool to guide the selection of explanatory variables, but rather as the primary forecasting technique itself. It's important to highlight that the assessments of variable importance derived from applying random forests provide an opportunity to enhance other forecasting methodologies within the Banco Central de Costa Rica (BCCR). Nonetheless, we are of the view that pursuing such enhancements would be better suited for a separate research study.

Additionally, in a different context, Bajari et al. [18] integrated random forests into their machine learning toolkit for estimating grocery demand. Rodríguez et al. [19] present an initial evaluation of the predictive capabilities of machine learning methods for forecasting inflation in Costa Rica. The analysis involves generating forecasts using two variations of k-nearest neighbors, random forests, extreme gradient boosting, and a long short-term memory (LSTM) network. The evaluation is based on criteria derived from optimal forecast literature, and they juxtapose their performance with that of the average of univariate inflation forecasts currently employed by the Central Bank of Costa Rica. The findings reveal that the most effective forecasts emerge from LSTM, followed by univariate k-nearest neighbors, and to a lesser degree, random forests. Additionally, a combination of forecasts outperforms the individual forecasts contained within it, as well as the average of the univariate forecasts. Importantly, this combination is unbiased, displaying appropriate forecast error characteristics, and notably enhances forecast accuracy across all horizons.

This pertains to both predicting the inflation level and anticipating the direction of its fluctuations. Yang et al. [20] used the recurrent neural networks with a gated recurrent unit (GRU-RNN) model are employed to train and analyze Consumer Price Index (CPI) indicators, aiming to derive predictions related to inflation. The empirical findings, obtained from historical data, highlight the strong performance of the GRU-RNN model in forecasting China's inflation rate. Remarkably, when compared to certain traditional models, the proposed method demonstrates a substantially superior level of effectiveness. The focal point of the study by Simionescu et al. [21] is to refine the accuracy of quarterly inflation forecasts provided by the National Bank of Romania by employing sentiment analysis. Importantly, forecasts generated from sentiment analysis of textual narratives within official central bank reports exhibited notable superiority compared to numerical predictions from the central bank itself and various combined forecasts, spanning the temporal range from 2008:Q1 to 2021:Q4. Additionally, the study introduces novel forecasting models that amalgamate machine learning techniques, sentiment indices, and an array of forecasts from the National Bank of Romania.

Among these models, the most striking performer emerged as the forecasting model utilizing Fourier transform-derived signals as inputs for an artificial neural network and support vector machine. This particular model consistently showcased elevated forecast accuracy, surpassing the performance of all other models considered in the analysis. Theoharidis et al. [22] introduce an innovative hybrid deep learning model, known as Variational Autoencoders and Convolutional LSTM Networks (VAE-ConvLSTM), as an effective solution for predicting inflation. Drawing upon a publicly accessible macroeconomic dataset comprising 134 monthly US time series spanning from January 1978 to December 2019, the proposed model undergoes meticulous evaluation in comparison to a spectrum of established econometric and machine learning benchmarks. These encompass Ridge regression, LASSO regression, Random Forests, Bayesian methods, VECM, and multilayer perceptron. In terms of both consistency and out-of-sample performance, the VAE-ConvLSTM model distinctly outperforms its rival models.

This conclusion is robustly supported through cross-validation and Monte-Carlo simulations, which encompass a variety of training, validation, and test samples. The findings accentuate the potential of deep learning models in macroeconomic forecasting, especially when addressing intricate nonlinearities and nonstationarity. This heightened performance capability underscores the motivation to explore deep learning methodologies, which exhibit the capability to excel beyond the confines of conventional econometric approaches centered on linear and stationary models.

### **3. Methodology**

#### **A. Dataset**

The dataset used in this study is available at <https://data.worldbank.org/>. The data presents a chronological sequence of annual inflation rates spanning the years 1960 to 2022 in Egypt. Each year

is associated with a corresponding inflation rate, expressed as a percentage. The dataset reveals fluctuations in inflation over time, showcasing both periods of relatively low inflation and instances of heightened inflation. Here is a brief description of the key characteristics of the data:

- **Trend:** The data illustrates an evident variability in inflation rates throughout the years, with some years experiencing higher inflation rates and others witnessing lower rates.
- **Volatility:** The inflation rates are subject to significant variability, with periods of both moderate and pronounced fluctuations in inflation. Notable spikes in inflation are visible in certain years.
- **Periods of High and Low Inflation:** The dataset showcases years with substantially high inflation rates, particularly during certain periods like the mid-1970s, early 1980s, and the late 2010s. Conversely, there are also years marked by relatively low inflation rates.
- **Recent Trends:** In recent years, the dataset portrays a mix of moderate and somewhat higher inflation rates, indicating a degree of economic variability and potential influence from recent global events.
- **Recent Increases:** The dataset indicates a notable increase in inflation rates in the year 2021 and continues into 2022, suggesting a recent period of rising inflation.
- **Data Coverage:** The dataset spans over six decades, providing a comprehensive overview of inflation trends over a long period.

The dataset contains two features, namely, inflation and years. Figure 1 presents a plot between the years and the inflation.

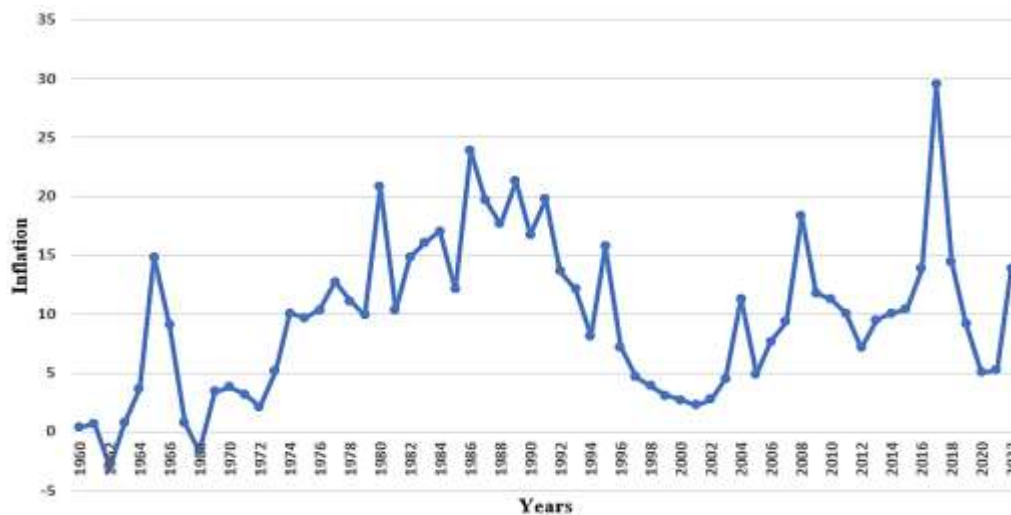


Figure 1: Plot of the data features between years and inflation

## B. Data Normalized

Data normalization, referred to as feature scaling or standardization, plays a pivotal role in machine learning preprocessing [23]. Its purpose is to convert dataset features into a consistent scale. This procedure significantly enhances the effectiveness and convergence of diverse machine learning algorithms, particularly those dependent on distance-related computations or optimization methods [24]. The primary objective of data normalization is to establish equitable contributions from various features during the learning phase, mitigating the potential dominance of a single feature due to its larger numerical values. In this study, Min-Max scaling normalization is used. This method scales the features to a specific range, often between 0 and 1. Min-max scaling is vulnerable to outliers since exceptionally large or small values have the potential to disrupt the scaling of the entire feature. The formula for min-max scaling is:

$$X_{normalized} = \frac{X - X_{min}}{X - X_{max}} \quad (1)$$

where,  $X$  is the original feature value,  $X_{min}$  is the minimum value of the feature, and  $X_{max}$  is the maximum value of the feature.

### C. Moving Average Regression

Moving Average (MA) regression is a fusion of the moving average (MA) model and linear regression. MA model is a statistical tool employed in time series analysis for predicting forthcoming values through the examination of past observations [25, 26]. This approach falls under the umbrella of Autoregressive Integrated Moving Average (ARIMA) models, amalgamating elements from both Autoregressive (AR) and Moving Average (MA) models [27]. This composite approach is frequently symbolized as an ARIMA ( $p, d, q$ ) model, where ' $p$ ' signifies the autoregressive order, ' $d$ ' signifies the differencing order, and ' $q$ ' signifies the moving average order. The MA component captures the relationship between the current value of the time series and past error terms (residuals). It's represented as follows:

$$x_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2)$$

where,  $x_t$  is the value of the time series at time ' $t$ ',  $\mu$  is the mean of the time series,  $\varepsilon_t$  represents the white noise at time ' $t$ ', which is an error term, and  $\theta_1, \theta_2, \dots, \theta_q$  are the coefficients of the MA terms, indicating the influence of past error terms on the current value.

The regression component introduces the influence of other independent variables on the time series. It's represented as follows:

$$x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (3)$$

where,  $\beta_0, \beta_1, \dots, \beta_p$  are the coefficients of the regression terms, and  $x_{t-1}, x_{t-2}, \dots, x_{t-p}$  are the lagged values of the time series.

The complete model combines both the MA and regression components:

$$x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_p x_{t-p} + \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (4)$$

The parameters  $\beta_0, \beta_1, \dots, \beta_p$  for the regression component and  $\theta_1, \theta_2, \dots, \theta_q$  for the MA component are estimated using techniques like maximum likelihood estimation.

### D. Polynomial Regression

Polynomial Regression (PR) is an analytical method employed to depict the connection between a dependent variable and one or more independent variables [28]. It expands upon the principles of linear regression by incorporating polynomial components, enabling the model to encapsulate non-linear associations among the variables [29]. Polynomial regression overcomes this restriction by incorporating polynomial expressions such as squares, cubes, and higher-order terms. Polynomial regression can be expressed as follows:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_n x^n + \varepsilon \quad (5)$$

where,  $y$  is the dependent variable,  $x$  is the independent variable,  $\beta_0, \beta_1, \dots, \beta_n$  are the coefficients of the polynomial terms, and  $\varepsilon$  is the error term.

The polynomial's degree ( $n$ ) establishes the model's level of complexity. A greater degree permits the model to closely match the data, yet it concurrently elevates the possibility of overfitting, wherein the model captures noise instead of genuine patterns [30]. Selecting the suitable polynomial degree holds utmost importance. A polynomial with a low degree could lead to underfitting, rendering the model too simplistic to grasp the fundamental patterns. Conversely, a polynomial with a high degree might induce overfitting, causing the model to conform excessively to the training data and subsequently

struggle to generalize effectively to novel data. In the context of polynomial regression, a consistent independent variable is elevated to diverse powers in order to generate novel features.

### E. Linear Regression

Linear Regression (LR) serves as a foundational statistical and machine learning method employed to model the connection between a dependent variable (referred to as the target) and one or multiple independent variables (known as predictors or features). In simple linear regression, there is only one independent variable, and the relationship is represented by a straight-line equation:

$$y = mx + b \quad (6)$$

where,  $y$  is the dependent variable (target),  $x$  is the independent variable (feature),  $m$  is the slope of the line, indicating the change in  $y$  for a unit change in  $x$ , and  $b$  is the  $y$  – intercept, which is the value of  $y$  when  $x$  is 0.

In multiple linear regression, there are multiple independent variables, and the relationship is represented by the equation:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (7)$$

where,  $y$  is the dependent variable (target),  $x_1, x_2, \dots, x_n$  are the independent variables (features),  $b_0$  is the intercept, and  $b_1, b_2, \dots, b_n$  are the coefficients representing the impact of each independent variable on the dependent variable.

The central objective of linear regression is to identify the optimal linear association that characterizes how changes in the independent variables impact the dependent variable. In linear regression, the objective is to identify the most suitable coefficients that minimize the disparity between the anticipated values and the real values (referred to as residuals). This endeavor is commonly accomplished through techniques like the least squares method, which endeavors to locate the line that minimizes the sum of squared residuals. Linear regression is extensively employed for both prediction and comprehending the connections among variables, even though it presupposes a linear relationship. It establishes the groundwork for more sophisticated regression methods and functions as a pivotal instrument across diverse domains, encompassing economics, social sciences, and machine learning.

### F. Waterwheel Plant Optimization Algorithm

The Waterwheel Plant Optimization Algorithm (WWPA) draws inspiration from the distinctive growth and energy utilization mechanisms of the waterwheel plant. This nature-inspired optimization technique aims to address intricate optimization challenges by emulating the growth and movement behaviors exhibited by the waterwheel plant. Salient features of the Waterwheel Plant Optimization Algorithm encompass:

- **Growth and Energy Utilization:** The carnivorous aquatic plant, *Aldrovanda vesiculosa*, possesses a specialized waterwheel-like structure enabling it to capture and consume prey using rapid trap movement. WWPA emulates this plant's energy-gathering process and growth dynamics to steer optimization strategies.
- **Population Dynamics:** In parallel with other optimization algorithms, WWPA operates through a population of potential solutions pertaining to the optimization problem. Each solution corresponds to a virtual waterwheel plant within the simulation.
- **Movement and Interaction:** The algorithm replicates the movement patterns of waterwheel plants in their aquatic habitat. Analogous to exploration in a search space, the plants navigate towards regions abundant in prey to optimize their energy acquisition.

- **Prey Capture and Energy:** Waterwheel plants employ their traps to ensnare prey. Within the algorithm, these prey instances symbolize potential solutions for the optimization problem. Capturing prey corresponds to the selection and evaluation of potential solutions.
- **Adaptation:** As waterwheel plants consume prey and accumulate energy, their growth and mobility capacities increase. This concept is harnessed within the algorithm to amplify the fitness of solutions based on their performance.
- **Solution Refinement:** The algorithm's iterations mimic the growth cycles of actual waterwheel plants. Progressing through the iterations, the algorithm enhances solutions by maneuvering them toward areas of heightened prey concentration, equivalent to improved solutions in the realm of optimization problems.

#### 4. Results and Discussion

Linear regression is a widely used statistical technique that aims to estimate the relationship between a dependent variable and one or more independent variables. When applying linear regression, the model generates predicted values based on the input variables. These predicted values are an approximation of the true values or actual values of the dependent variable as shown in figure 2. The accuracy of the linear regression model is often evaluated by comparing the predicted values to the actual values. If the predicted values closely align with the actual values, the model is considered to have a good fit. However, if there is a significant disparity between the predicted and actual values, it indicates that the model may not accurately capture the underlying relationship between the variables. This comparison between actual and predicted values helps assess the effectiveness and reliability of the linear regression model in making predictions and understanding the behavior of the dependent variable.

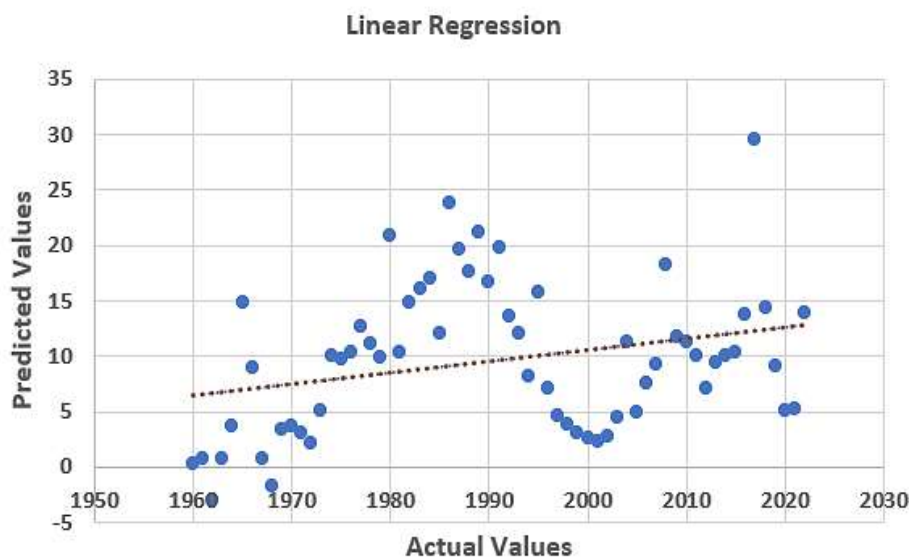


Figure 2: Actual values vs. predicted values using linear regression

Forecasting future inflation is a complex task, but linear regression can provide a useful framework for making predictions. By analyzing historical data on inflation rates and relevant economic variables, a linear regression model can estimate the relationship between these factors and inflation. This model can then be used to project inflation rates into the future, such as up to the year 2050. However, it is important to note that linear regression assumes a linear relationship between the variables, which may not hold true in the long term. Additionally, there are numerous factors that can influence inflation, including government policies, global economic conditions, and unforeseen events. Therefore, while linear regression can provide a starting point for forecasting inflation, it

should be used in conjunction with other forecasting techniques and careful consideration of other influential factors to obtain more accurate and reliable predictions for the future as shown in Figure 3.

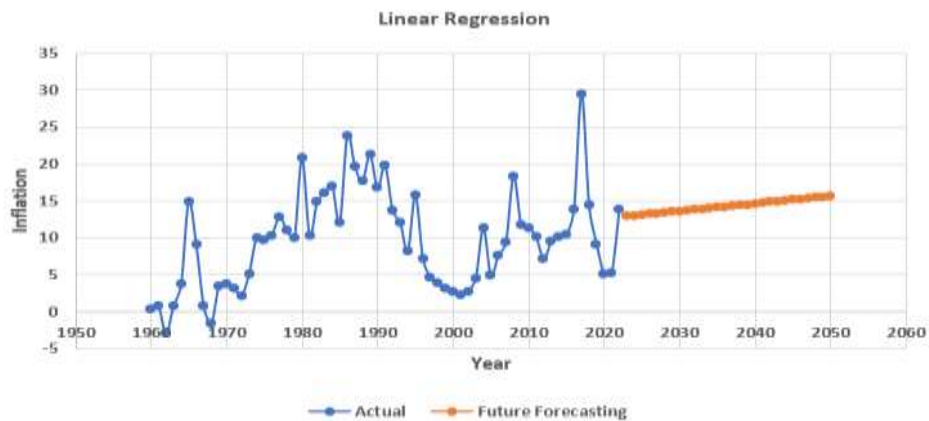


Figure 3: Future forecasting for inflation to 2050 using linear regression.

Polynomial regression is an extension of linear regression that allows for the modeling of nonlinear relationships between variables. When employing polynomial regression, the model generates predicted values based on the input variables, just like in linear regression. However, the key difference lies in the ability of polynomial regression to capture more complex patterns in the data by introducing polynomial terms, such as quadratic or cubic terms. By including these additional terms, the model can better fit the data and potentially improve the accuracy of predictions. When comparing actual values to predicted values in polynomial regression, it is important to assess the overall fit of the model. If the predicted values closely align with the actual values and exhibit patterns similar to the underlying data, it suggests that the polynomial regression model captures the nonlinear relationship well. However, if there are significant deviations or discrepancies between the predicted and actual values, it may indicate limitations in the model's ability to capture the complexity of the relationship or potential overfitting. Regular model evaluation and validation techniques are essential to ensure the reliability and effectiveness of polynomial regression in predicting future values as shown in Figure 4.

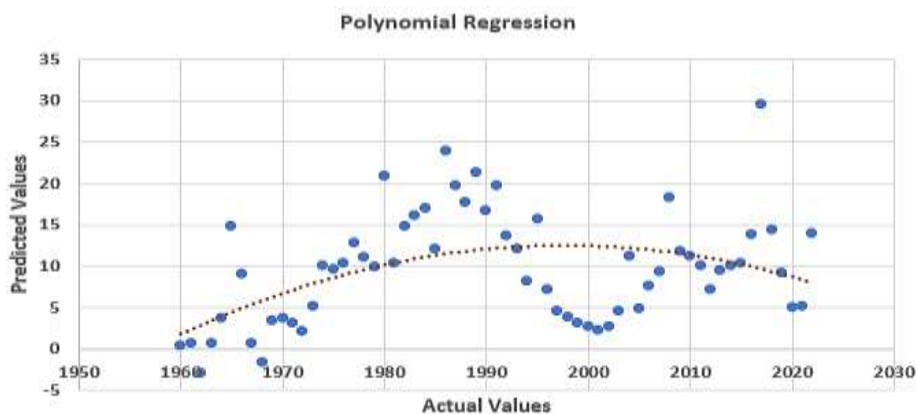


Figure 4: Actual values vs. predicted values using polynomial regression.

Forecasting future inflation to the year 2050 using polynomial regression can provide a valuable tool for understanding the potential trends and patterns in inflation rates. By incorporating polynomial terms into the regression model, it becomes possible to capture nonlinear relationships between economic variables and inflation, allowing for more flexibility in predicting future inflation rates. By analyzing historical data and identifying patterns in inflation fluctuations, the polynomial regression model can estimate the best-fitting curve to the data and project it into the future. However, it is crucial

to exercise caution when making long-term predictions, as the accuracy of the forecast relies heavily on the assumption that historical patterns will continue to hold true. Moreover, unforeseen events, shifts in economic policies, or changes in global dynamics can significantly impact inflation trends, potentially rendering the polynomial regression model less accurate over an extended forecast horizon like 2050. Therefore, it is advisable to combine polynomial regression with other forecasting techniques and regularly update the model with new data and variables to enhance the robustness and reliability of inflation predictions for the distant future as shown in Figure 5.

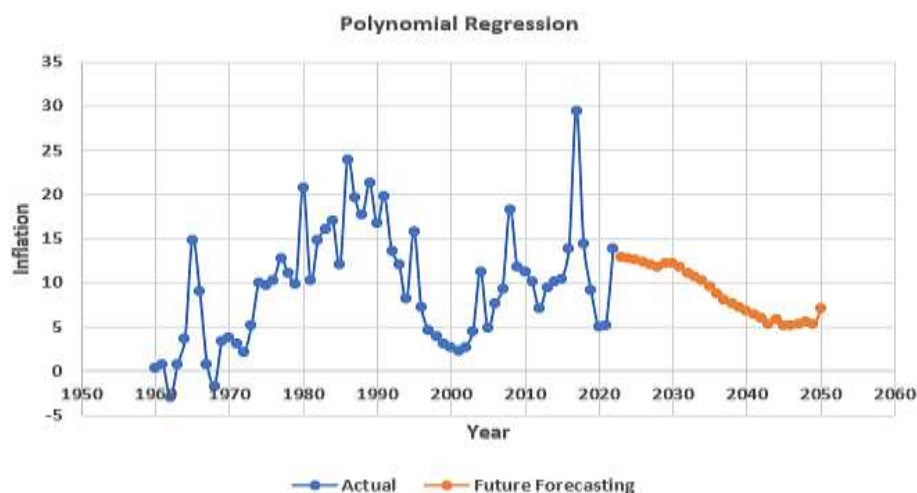


Figure 5: Future forecasting for inflation to 2050 using polynomial regression.

Moving average regression is a statistical technique commonly used for time series analysis and forecasting. In this approach, a moving average is calculated by taking the average of a fixed number of previous observations. When applied to regression, the model generates predicted values based on the moving average of historical data. When comparing actual values to predicted values using moving average regression, it is important to consider the smoothing effect inherent in the technique. The predicted values tend to exhibit a lag in response to changes in the actual values, as the moving average incorporates past observations. This lag can result in a less accurate representation of short-term fluctuations or sudden shifts in the data. However, moving average regression can be effective in capturing and forecasting the underlying trend of a time series, particularly when dealing with data that exhibits consistent patterns over time. It is important to note that the accuracy of predictions using moving average regression can vary depending on the choice of the moving average window size and the stability of the underlying data. Regular evaluation and refinement of the model are essential to ensure the reliability and effectiveness of predictions when using moving average regression as shown in Figure 6.



Figure 6: Actual values vs. predicted values using moving average regression.

Forecasting future inflation to the year 2050 using moving average regression can serve as a useful tool for understanding the long-term trends and patterns in inflation rates. By applying a moving average window to historical inflation data, the model calculates the average inflation rate over a fixed period. This moving average is then used to generate predicted values for future inflation. However, it is important to note that moving average regression tends to smooth out short-term fluctuations and may not capture sudden changes or shocks in the data. When making long-term predictions, it is crucial to consider the potential impact of various factors such as economic policies, global events, and structural changes in the economy. Additionally, the accuracy of predictions using moving average regression can be influenced by the choice of window size, which determines the length of the historical period used to calculate the moving average. Regular evaluation and refinement of the model, along with the incorporation of other forecasting techniques and external factors, are essential for enhancing the reliability and accuracy of inflation predictions up to 2050.

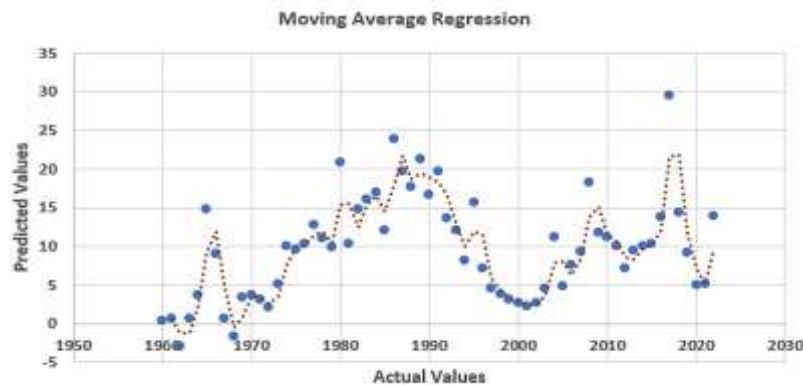


Figure 7: Future forecasting for inflation to 2050 using moving average regression.

The mean squared error (MSE), mean absolute error (MAE), median absolute error (MedAE), and determination of coefficient ( $R^2$ ) are calculated in below table 1.

Table 1: MSE, MAE, MedAE,  $R^2$

Model	MSE	MAE	MedAE	$R^2$
Linear regression	0.0062	0.0073	0.0069	85.37%
Polynomial regression	0.0048	0.0058	0.0055	89.72%
Moving average regression	0.0024	0.0049	0.0045	91.58%

## 6. Conclusion

Inflation is a widespread economic phenomenon caused by rising costs for goods and services. A data-rich environment makes inflation prediction challenging and raises several methodological problems. The selection and extraction of data having the highest predictive ability for a specific variable is critical, especially in cases with several closely linked variables, such as inflation prediction. Recently, inflation parameter prediction has drawn attention to machine learning (ML). The framework used determines how well such models forecast. Creating an "ensemble model" from multiple models improves forecast accuracy. This paper introduces a weighted average ensemble model for inflation forecasting. A linear regression (LR), polynomial regression (PR), and moving average regression (MA) model underpin the suggested technique. The optimization of base model weights is the most crucial aspect of this ensemble. We employ the strong Waterwheel Plant Optimization approach (WWPA) to establish the appropriate weighting strategy for our core models. Three baseline models are compared to the suggested model. Empirical evidence supports the weighted average ensemble model's highly accurate inflation prediction.

**Funding:** "This research received no external funding"

**Conflicts of Interest:** "The authors declare no conflict of interest."

## References

- [1] Di Vaio A, Hassan R., Alavoine C, Data intelligence and analytics: A bibliometric analysis of human–Artificial intelligence in public sector decision-making effectiveness. *Technological Forecasting and Social Change*, 174, 121201, 2022.
- [2] Araujo, G S Gaglianone, W P, Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. *Latin American Journal of Central Banking*, 4(2), 2023.
- [3] Chavarría J D, Morales C C, Pronósticos de inflación mediante técnicas bayesianas. *Economía y Sociedad*, 20(48), 1-27, 2015.
- [4] Muñoz Salas E, El modelo macroeconómico de proyección trimestral del Banco Central de Costa Rica. *Modelos macroeconómicos de la banca central: Centroamérica y República Dominicana*. México, DF: CEPAL, 16-43, 2008.
- [5] Aras, S. and Lisboa, P.J., Explainable inflation forecasts by machine learning models. *Expert Systems with Applications*, 207, 117982, 2022.
- [6] Aranda R A, Camacho M, Pérez-Quirós G, Finite sample performance of small versus large scale dynamic factor models. *Documentos de trabajo del Banco de España*, (4), 1-53, 2012.
- [7] Boivin, J. and Ng, S., Are more data always better for factor analysis?. *Journal of Econometrics*, 132(1), 169-194, 2006.
- [8] D'Agostino, A. and Giannone, D., Comparing alternative predictors based on large-panel factor models. *Oxford bulletin of economics and statistics*, 74(2), 306-326, 2012.
- [9] Konzen, E. and Ziegelmann, F.A., LASSO-Type Penalties for Covariate Selection and Forecasting in Time Series. *Journal of Forecasting*, 35(7), 592-612, 2016.
- [10] Medeiros, M.C. and Mendes, E.F.,  $\ell_1$ -regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors. *Journal of Econometrics*, 191(1), 255-271, 2016.
- [11] Bai, J. and Ng, S., Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2), 304-317, 2008.
- [12] Kim, H.H. and Swanson, N.R., Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting*, 34(2), 339-354, 2018.
- [13] Ahmed, N.K., Atiya, A.F., Gayar, N.E. and El-Shishiny, H., An empirical comparison of machine learning models for time series forecasting. *Econometric reviews*, 29(5-6), 594-621, 2010.
- [14] Masini, R.P., Medeiros, M.C. and Mendes, E.F., Machine learning advances for time series forecasting. *Journal of economic surveys*, 37(1), 76-111, 2023.
- [15] Breiman, L., Random forests. *Machine learning*, 45, 5-32, 2001.

- [16] Biau, O. and D'Elia, A., 2009. Euro area GDP forecasting using large survey datasets. In A random forest approach, 2009.
- [17] David, B., Model economic phenomena with CART and Random Forest algorithms (No. 2017-46). University of Paris Nanterre, EconomiX, 2017.
- [18] Bajari, P., Nekipelov, D., Ryan, S.P. and Yang, M., Machine learning methods for demand estimation. *American Economic Review*, 105(5), 481-485, 2015.
- [19] Rodríguez-Vargas, A., Forecasting Costa Rican inflation with machine learning methods. *Latin American Journal of Central Banking*, 1(1-4), 100012, 2020.
- [20] Yang, C. and Guo, S., Inflation prediction method based on deep learning. *Computational Intelligence and Neuroscience*, 2021.
- [21] Simionescu, M., Econometrics of sentiments-sentometrics and machine learning: The improvement of inflation predictions in Romania using sentiment analysis. *Technological Forecasting and Social Change*, 182, 121867, 2022.
- [22] Theoharidis, A.F., Guillén, D.A. and Lopes, H., Deep learning models for inflation forecasting. *Applied Stochastic Models in Business and Industry*, 2023.
- [23] Alkhamash, E.H., Hadjouni, M. and Elshewey, A.M., A hybrid ensemble stacking model for gender voice recognition approach. *Electronics*, 11(11), 1750, 2022.
- [24] Shams, M.Y., El-kenawy, E.S.M., Ibrahim, A. and Elshewey, A.M., A hybrid dipper throated optimization algorithm and particle swarm optimization (DTPSO) model for hepatocellular carcinoma (HCC) prediction. *Biomedical Signal Processing and Control*, 85, 104908, 2023.
- [25] Benjamin, M.A., Rigby, R.A. and Stasinopoulos, D.M., Generalized autoregressive moving average models. *Journal of the American Statistical association*, 98(461), 214-223, 2023.
- [26] Rocha, A.V. and Cribari-Neto, F., Beta autoregressive moving average models. *Test*, 18, 529-545, 2009.
- [27] Petukhova, T., Ojkic, D., McEwen, B., Deardon, R. and Poljak, Z., Assessment of autoregressive integrated moving average (ARIMA), generalized linear autoregressive moving average (GLARMA), and random forest (RF) time series regression models for predicting influenza A virus frequency in swine in Ontario, Canada. *PloS one*, 13(6), e0198313, 2018.
- [28] Ostertagová, E., Modelling using polynomial regression. *Procedia Engineering*, 48, 500-506, 2012.
- [29] Cheng, C.L. and Schneeweiss, H., Polynomial regression with errors in the variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1), 189-199, 1998.
- [30] Bickel, P.J. and Li, B., Local polynomial regression on unknown manifolds. *Lecture Notes-Monograph Series*, 177-186, 2007.