UNDER PEER REVIEW

OriginalResearchArticle

Enhanced Inflation Forecasting by Using a Novel Hybrid Model:EmpiricalEvidencefromSriLanka

Comment [M1]: Evaluating the predictive performance of monthly inflation rates in Sri Lanka using the hybrid model (HB)

Abstract

Aims/ objectives:This study aims to develop and evaluate the forecasting performance of a novel hybridmodel(HB)forpredictingmonthlyinflationrates.Theprimaryobjectivesincludeassessing the predictive accuracy of the HB model in comparison to established models and demonstrating itsadaptabilityandrobustnessoveranextendedperiodusinghistoricaldatafromSriLanka(1988- 2021).

Studydesign:Thisresearchadoptsatimeseriesforecastingapproach,specificallyemploying a unique hybrid model that combines Autoregressive Integrated Moving Average (ARIMA) and ArtificialNeuralNetworks(ANNs). Themodel'sefficacyisevaluatedthroughanextensiveanalysis of monthly inflation rates.

Methodology:Thehybridmodelisappliedtopredictmonthlyinflationrates, leveragingthestrengths ofbothARIMAandANNcomponents. HistoricaldatafromSriLankaisutilized,andtheforecasting accuracyiscomparedagainstotherestablishedmodels. Theevaluationconsidersmetricssuchas MeanAbsolutePercentageError(MAPE)toquantifythepredictiveperformanceoftheHBmodel.

Results:The empirical results reveal the superior forecasting accuracy of the hybrid model (HB) compared to other established models.The MAPE on the validation data is notably reduced to7.10%,demonstratingtheheightenedpredictiveprowessoftheHBmodel. Furthermore,themodel exhibits adaptability and robustness over different periods, enhancing its credibility as a reliable forecasting tool.

Conclusion:The innovative fusion of ARIMA and ANN methodologies in the hybrid model represents a significant advancement in time series forecasting.The demonstrated effectiveness and reliability of the HB model suggest its potential applicability in diverse economic and financial forecasting domains.By providing more accurate predictions, the HB model empowers decision-makers to make informed choices, contributing to improved decision-making processes.

Keywords: InflationRate;HybridModels;NeuralNetwork;Forecasting;NonlinearModels

2010MathematicsSubjectClassification:53C25;83C05;57N16

1 Introduction

The rate of inflation is an essential economic gauge that signifies the proportional fluctuation in the comprehensivevaluationofgoodsandservicesinagiveneconomythroughoutasetinterval, typically yearly Mankiw(2014).When the inflation rate is positive, it implies that prices are on an uptrend, thereby eroding the purchasing power of money.In contrast, a negative inflation rate, or deflation, suggests that prices are trending downwards Mankiw(2014).For policy-makers, monitoring the inflationrateiscrucialasitinformsdecisionsaboutmonetarypolicies,suchasmanagingthemoney supplyandadjustinginterestrates,withtheultimategoalsofmaintainingeconomicstability,fostering growth, and staving off recessions Mishkin(2010).

Fromamicroeconomicperspective, understandinginflationtrendsisessentialasitimpactsarange of financial decisions Mankiw(2014).For businesses, insights into the trajectory of inflation can aid in forecasting future costs and setting suitable prices for their goods or services.For investors, the inflation rate has implications for returns on investments, bond yields, and the general health of the stockmarket.Forindividuals, inflationaffectstheircostoflivingandinformsdecisionsaboutsaving and consumption Mishkin(2010).Given these considerations, accurate forecasting of the inflation rate, the central aim of our study, has substantial benefits for a broad range of stakeholders.

Our investigation delves into the realm of time series forecasting, a branch of data analysis that predicts future data points by examining past observations Hyndman and Athanasopoulos(2014). Thisprognosticationapproachhasseenextensiveuseindiversefields,encompassingeconomics, finance,agriculture,meteorology,andbiomedicalresearch,largelyduetotheabundantandcomprehensive datasetsandtheevolutionininformationprocessingtechnologyHyndmanandAthanasopoulos(2014). Acriticalelementofsuccessfultemporalsequencepredictionsisthedetectionandunderstandingof patterns and interconnections present in the data.These insights facilitate well-grounded forecasts about upcoming events.

TheBoxandJenkinsBoxandJenkins(1976)introducedmodel,knownastheAutoregressiveIntegrated MovingAverage(ARIMA),hasemergedasawell-acceptedapproachintimeseriesforecasting. The attractiveness of ARIMA models is anchored in their simplicity of use and versatility,makingthem a common selection for these predictive analyses.However, these models assume linearity, which posits a linear relationship among the time series observations Hyndman and Athanasopoulos(2014).Thispresumptionoftenfallsshortwhendealingwithnonlineardata.Toovercomethis , various nonlinear models like bilinear models, threshold autoregressive models, and autoregressive conditional heteroskedasticity models have been proposed Tong(1990).

TheusageofArtificialNeuralNetworks(ANNs)hasalsobeenobservedinthedomainoftimeseries predictionZhang(2003).ANNsofferaflexiblecomputationalframeworkcapableofmodellingabroad spectrumofnonlinearproblemswithoutnecessitatingpriorassumptionsaboutthemodel'sstructure Haykin(2009). AlthoughbothARIMAandANNshavetheirrespectivestrengthsindealingwithlinear and nonlinear data, it has proven challenging to identify a single model that can provide accurate predictionsforbothtypesofdataZhang(2003). Thishasledscholarstoinvestigatethepossibilities of mixed models that integrate the advantages of both ARIMA and ANNsZhang(2003).

Severalstudieshavedemonstratedtheeffectivenessofhybridmodels.ForexampleZhang(2003)proposedanANN -ARIMAhybridmodelforforecastingnonlineartimeseriesdata,achievingaRootMean Square Error (RMSE) of 0.0072.Similarly, Cadenas and Rivera Cadenas and Rivera(2009) utilized anANN-ARIMAhybridmodeltoforecastthespeedofwindinMexico,withtheirhybridmodeloutperforming both standalone ANN and ARIMA models. Another study by Perera et al. Perera et al.(2012) applied an ANN-ARIMA hybrid model to forecast stock data in Colombo, showing that the hybrid model outperformed both ANN and ARIMA models in terms of Mean Absolute Percentage Error (MAPE) and RMSE.

Comment [M2]: Ibid

Comment [M3]: Ibid

Comment [M4]: Old reference

Comment [M5]: Old reference

In our study, we aim to assess the performance of a novel hybrid model built upon an auto-tunedalgorithm,comparingitagainstthehybridmodelproposedbyZhangZhang(2003). Weusehistorical inflation data from Sri Lanka, spanning from 1988 to 2021, as the benchmark for evaluating these models of Sri Lanka(2021).The implementation of the models is carried out using the R software package(version4.1.3)andtheKeraslibraryforRTeam(2022);Cholletetal.(2018).Theperformance of our hybrid model will be compared with the Zhang,s model suggested by Zhang Cadenas andRivera(2009);Zhang(2003)intermsofforecastingaccuracyandotherrelevantperformancemetrics.

Comment [M6]: It is preferable to add more recent references

2 Methodology

2.1 Brock-Dechert-ScheinkmanTest(BDSTest)

TheBrock-Dechert-Scheinkman(BDS)Test,introducedbyBrocketal.(1996).in1987,isarenowned approachwithinchaostheoryusedtoidentifynonlinearity.Originallyaimedattestingforindependence andidenticaldistribution(iid),theBDStesthasprovenusefulfordiscoveringavarietyoflinear andnonlinearstructures.Itcanalsobeappliedtoresidualsfromafittedmodelforconducting portmanteau or misspecification tests.

Theunderpinningconceptoftheabovetestistheintegralcorrelation,whichquantifieshowoften temporalpatterns recur in the data.Given a time series X_t,defined as $X_t^m=(x_t,x_{t-1},\ldots,x_{t-m+1})$ anditsm-historydenotedas$X_t^m=(x_t,x_{t-m+1})$,them-dimensionalcorrelationintegralisexpressed as:

$$C_{m,T}(\epsilon) = \sum_{i<s} I(X_t^m,X_s^m) * \frac{2}{T_m(T_m-1)} \tag{2.1}$$

Here,$T_m=T-(m-1)$and$I_{X_t^m},I_{X_t^s}$areindicatorfunctions.ThefunctionvaluesofX_t^m,X_t^sare1 ifthesupnorm$\|X_t^m,X_t^s\|$and0inallothercases.Essentially,$C_{m,T}(\epsilon)$recordshowfrequentlym-historiesarefoundinproximitywithinahyper-cubeofaspecificsize.Inotherwords, thecorrelation integral computes the probability that any two m-dimensional points are near each other.

$$P(|X_t-X_s|<\epsilon, |X_{t-1}-X_{s-1}|<\epsilon, \ldots, |X_{t-m+1}-X_{s-m+1}|<\epsilon) \tag{2.2}$$

In the limiting case, assuming that the $X_t$are iid, this probability should equate to the following:

$$C_{1,T}(\epsilon)^m=P(|X_t-X_s|<\epsilon)^m \tag{2.3}$$

Brocketal.(1996)specifiedtheBDSstatisticsasfollows:

$$V_{m\epsilon}=T^{1/2}*\frac{C_{m,T}(\epsilon)-C_{1,T}(\epsilon)}{S_{m,T}} \tag{2.4}$$

Where $S_{m,T}$is the standard deviation, which, according to Brock et al.(1996), can be consistently estimated.ThedistributionoftheBDSstatisticapproachesanormaldistribution,withanaverageof 0 and a variance of 1.

Comment [M7]: The stages of estimation of the ARIMA model and the neural network mode should be detailed
• First: StageIdentification: examining the stability of the time series, and applying the necessary transformations to make it stable even if it is not.
• Second: Model Specification stage: Identify the appropriate model from the ARIMA model family.
• Third: Estimation stage of the model parameters.
• Fourth: Diagnostic Stage: Examining the model to verify its suitability for the time series - the subject of the research - and when it is not appropriate, we return to the second stage.
• Fifth: Forecasting Stage

3 DataSet

Inourresearch,weexaminethedataonSriLanka’smonthlyinflationratefromJanuary1988through August 2021 TradingView(2023). Figure 01 depicts these inflation rate data.

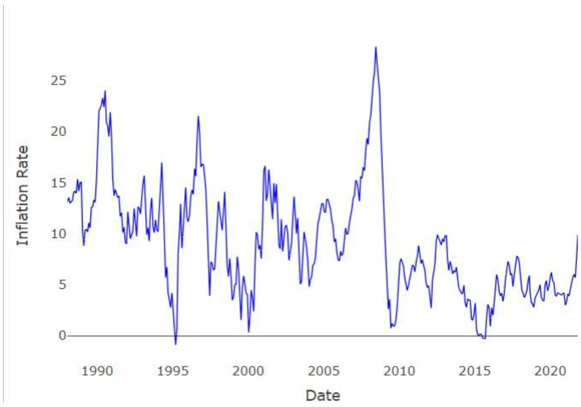

Figure1:MonthlymeanInflationrateofSriLanka(1988-2021)

Thedatasetemployedinthisresearchcomprisesatotalof407datapoints,withanaverageof9.61, variance 29.91554, minimum value of -0.890, and a peak value of 28.310.The data's time series representation shows a stochastic pattern with occasional outlier points, one noteworthy instance being in June 2008, likely associated with the nation's political climate at that moment.We divided thedatasetintothreesegments: trainingdataspanningfromJanuary1988toFebruary2020,which aids in checking the precision of the modeled fits; test data extending from March 2020 to February 2021; and validation data ranging from March 2021 to November 2021, which helps evaluate the accuracy of the projected values.

Comment [M8]: The time series stationarity test should be performed using the Dickey-Fuller test before the modeling procedures

4 Models

4.1 AutoRegressiveIntegratedMovingAverageofOrder(p,d,q),$(ARIMA(p,d,q))$

TheARIMAmodelBoxandJenkins(1976)isanexpansionoftheARMAmodelthatincludesscenarios of non-stationarity as well.In ARIMA models, a non-stationary time series is transformed into a stationaryonebyimplementingfinitedifferencestothedatapoints. Themathematicalrepresentation of the ARIMA(p, d, q) model using lag polynomials is provided as follows:

Comment [M9]: The researcher did not follow the Box-Jenkins method in estimating the ARIMA model

$$(1-\sum_{i=0}^{p}\phi_i L^i)(1-L)^d y_t=(1+\sum_{j=0}^{q}\vartheta_j L^j)\epsilon_t \tag{4.1}$$

In this context,p,d,and q are integers that are equal to or greater than zero.They denote the orderoftheautoregressive,integrated,andmovingaveragecomponentsofthemodel,respectively. The integer d determines the degree of differentiation.When d equals 0, the model simplifies to an ARMA(p,q)model.AnARIMA(p,0,0)correspondstoanAR(p)model, whereasARIMA(0,0,q)aligns with an MA(q) model.

4.2 ArtificialNeuralNetworks(*ANN*)

Thesinglehiddenlayerfeedforwardneuralnetworkservesasthemostprevalentapproachformodeling andforecastingtimeseriesHaykin(1999).Thearchitectureencompassesthreetiersofsimple processing units interconnected by acyclic links.There exists a mathematical connection between the output $X_t$and the inputs $X_{t-1}, X_{t-2}, \ldots\ldots\ldots X_{t-p}$:

$$X_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j H(\beta_0 + \sum_{i=1}^{q} \beta_{i,j} X_{t-i}) + \epsilon_t \quad (4.2)$$

Here, α and $\beta$are the model parameters, p denotes the count of input nodes, and q symbolizes the numberofhiddenlayers. Thelogisticfunction$H(t)$,typicallyemployedasthetransferfunctionforthe hidden layer, is given as follows:

$$H(X) = \frac{1}{1 + exp(-X)} \quad (4.3)$$

Basedonequation(1)andequation(2),theANNmodelessentiallycarriesoutanon-linearfunction that maps previous observations ($X_{t-1}, X_{t-2}, \ldots, X_{t-p}$) to future values ($X_t$).This can be expressed as:

$$X_t = f(X_{t-1}, X_{t-2}, \ldots, X_{t-p}, V) + \epsilon_t \quad (4.4)$$

where $V$is a vector of all parameters, and f is determined by the network structure and the weights of connections. Therefore, the neural network corresponds to a nonlinear auto-regressive model.

Comment [M10]: The researcher did not explain the stages of estimating the neural network model, starting from the training and testing phase, and we cannot assume the logistic function before training the network well.

4.3 ReiterationofG.PeterZhang'sModel(PZM)

G. Peter Zhang's hybrid model(PZM)Zhang(2003) harnesses the combined power of ARIMA and ANN models to enhance forecasting performance.The model initially applies an ARIMA model to scrutinizethelinearcomponentsofthedata,thenanANNmodelisdevelopedtomodeltheresiduals generated from the ARIMA model, which encapsulate information about the data's non linearity.

By separately modeling linear and nonlinear patterns and integrating the forecasts, the hybrid model canbolsteroverallmodelingandforecastingprecision.Interestingly,thehybridmethodmayincorporate sub-optimal models, which paradoxically can enhance the model's utility as it's often more effectiveto amalgamate forecasts derived from varying information sets.

4.4 TheProposedARIMA-ANNHybridModel(HB)

InaccordancewithZhangetal.(2008),itissuggestedthatatimeseriesisamixtureofalinear component and a non-linear component. Therefore, we can represent a time series X at time t as:

$$X_t = L_t + N_t \quad (4.5)$$

Here, Ldenotes the linear part of the time series Xat time t, while Nstands for the non-linear part. In the preliminary application of the ARIMA model to X, the non-linear part of Xcan be estimated as follows:

$$N_t = X_t - L_t \tag{4.6}$$

Where L_t is the estimated linear part of time series X at time t. In this study, 12 data points ahead were chosen as our test data length, and an "Akaike Information Criteria" based auto-tuned ARIMA model was used to forecast ten data points ahead using training data lengths ranging from 12 to 386, leveraging the power of Algorithm 2. The initial ten data points were omitted as the ARIMA model requires some data to make its first prediction. The advantage of this method is that it increases the length of training data, ensuring successful estimation of N_t. As pointed out by Faraway and Chatfield **?**, the accuracy of ANN models is heavily dependent on training data length. N_t is estimated at each step as follows:

$$N_t = [X_{t_1}, X_{t_2} \ldots . X_{t_{12}}] - [L_{t_1}, L_{t_2} \ldots . L_{t_{12}}] \tag{4.7}$$

Let $X_t - L_t = \epsilon_t$ and denote N_t as:

$$N_{352\times 10} = \begin{bmatrix} \epsilon_{1,1} & \epsilon_{1,2} & \cdots & \epsilon_{1,h} \\ \epsilon_{2,1} & \epsilon_{2,2} & \cdots & \epsilon_{2,h} \\ . & . & \ddots & . \\ . & . & & . \\ \epsilon_{L,1} & \epsilon_{L,2} & \cdots & \epsilon_{L,h} \end{bmatrix} \tag{4.8}$$

We intend to model N_t based on univariate ANN modeling techniques and transform $N_{386\times 12}$ to $N_{1\times 4632}$:

$$N'_t = (N_t)^T \tag{4.9}$$

By treating N'_t as a univariate time series, we can model the non-linear component of X_t. Finally, we can modify [equation 6] as follows:

$$X_t = L'_t + N'_t + e_t \tag{4.10}$$

Here, e_t represents the error term generated from estimating N_t using ANN modeling processes. The objective of making an accurate forecast for this task [Algorithm 2] is to minimize e_t.

4.5 Error Calculation Methods

Let n denote the number of fitted points, Y_t represent the actual value of the response variable Y at time t, and $\hat{Y}_t$ correspond to the predicted value of Y_t.

4.5.1 Mean Absolute Percentage Error (MAPE)

The Mean Absolute Percentage Error (MAPE) can be computed using the formula provided below.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_t - \hat{Y}_t|}{Y_t}.$$

MAPEisthemostcommonlyusedmeasureoftheforecasterroranditworksbestintheabsenceof an extremes values in the data set.

5 SimulationResults

5.1 PseudoCodeforOptimal*ARIMA*Order

Comment [M11]: A time series stability test must be performed before simulation which is a requirement for building an ARIMA model

Algorithm1: PseudoCodeforOptimal*ARIMA*Order

Input: X, p_max, d_max, q_max
Output:Optimum*ARIMA*Order

- For$p \leftarrow -0$ *to pmax*
- For$d \leftarrow -0$*todmax*
- For$q \leftarrow -0$*toqmax*
- Try:

  model←fit(arima(p,d,q))
- currentAIC←*AIC*(*model*)
- If*currentAIC<AIC*
- *AIC←OptimumARIMAOrder*
- optimum_model←model
- Cat:*error*=function(e)
- End If
- End For
- **return***OptimumARIMAOrder*(*p,d, q*)

The algorithm detailed here delineates the procedure for determining the ideal order of an ARIMA model, acriticalcomponentoftimeseriesforecasting.TheARIMAmodelnecessitatesthreeparameters: p (autoregressive part's order), d (degree of first differencing involved), and q (moving average part's order).The algorithm launches a sequence of loops for each parameter,within the bounds of 0 totheir respective maximum limits (p_{max}, d_{max}, and q_{max}).It then attempts to fit the ARIMA model witheachparametercombination.TheAkaikeInformationCriterion(AIC)foreachmodeliscomputed and contrasted with the best (lowest) AIC noted so far.If a model results in a lower AIC, it supplants the current optimal model. Ultimately, the algorithm returns the parameters of the ARIMA model that produced the lowest AIC, delivering the optimal order for the ARIMA model for the provided timeseries data.

5.2 Pseudocodeforproposedhybridmodel

Comment [M12]: Where is the neural network model?

Algorithm2: Pseudocodeforproposedhybridmodel

Input: D-Dataset,p max,d max,q max,H-testdatalength,L-thelearningrate,K-Initialindex train *ARIMA* model to , Network structure
Output:Trainedmodel

- For$i\leftarrow-$K*tolength*(D)
- For$p$ $\leftarrow-$0*to pmax*
- For$d\leftarrow-$0*todmax*
- For$q\leftarrow-$0*toqmax*
- Try:

  model←fit(arima(p,d,q))
- currentAIC←*AIC*(*model*)
- If*currentAIC<AIC*
- *AIC*←*OptimumARIMAOrder*
- optimummodel←model
- Cat:*error*←−function(e)
- End If
- **return***OptimumARIMAOrder*(*p,d, q*)
- Predict:(ARIMA$h\leftarrow-H$)
- $E\leftarrow-$D-Predict
- End For
- Transform(E)←−UnivereatT.S
- Initializeallweightsandbiases;
- whileterminatingconditionisnotsatisfied

  for each training tuple X in D

  Foreachinputlayerunit$j$

  $O_j=I_j$;
- $I_j=\Sigma_i W_{i,j} O_i+\Theta_j$
- $O_j=\frac{1}{1+e^{-1}j}$;
- Foreachunitjintheoutputlayer;

  $E_j=O_j(1-O_j)(T_j-O_j)$
- Foreachunitjinthehiddenlayers,fromthelasttothefirsthiddenlayer;

  $E_j=O_j(1-O_j)\sum E_k W_{j,k}$
- Foreachweight$W_{i,j}$innetwork

  $\Delta W_{i,j}=L(E_j O_i)$;

  $W_{i,j}=Wi,j+\Delta W_i$;
- ForeachbiasΘ$_j$innetwork

  $\Delta\Theta_j= L(E_j)$;

  $\Theta_j=\Theta_j+\Delta\Theta_j$;

Algorithm 2 shows that the pseudo code for proposed model.Hear H is the test data length and Kis the initial index of univareate data set and integer H needs minimum values to train the *ARIMA* model. In this case we selected 10 data points to train the *ARIMA* model.

5.3 BDSTestResults

m(Embeddingdimension)=2,3,4
Epsilonforclosepoints=2.7348,5.4695,8.2043,10.9390

Table1:ResultsofBDStest

m	[2.7348]	[5.4695]	[8.2043]	[10.939]
[2]	p<2.2e-16	p<2.2e-16	p<2.2e-16	p<2.2e-16
[3]	p<2.2e-16	p<2.2e-16	p<2.2e-16	p<2.2e-16
[4]	p<2.2e-16	p<2.2e-16	p<2.2e-16	p<2.2e-16

Table 1 shows that results of BDS test.According to those results our data set is not a indentation identicaldistribution(*iid*). BecauseP-valuesofclosepointsareclosetozeroinfirstfourEmbedding dimension and we can reject null hypothesis.

5.4 Predictedvaluesfortestdata

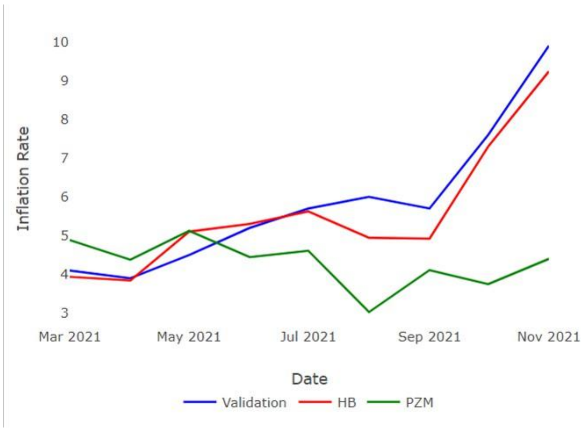

Figure2:PredicteddatacomparisonwithTestData

Thedisplayedplot(Figure02)isamulti-linechartrepresentingtheInflationRatepredictionsmadeby fourdifferentmodels–Test,HBandPZM,overaperiodfromMarch2020toFebruary2021.Eachline, distinguishedbyauniquecolor,representsaspecificmodel'spredictiontrajectoryoverthespecified period.By observing the line trends, one can gain insights into the behavior of each model.For instance, a line consistently above others indicates a model with a higher inflation rate prediction. Conversely, a jagged line with noticeable peaks and troughs suggests volatile model predictions withsignificantfluctuationsovertime.Thus,thisplotoffersacomprehensivevisualdepictionof the models' performances, allowing for an easy comparison and identification of patterns in their respective predictions.

Table2:Testdatacomparisonofpredicteddata

Date	Test	HB	PZM
Mar-2020	5.4	5.44	5.96
Apr-2020	5.2	5.15	5.29
May-2020	4	3.97	4.22
Jun-2020	3.9	3.98	3.50
Jul-2020	4.2	4.04	4.38
Aug-2020	4.1	4.08	4.16
Sep-2020	4	3.86	4.29
Oct-2020	4	4.07	4.19
Nov-2020	4.1	4.06	3.23
Dec-/2020	4.2	4.15	4.41
Jan-2021	3	2.90	3.43
Feb-2021	3.3	3.29	3.53
MAPE	-	1.60	7.83

The Table 02 presents the predicted inflation rates(test data) by HB and PZM,across a time frame from March 2020 to February 2021.Each row corresponds to a specific month, while each column signifies the inflation rate predicted by a particular model for that month.It allows for a thorough examination of the individual model's performance over time by comparing the data vertically within each column.

Most notably, the table includes the Mean Absolute Percentage Error (MAPE) at the end, a key indicator of the accuracy of the model's predictions.The MAPE for HB, and PZM, is provided as 1.60% and 7.83%, respectively.These MAPE values serve as benchmarks for model comparison, withlowervaluessignifyinghigherpredictiveaccuracy. Forinstance,withaMAPEof1.60%,theHB model exhibits the most accurate predictions among the two models in this case.

Insummary,thistablenotonlyprovidesagranularviewofeachmodel'smonthlypredictionsbutalso quantitatively compares their performance via the MAPE metric.This approach, in conjunction with thevisualcomparisonofferedbythelinegraph,facilitatesacomprehensiveevaluationofthemodels' inflation rate predictions.

5.5 Froecastedvaluesforvalidationdata

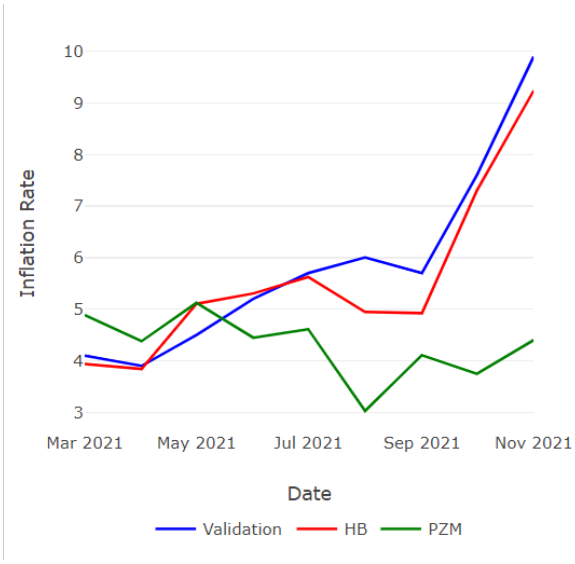

Figure3:ValidationDatawithForecastedvalues

The Figure 03 depicted here showcases the inflation rate forecasts from March 2021 to November 2021usingtwodifferentmodels: HBandPZM,. Thebluelinerepresentsthevalidationdata, which serves as the actual observed data against which the forecasts are validated.The other lines - red forHB,greenforPZM,theforecastsmadebyeachrespectivemodel. Overthegiventimeperiod,itis evidentthatalltwomodelsexhibitvaryingdegreesofdeviationfromthevalidationdata. Forinstance, thePZMmodel(Greenline)demonstratesahigherdeviationinlatermonths,whereastheHBmodel (Red line) appears to track relatively closer to the actual data.

Table3:Validationdatacomparisonofforecasteddata

Date	Validation	HB	PZM
Mar-2021	4.1	3.93	4.88
Apr-2021	3.9	3.84	4.37
May-2021	4.5	5.10	5.12
June-2021	5.2	5.30	4.44
Jul-2021	5.7	5.62	4.60
Aug-2021	6	4.94	3.02
Sep-2021	5.7	4.92	4.10
Oct-2021	7.6	7.30	3.74
Nov-2021	9.9	9.23	4.40
MAPE	-	7.10	27.17

This table presents a comparison of two forecasting models (HB, PZM) against the actual validation dataforaperiodfromMarch2021toNovember2021.Foreachmodel,theMeanAbsolutePercentage Error (MAPE) is also provided, which serves as an indicator of the model's accuracy.

TheHBmodelhasthelowestMAPEat7.10,whichindicatesthatonaverage,themodel'spredictions are approximately 7.10% off from the actual values.The PZM model's MAPE is 27.17, suggesting that its predictions deviate by about 27.17% from the actual values.These MAPE values can be usefulfordeterminingwhichmodelprovidesthemostaccuratepredictionsforthegivendata. Based onthistable,theHBmodelappearstoprovidethemostaccurateforecasts,withthelowestaverage deviation from the actual validation data.

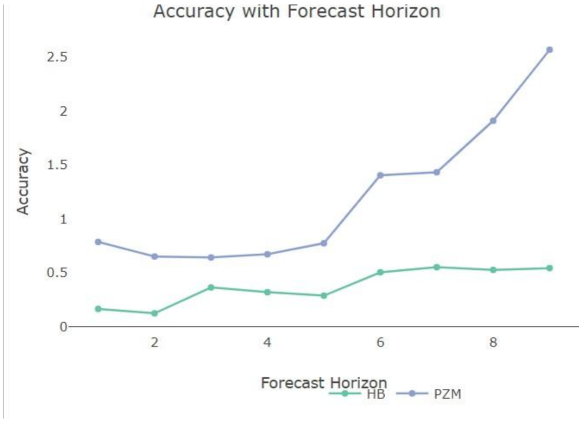

Figure4:RMSEChangewithforecasthorizon-Validationdata

TheRMSEplotfurthersubstantiatestheassessmentofmodelperformancebyvisualizingthevariation of error as a function of the forecast horizon.As illustrated in the plot,the RMSE of all modelstends to increase with forecast length, indicating a reduced predictive accuracy over time.The PZM model (Blue line) shows the highest increase, corroborating the comparatively higher MAPE in the earliertable. Conversely,theHBmodel(Greenline)presentsamoresteadyerrorincrease,affirming its superior forecast precision.This graphical representation thus provides an intuitive means to compare forecast model robustness over an expanding forecast horizon.

6 ConclusionandDiscussion

Thisstudypresentedacomparativeevaluationoftwoforecastingmodels(HB,PZM)usingreal-world validationdata. TheperformanceofthesemodelswasassessedbytheirMeanAbsolutePercentage Error(MAPE),acommonlyusedmetricforevaluatingtheaccuracyofforecastingmodels.Theresults of the study revealed significant differences in the forecasting performances of the models.

The proposed model (HB) for this study, demonstrated the best performance with the lowest MAPE 7.10% meaning it had the smallest average deviation from the actual validation data.The other model,PZMshowedlargerdeviationsfromtheactualvalues,withMAPEsof27.17%. Thisindicates that the HB model was more accurate and reliable in forecasting the data in this specific instance.

ItisessentialtonotethatwhiletheHBmodelperformedthebestinthiscontext,theoptimalforecasting model can vary depending on the specific data set and the characteristics of the data. For instance, different models may perform better for data with different levels of noise, seasonality, or trend patterns.Therefore,thechoiceofthebestforecastingmodelshouldalwaysbeinformedbyathorough understanding of the data and the specific requirements of the forecasting task.

Thisstudyprovidesagoodbasisforfurtherresearchinimprovingforecastingaccuracy.Future work could explore the use of more advanced techniques, such as machine learning or deep learning models,orthecombinationofmultipleforecastingmodelstoenhancepredictionperformance.Moreover, furthervalidationoftheHBmodelusingdifferentdatasetswouldalsobevaluabletoconfirmits robustness and generalizability.

References

Box,G.E.andJenkins,G.M.(1976).*Timeseriesanalysis:forecastingandcontrol*.Holden-Day.

Brock,W.A.,Dechert,W.D.,andLeBaron,B.(1996).Thebdsstatisticasatestfortheadequacy of a linear time series model.*Journal of Time Series Analysis*, 17:215–235.

Cadenas, E. and Rivera, W. (2009).Wind speed forecasting in three different regions of mexico, using a hybrid arima-ann model.*Renewable Energy*, 34:2732–2738.

Chollet,F.etal.(2018).Keras:Thepythondeeplearninglibrary.

Haykin,S.(1999).*NeuralNetworks:AComprehensiveFoundation*.PrenticeHall.

Haykin, S. (2009).*Neural Networks and learning machines*.Pearson Education, 3 edition.

Hyndman,R.J.andAthanasopoulos,G.(2014).*Forecasting: principlesandpractice*.OTexts. Mankiw,

N. G. (2014).*Principles of Economics*.South-Western Cengage Learning.

Mishkin,F.S.(2010). *TheEconomicsofMoney,Banking,andFinancialMarkets*. Pearson. of

Sri Lanka, C. B. (2021).Central bank of sri lanka statistical information.

Perera, N., Senadeera, M., and Sivakumaran, N. (2012).Stock market forecasting using arima-based neural networks:Evidence from sri lanka.*International Journal of Business and Economics Research*, 1:53–59.

Team,R.C.(2022).R:Alanguageandenvironmentforstatisticalcomputing.

Tong,H.(1990).*Non-lineartimeseries: adynamicalsystemapproach*.OxfordUniversityPress. TradingView

(2023). Economics-lkiryy.

Zhang, G., Patuwo, B. E., and Hu, M. Y. (2008).Forecasting with artificial neural networks:The state of the art.*International Journal of Forecasting*, 14:35–62.

Zhang,G.P.(2003).Timeseriesforecastingusingahybridarimaandneuralnetworkmodel. *Neurocomputing*,50:159–175.