WORKING PAPER 206/2021

INTEGRATION OF ECONOMETRIC MODELS AND MACHINE LEARNING- STUDY ON US INFLATION AND UNEMPLOYMENT

Sri Rajitha Tattikota
Naveen Srinivasan

MSE
Mentoring Excellence

MADRAS SCHOOL OF ECONOMICS
Gandhi Mandapam Road
Chennai 600 025
India

June 2021

Integration of Econometric Models and Machine Learning- Study on US Inflation and Unemployment

Sri Rajitha Tattikota
Madras School of Economics, Chennai, India
pgdm19srirajitha@mse.ac.in

and

Naveen Srinivasan
(Corresponding author)
Madras School of Economics, Chennai, India
naveen@mse.ac.in

WORKING PAPER 206/2021

June 2021

Price : Rs. 35

MADRAS SCHOOL OF ECONOMICS
Gandhi Mandapam Road
Chennai600 025
India

Phone: 2230 0304/2230 0307/2235 2157
Fax: 2235 4847/2235 2155
Email : info@mse.ac.in
Website: www.mse.ac.in

Integration of Econometric Models and Machine Learning- Study on US Inflation and Unemployment

Sri Rajitha and Naveen Srinivasan

Abstract

In this study we compare the in-sample-accuracy to evaluate the performance of Econometric models and Machine Learning models on the Time Series data. Enclosed to explore techniques which perform better for Time Series Classification to predict the state (High, Medium, or Low) of each quarter by studying macroeconomic variables in the United States: Inflation and Unemployment. In the direction of improving the models using machine learning techniques and investigating how they are incorporated in time series data to improve the efficiency of the predictions. We perform a comparative analysis of various models for this classification problem. In ML, Logistic regression, K-Nearest neighbors, Support vector machines, Gradient boosting and Random forest models were explored. In Econometrics, Autoregressive Moving Average and Autoregressive Conditional Heteroskedasticity models were explored. The results showed that Machine learning models are superior compared to the traditional Econometric models for time series data. The best model for Unemployment data was EGARCH in Econometrics and K- Nearest Neighbors to predict both 2 states and 3 states in ML. The best model for Inflation data was EGARCH in Econometrics and Linear SVM, Random forest to predict 2 states and 3 states respectively in ML. Even though the ML models lack the interpretability and clarity in the exact internal process, these models have resulted exceptional in terms of accuracy in predictions. Econometric modelling would be more suitable, if we focus to only understand the effect and interpret the casual effect of the data.

Keywords: *Inflation, Unemployment, Econometric models, Machine Learning*

JEL Codes: *C5, E24, E27, E31, E37*

Acknowledgement

This paper is a revised version of the first author's master's dissertation thesis at the Madras School of Economics. The authors acknowledge the access to the OECD dataset of Unemployment and Inflation rates for the purpose of the study's analysis.

Sincerest thanks to Dr. Rakesh Nigam, Principal, Department of Research and Business Analytics, Madras School of Economics for his valuable suggestions and comments throughout this research work.

Sri Rajitha Tattikota
Naveen Srinivasan

INTRODUCTION

Every institution either government or non-government, small scale or large scale need good fitting forecasting models. Better forecasting processes facilitate the institution to estimate the future and optimize their actions accordingly. Inflation and Unemployment are most important macroeconomic variables which have significant influence on the functioning of the country and the economic well-being of people. Developing highly accurate techniques for predicting the inflation and the unemployment is crucial.

Hence, forming appropriate expectations and accurate predictions about the future movements of these variables is very essential both for the policymakers, bankers, and economists. However, it is a generally agreed fact that Inflation and Unemployment data is very noisy, non-linear, and unstable. Modeling these variables has been a challenging, often a daunting, task, more so, when there are frequent business cycles and seasonal variations in the economy. To find an appropriate model, which could make accurate forecasts and could be flexibly reapplied to shifting cyclical data has been a research question persistently pursued by Researchers.

Traditionally, Linear Time Series (TS) models, principally Univariate, was considered suitable for modelling the data. Non-linear TS into vogue to accommodate the non-linear and volatile part of the data. however, the performance and precision of these models have been falling short of the requirement of policy makers and bankers. (Andreas, Konstantinos, Georgios, Sermpinis, Stasinakis, 2014). Even though researchers seem to capture the pattern of the macroeconomic variables given common market conditions these models are not effective in providing accurate results for the periods of economy shocks and recessions. (Andreas et.al, 2014 and Maehashi and Shintani, 2020).

Machine learning (ML) is a subset of the field of Artificial Intelligence which is targeted to develop algorithms that automatically learn from a given set of evidence. This area has been at the center of importance for the advances in science and technology. However, Econometric models are mainly ingrained and built using the economic theories. These economic theories provide an insight to understand the cause-and-effect relationship between observed variables involved in these models which facilitates our forecasts to be logical, reliable, and consistent. On the other hand, modelling through ML is using the past data, building structure of the model, and generating forecast value as a result. ML searches for the best fit of model for a certain problem and decides the pattern, functions, and parameters of the model. These models are used to predict future values of the variables using a known value obtained from the historical data or experience. (Gutierrez, Hervas-Martinez, Martinez-Estudillo and Fernandez, 2009).

Either of these methods Econometric and ML modelling have their own virtues and shortcomings. The major advantage when we use econometric models is the ability of the model to analyze a causal relationship between variables and make them clear. These models not only predict the values but also help us to understand economies of the function and reasons for the failures in strategies. In case of ML models, the main advantage is of accuracy that the models result i.e., the difference between the estimated value and the actual value is smaller. These models do not require any prior information about the distribution or probability of data, instead they just learn from the given past information or experience which enables them to work with nonlinear data and imperfect data.

Despite this increasing attention towards ML, they are still not very well established in the literature of forecasting. Especially in the case of univariate time series, it is mostly dominated by the statistical methods based on linear processes like Exponential smoothing or Autoregressive integrated moving average (ARIMA). The main objective of this paper is to analyze the performance of ML models in modelling the macroeconomic data- Inflation and Unemployment. (Cerqueira, Torgo and Soares, 2019).

Consequently, this study focusses on exploring how Time Series classification can be performed using the univariate time series data of the United States- Inflation rate and Unemployment rate, by deploying the most widespread methods in classification: Logistic regression, Decision trees, K- Nearest Neighbors, Support Vector Machines using kernels, Regularization techniques, Random forest, and Gradient Boosting methods. This study results in Random Forest to be the best model for the univariate Inflation data and K-NN to be the best model for Unemployment data, for predicting the state in which these rates will be- whether high state, medium state or low state based on the allotted threshold value.

In the further sections of the study, we will focus on what areas the previous researchers have worked on and the gap which we cover through this study. Also, covering the description of the data and methodology used in the study. Later, the results after exploring how the ML models work when classifying the given TS data and how the accuracy of these models can be improved. Finally, concluding based on the results and suggestion on further scope of the study.

REVIEW OF LITERATURE

Macroeconomic analysis, modelling, and forecasting are a widely researched areas in the applied economic literature. The accuracy of various forecasting methods known is a topic of continuing interest and research. While the use of ML methods for forecasting exchange rates, stock prices and incomes widely popular, ML for inflation forecasting has started recently. The use of ML for many other macroeconomic variables has motivated the study of comparison of efficacy with TS modelling. This is useful in understanding the efficiency of technical and computational intelligence techniques in forecasting macroeconomic variables.

A paper on US inflation concentrated on modelling the volatility of inflation using the Generalized Auto Regressive Conditional Heteroskedasticity (GARCH) models. The results have shown that the AR-GARCH model can approximate reality quite well. And demonstrate how the theoretic results on the correlation structure of the inflation model can be used. (Karanasos, Karanassou and Fountas, 2004) An article based on the forecasting of US Unemployment rates deployed the Auto Regressive (AR), Moving Average (MA) and GARCH models which compared the accuracy of TS models. They found that MA-ARCH has given better forecasts. The results of the comparisons of these forecasts in TS models show that the simplest models are the best. Also suggested that the future research should seek to investigate more complex forecasting methods to predict the unemployment series. (Floros, 2005).

One of the papers highlights the advantage of Support Vector Regression (SVR) model was able to adapt efficiently to the structural breaks for both Inflation and Unemployment. According to this study, Genetically Adapt (GA)- SVR has forecasting superiority even during periods of economic instability. GSVR selection of estimators has resulted both statistically and computationally efficient. (Andreas et.al, 2014).

Another study aimed towards investigating how ML to forecast US Unemployment has applied Radial Basis Neural Networks (RNN) and SVR regressions. Also checked how the traditional methods of moving averages, smoothing techniques work in this scenario. The results showed that SVR techniques improves the statistical capabilities and accuracy of the model in comparison with Simple Average techniques. (Andreas et.al, 2014).

A study on US inflation compared the Univariate and Multivariate models along with ML models in forecasting inflation. Analysis on different measures of inflation to find the best forecasting model have shown that K-Nearest Neighbors algorithm, AR Distributed Lag model and SVR fitted well for the measures that were studied. It was concluded that ML models are most appropriate for inflation measures with high volatility and TS modelling for core inflation forecasting. (Müller and Guido, 2016 and Ulke, Sahin and Subasi, 2018).

With the aim of studying relatively better performance of TS over ML models, a study has concluded that ML models tend to underperform when the sample size is very small as they assume a more flexible functional form, hence tend to overfit. Thus, the study suggests that TS models are more suited for small sample size, while ML models outperform as the sample size increases. (Cerqueira, Torgo and Soares, 2019).

An article written on examining whether ML is better on economic data or not has used ML methods to study and model unemployment rate in US. A ML method study to model and forecast unemployment rate in US was undertaken by Oberlin and Nomura. Naive Bayes Forecast, Lasso Regression and ANN were used simultaneously to decide the optimal model. It was observed that Navie’s Forecast underperformed, compared to other methods and NN has higher precision. (Kreiner and Duca, 2020).

A journal studying the application of macroeconomic forecasting to Japan employed multiple factors, ML and Deep learning algorithms to analyze their accuracy in modelling multiple macro-economic variables which concluded that ML models perform better than traditional TS models. ML models particularly have resulted to be more efficient in the presence of non-linearity and work well for forecasts for longer periods of time. (Maehashi and Shintani, 2020).

There are no formal studies on how ML can be used to forecast the Unemployment and Inflation rates considering the Univariate TS data. We often try to predict the absolute values of these rates which are vulnerable to change in the economy. The interest of this study would be focusing on the domain of Time Series Classification which does not have much literature or research in macroeconomics variables. Also, to check how well the macroeconomic data can be classified into states representing the cycle they are in.

DATA AND METHODOLOGY

Aim: Time Series Classification using Machine Learning.

Objective of the study:

1. To evaluate the performance of econometric models and machine learning models on the time series data.
2. To explore techniques which perform better and determine the best fit.
3. To improvise the models using machine learning techniques to get accurate predictions.

Data: Secondary Data. The study is based on the data collected from the FRED database.

1. Unemployment data- 204 observations. The data has quarterly rates ranged from 1970-01-01 to 2020-10-01. (OECD, 2021)
2. Inflation data- 243 observations. The data has quarterly inflation rates ranged from 1960-01-01 to 2020-10-01. (OECD, 2021)

Forecasting Methodology

For achieving the objectives of the study Explorative research was done.

UNIVARIATE TIME SERIES MODELS

Univariate time series models are a class of specifications where we attempt to model and to predict the target/ dependent variable using only information from their own past values and feasibly current and past values of an error term. The time series models are usually a-theoretical which imply that their structure and use is not based on any underlying theoretical model.

Auto Regressive Moving Average – ARMA (p, q)

By combining the AR with p lags and MA with q lags models, an ARMA (p, q) model is obtained. Such a model states that the current value of some series y depends linearly on its own previous values plus a combination of current and previous values of a white noise error term. With p as the own lags and q as the error lags. (Brooks, 2014 and Chen and Deo, 2004).

When the variable is level stationary, we can use ARMA model given by:

$$Y_t = a_0 + a_1 Y_{t-1} + \cdots + a_p Y_{t-p} - b_1 e_{t-1} - \cdots - b_q e_{t-q} + u_t$$

Auto Regressive Integrated Moving Average – ARIMA (p, d, q)

This is a series which is modeled only in terms of its own past values and some disturbances. The 'I': Integration here means the number of times we are differencing or integrating the series to get stationarity; usually denoted by I (r) or I (d). (Johnston and Dinardo, 1996 and Brooks, 2014).

Seasonal ARIMA- SARIMA (p, d, q) x (P, D, Q)

This model is an extension to ARIMA model which would support the modeling of the seasonal component of time series. This includes three trend elements that are the same as the ARIMA model. There are four more elements that are seasonal elements which are not part of ARIMA model are: 'm' which is seasonal period, 'P' seasonal autoregressive order, 'D' seasonal difference order and 'Q' seasonal moving average model. (Johnston and Dinardo, 1996).

$$\Phi(A^s)\phi(\mathrm{A})(x_t - \mu) = \Theta(A^s)\theta(A)\ \omega_t$$

Autoregressive Conditional Heteroscedasticity - ARCH (p)

This model says that the variance of the error term at time t depends on the squared error terms from previous periods. Both the mean and variance are modelled simultaneously.

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \ \alpha_2 e_{t-2}^2 + \cdots + \ \alpha_p e_{t-p}^2$$

Where σ_t^2 is the variance of the error term and p is the number of lagged terms.

For the mean models we could consider Zero mean using model with zero conditional mean estimation and simulation, Constant mean, AR mean using autoregressive model with optional exogenous regressors, HAR mean using heterogeneous autoregression model or the LS mean using least squares model estimation.

Generalized ARCH – GARCH (p, q)

The generalized ARCH is a weighted average of long-term average. It allows conditional variance to be dependent on previous own lags. (Brooks, 2014).

The specification is as follows:

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \ \alpha_2 e_{t-2}^2 + \cdots + \ \alpha_p e_{t-p}^2 + \cdots + \gamma\sigma_{t-1}^2 + \cdots + \gamma\sigma_{t-q}^2$$

Where, α_i is the ARCH coefficient and $\gamma_i$is the GARCH coefficient.

Threshold GARCH- TARCH

This one of the asymmetric GARCH models where we try to factor in the leverage effect.

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \gamma\sigma_{t-1}^2\ \lambda e_{t-1}^2 d_{t-1}$$

Where d_t =1 if $e_t < 0$ (bad news) and d_t =0 if $e_t > 0$ (good news). They have differential effects on conditional variance. If $\lambda > 0$, there is a leverage effect i.e., bad news increases the volatility and $\lambda \neq 0$, the impact of the new is asymmetric.

Exponential GARCH- EGARCH

This is also one of the asymmetric GARCH models where we try to factor the asymmetric effect which is non-linearly effecting the volatility. The log of conditional variance equation is taken which means leverage effect is exponential than quadratic. This condition also ensures that the forecasts of conditional variances are positive.

$$\log \sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 \left|\frac{e_{t-1}}{\sigma_{t-1}}\right| + \lambda \frac{e_{t-1}}{\sigma_{t-1}}$$

They have differential effects on conditional variance. If $\lambda > 0$, there is a leverage effect i.e., bad news increases the volatility and $\lambda \neq 0$, the impact of the new is asymmetric.

MACHINE LEARNING MODELS

Classification is a process of classifying, grouping, or categorizing a given set of the data into classes. This can usually be performed on both structured and unstructured data. The aim of this process is to predict the class of given set of data points. These classes are referred to as target, label, or categories.

Logistic Regression

This method is used when the target/ dependent variable is a categorical variable. Utilizing the Logit function and the concept of log odds which models the probability of class allocation of various observations. Data points are assigned to classes based on whichever likelihood is larger.

$$\log(\frac{p(Y=1)/X}{1-p(Y=1)/X}) = \beta_0 + \beta_1 X$$

Decision Trees

Decision trees utilize the structure of Nodes and Branches to split observations based on certain optimal criteria - wherein optimal splits are defined using Gini Index or Entropy. The terminal nodes signify class allocation of observations. In the equation of Gini index note that p_{mk} is the proportion of values in the m^{th} node having the k^{th} class label.

$$Gini\ Index = \Sigma_{k=1}^{k}\, p_{mk}(1 - p_{mk})$$

K- Nearest Neighbor

KNN is a model that utilizes distance metrics and locality of data points to assign class labels. Each new observation is assigned the class corresponding to the predominant class of its 'K' nearest neighbors where 'nearness' is determined by distance metrics. In the equation N_0 represents the set of points in the locality of the point x_0 whose classification we compute by calculating the fraction of neighbors having j as class. (James, Witten, Hastie and Tibshirani, 2013).

$$P(Y = j|X = x_0) = \frac{1}{k}\Sigma_i I(\, y_i = j)$$

Support Vector Machines

These are mechanisms that use the concepts of vector spaces and separating hyperplanes to classify data points. This model displays its versatility when dealing with inherently non-linear structuring of datasets. Kernel tricks allow us to effectively expand the dimensionality of data till it becomes linearly separable.

Gradient Boosting Method-GBM

Used primarily in conjunction with decision tree models - using the concept of boosting to minimize predictive errors. Boosting is essentially the process of sequentially fitting various instances of models on top of the residuals of previously fit models to reduce error.

Random Forest

This is a technique uses decision trees as the baseline model. These are ensemble methods that run many instances of a model in parallel. The final predictions are then taken to be the mean of mode of class assignments predicted by the constituent models. These models keep the minimalism of decision trees but still utilizing the power of the ensemble. (Müller and Guido, 2016)

Regularization Techniques

Regularization is used to resolve the problem of over-fitting in the model. A penalty hyperparameter is included in the loss function minimization equation - essentially as a constraint on parameter values. Popular techniques are Ridge and Lasso methods.

$$Regularized\ Loss\ Function = Original\ Loss\ Function + Lambda * R(f)$$

R(f) can be a variety of functions, but this is where the separation of L-1 and L-2 methods is made according to how a penalty on the complexity of the model will be imposed. In L-1 regularization technique the weights for each parameter are assigned as a 0 or 1 which helps perform feature selection in sparse features spaces. This technique is good for high dimensional data since the coefficient zero would cause some features to be removed from the model. This process is done by taking the sum of absolute values of the weights of these feature coefficients.

TECHNIQUES FOR IMPROVISING MODELS

Basis Transformation

To make linear regression adapt to the nonlinear relationships between variables transform the data according to basis functions. We consider multidimensional linear model:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 \ldots$$

Then develop the $x_1, x_2, x_3, \ldots$ and so on from our single-dimensional input x then our model becomes a polynomial regression given by:

$$y = a_0 + \Sigma\alpha_i x + \Sigma\beta_i x^2 + \Sigma\gamma_i x^3 \ldots$$

where i = 1,2,3 This function would still give a linear model as the linearity in model means linearity in coefficients (as they never multiply or divide each other). We have effectively taken our 1-dimensional feature values and project them into higher dimension such that a linear fitting can fit further complicated relationships between the features and target variable y. (VanderPlas, 2016).

Principal Component Analysis- PCA

A principal component analysis is done with objective of explaining the variance– covariance structure of features through linear combinations of these features with goal of dimensionality reduction and interpretation. Even though n components would be required to replicate the variability of whole system but much of this variability can be explained by k principal components with $k < n$. There is a lot of information in these k components and the transformed data set is now reduced to a data set consisting of these components as features. The principal components are those uncorrelated linear combinations whose variances are very large. (Johnson and Wichern, 2018).

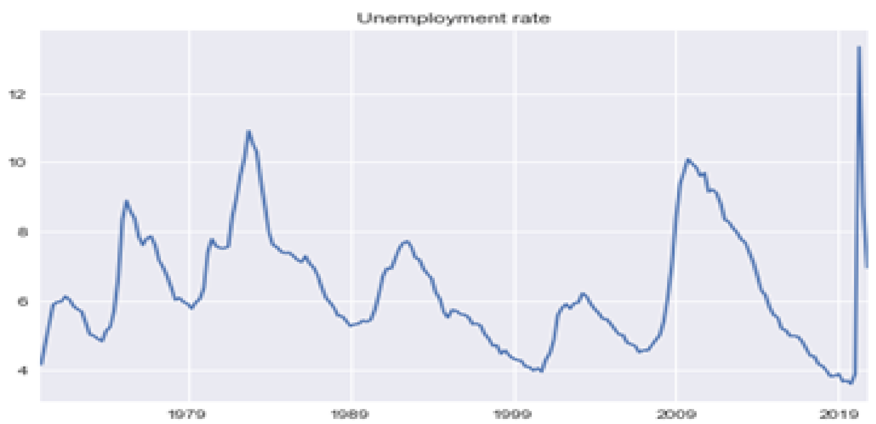
RESULTS AND DISCUSSION

Econometric Modelling

Unemployment Data

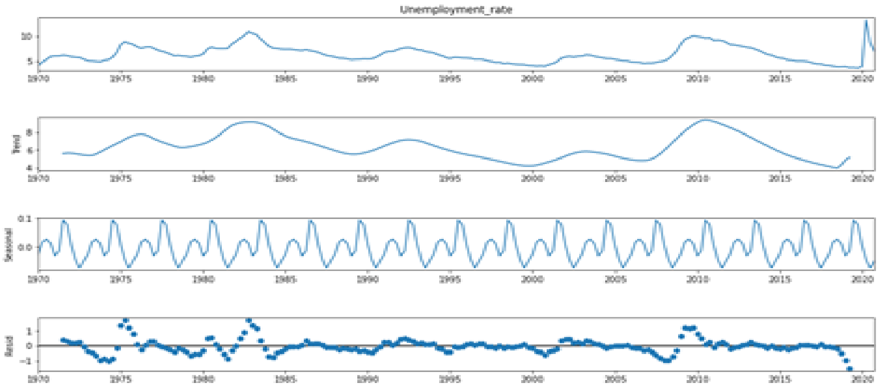
We get the subsequent results by modelling the Unemployment rate data. Figure 1 is the plot of the quarterly Unemployment rates in the United States from 1970. In Figure 2, we see the plot of how the components are decomposed into trend, seasonal and irregular for the given univariate data. This graph clearly shows us that there is no presence of trend in this data. Then we check the stationarity of the data using the ADF test which has the test statistic of -3.56 and a P Value- 0.00649. As the P value is very close to zero, we can reject the null hypothesis (H_0: The data is not stationary) of this test at 1 percent significance level. We can say with 99 percent confidence that the given data of Unemployment rate is Stationary. Once the data is stationary it is good to be modelled. To determine the number of lags to be used in the model, we look at the autocorrelation and partial autocorrelation functions (ACF and PACF) in Figure 3. The ACF function gives us the number of lags for the Moving Average process(q) and PACF function gives us the number of lags for Auto Regressive process (p).

Figure 1: Quarter-wise Unemployment vs Time



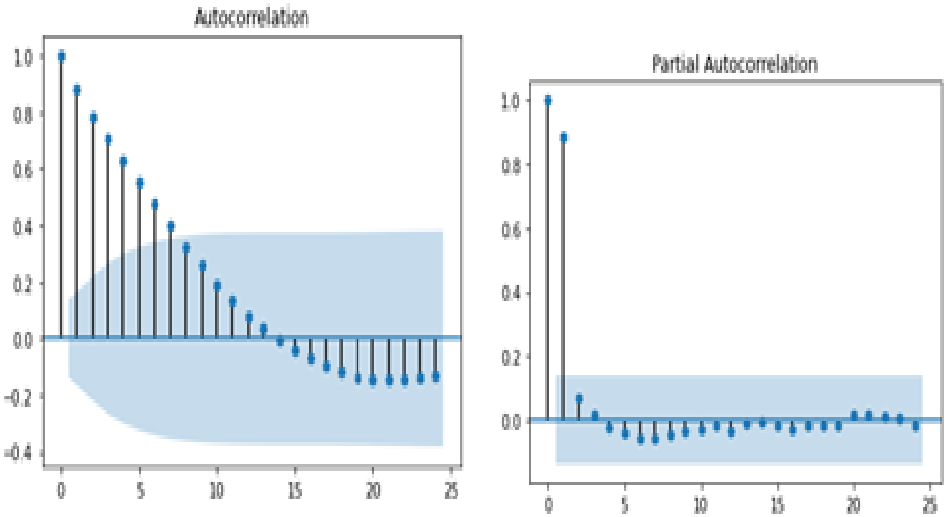
Source: Authors' Calculations.

Figure 2: Decomposed Components of Unemployment Rate



Source: Authors' Calculations**.**

Figure 3: ACF and PACF Plot of Unemployment Rate



Source: Authors' Calculations

Then we iteratively check for the p and q values that give us the lowest AIC. The value of p equal to 1 and q equal to 4 has given us the lowest AIC. Then we further apply other traditional econometric models to model unemployment rate. EGARCH (p=1, o=1, q=1) is the model with minimum AIC and RMSE value shown in the comparison Table 1.

Comparison of Models

Table 1: Comparison of Validation Metrics Across Models-Unemployment Rate

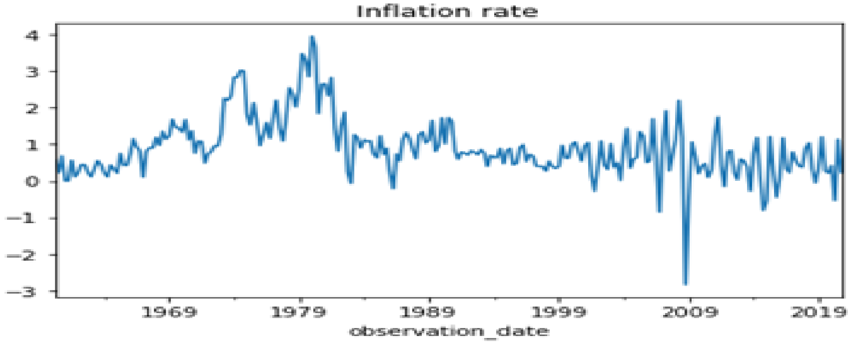
	Model	AIC	BIC	Test-RMSE
1	**ARIMA (1, 0, 4)**	41.348	59.722	1.75
2	**SARIMA (1, 0, 4)**	71.715	91.71	1.852
3	**ARCH (1)**	46.4	56.86	1.184
4	**AR(2)-ARCH (4)**	16.79	37.63	1.406
5	**GARCH (1,1)**	18.472	34.1	1.348
6	**GARCH (2,4)**	22.93	48.98	1.148
7	**TGARCH**	284.48	294.98	1.88
8	**EGARCH (4,0,1)**	10.03	33.38	1.31
9	**EGARCH (1,1,1)**	**3.658**	**21.89**	**1.069**

Source: Authors' Calculations **.**

Inflation Data

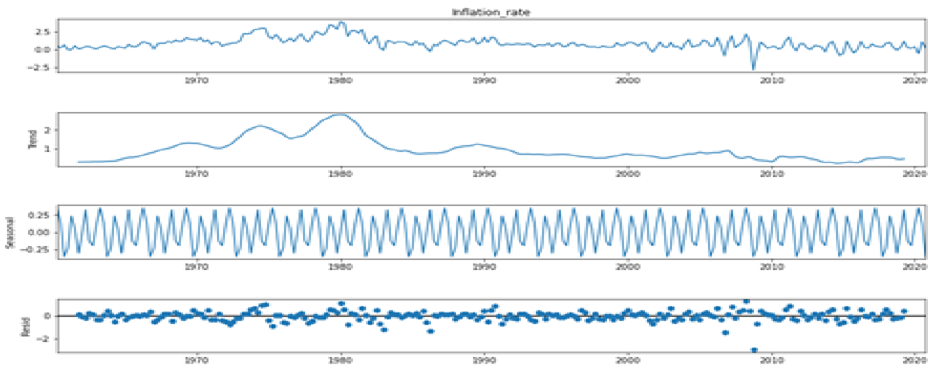
Figure 4 is the plot of the quarterly Inflation rates in the United States from 1960. In Figure 5, we see the plot of how the components are decomposed into trend, seasonal and irregular for the given univariate data. This graph clearly shows us that there is no presence of trend in this data. Then we check the stationarity of the data using the ADF test, but it has a lower test static and higher P value when taken as level variable. We then check the stationarity of the data using the first difference which has the test statistic of 6.621 and a P Value 0.00. As the P value is very close to zero, we can reject the null hypothesis (H_0: The data is not stationary) of this test at 1 percent significance level. We can say with 99 percent confidence that the given data of Inflation rate is Stationary with first difference. Once the data is stationary it is good to be modelled. Similarly, to determine the number of lags to be used in the model, we look at the autocorrelation and partial autocorrelation functions (ACF and PACF) using Figure 6.

Figure 4: Quarter-wise Inflation rate vs Time



Source: Authors' Calculations.

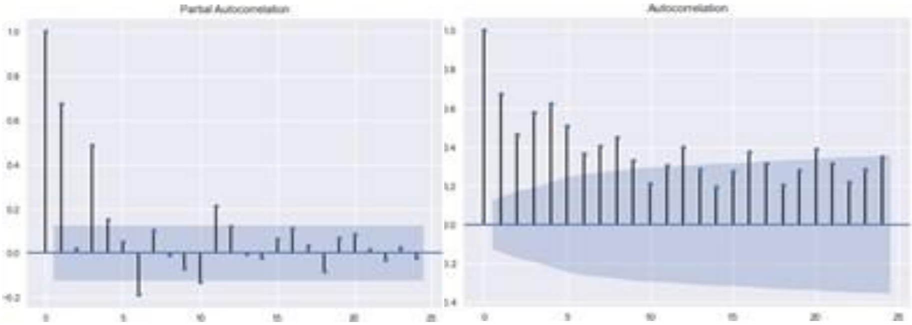
Figure 5: Decomposed Components of Inflation Rate



Source: Authors' Calculations **.**

Then we iteratively check for the p and q values that give us the lowest AIC. The value of p equal to 2 and q equal to 4 has given us the lowest AIC. Then we further apply other traditional econometric models to model Inflation rate. EGARCH (p=1, o=1, q=1) is the model with minimum AIC and RMSE value shown in the comparison Table 2.

Figure 6: PACF and ACF Plot of Inflation Rate



Source: Authors' Calculations

Comparison of models

Table 2: Comparison of Validation Metrics Across Models-Inflation Rate

	Model	AIC	BIC	Test-RMSE
1	ARIMA (2, 1, 4)	163.16	186.044	0.612
2	SARIMA (2, 1, 4)	175.4	200.27	0.627
3	ARCH (1)	195.58	207.02	0.745
4	AR (2)-ARCH (4)	367.45	384.86	0.601
5	GARCH (1,1)	162.064	179.17	0.585
6	GARCH (2,2)	165.12	187.94	0.589
7	TGARCH	202.12	212.59	0.755
8	EGARCH (1,0,1)	161.75	178.86	0.585
9	EGARCH (1,1,1)	**139.29**	**159.25**	**0.581**

Source: Authors' Calculations**.**

MACHINE LEARNING MODELLING

We will see the results of the performance of various ML models before and after carrying out certain data transformation operations that involve basis expansions and principal component analysis. At first however we present the baseline models wherein only a minimal amount of data pre-processing has been applied. This will then be contrasted with the ML models of subsequent sections wherein we obtain better accuracy metrics after applying the data transformations.

Data Pre-processing

We want to make this data classifiable, so the first step towards it is to get a threshold to classify the data. As predicting in terms of absolute values can be challenging, we now try to split the data into states i.e., to classify such that we get the state in which these macroeconomic variables will be at any point in time. we first consider a two-state model where the state space contains ’High and Low’ classes, then we extend the model to a three-state model wherein the state space contains ’High, Medium, Low’. Since the distribution of our data is skewed, we consider the log transformation of these rates to make it normal. The steps followed before fitting the data into the models.

(i) Applying a log transformation to the data and make it conform to normality assumptions of various models. Also, enabling us better separation of the states.

(ii) In case of Inflation data, we have negative values. Hence, we shift the values by 10 since adding a positive constant merely shifts the origin of all the data points, the statistical properties and structure of the data distribution is not affected. Hence, we shift the values by +10 to ensure that log transformation can be applied.

(iii) Then, we annotate the class labels as follows - for the two-state scenario, considering the mean as the threshold, all values below the mean are labeled as ’Low’ and all values above as ’High’. For the three-state scenario, the threshold for the ’High’ and ’Low’ states are the Mean ± Standard deviation, whereas the ’Medium’ state happens to be the region within the two thresholds.

(iv) Once the states are labelled, then we take 3 previous time lagged values for each observation in case of Unemployment data and 9 lagged values for Inflation data. These are essentially our 'features' or independent variables based on which, we will predict the state of the dependent/ target variable.

(v) Basically, we will be predicting the state of 'Next quarter' given the values of 'Previous quarters.'

UNEMPLOYMENT DATA

Two-State Modelling

The two states are taken as High and Low based on the mean (1.811) of the log transformed data where 2 states [(1.281, 1.811] < (1.811, 2.590]] is the set division. The values from 1.281 to 1.811 are grouped into the Low state and the values from 1.811 to 2.590 are grouped into the High state. The results in table 3 show that SVM using Radial kernel is the best model when we predict 2 states of unemployment rate (H or L) with the test data in-sample-accuracy of 91.67 percent. The second-best model is Logistic regression using the *L-1* regularization with the test data in-sample-accuracy of 91.66 percent.

Table 3: Comparison of 2-State Models Based on Accuracy- UR

	Model	Train	Test
1	Logistic Regression	90.55	83.33
2	Decision Trees	95.55	91.66
3	SVM- Linear	79.44	75
4	**SVM- Radial**	**93.33**	**91.67**
5	KNN	95	91.66
6	**L-1 Regularization**	**93.88**	**91.66**

Source: Authors' Calculations.

Three-State Modelling

The three states are High, Medium, and Low where 3 states [(1.281, 1.553] < (1.553, 2.069] < (2.069, 2.590]] is the set division. The values from 1.281 to 1.553 are grouped into the Low state, the values from 1.553 to 2.069 are grouped into the Medium state and the values from 2.069 to 2.590 are grouped into the High state. The results in table 4 show that SVM using Radial kernel is the best model when we predict 3 states of unemployment rate (H, M or L) with the test data in-sample-accuracy of 83.34 percent. The second-best model is K-Nearest Neighbors (k=9) with the test data in-sample-accuracy of 83.34 percent. Even though, K-NN has the same test accuracy we choose it as the second best because of the difference in accuracy of train and test data is higher for this model.

Table 4: Comparison of 3-State Models Based on Accuracy- UR.

	Model	Train	Test
1	Logistic Regression	91.66	83.33
2	Decision Trees	92.77	83.33
3	SVM- Linear	79.44	82.33
4	**SVM- Radial**	**86.11**	**83.34**
5	**KNN**	**89.44**	**83.34**
6	L-1 Regularization	80.55	83.33
7	Random Forest	95	83.33

Source: Authors' Calculations.

INFLATION DATA

Two-State Modelling

The two states are taken as High and Low based on the mean (2.386) of the log transformed data where 2 states [(1.970, 2.386] < (2.386, 2.635]] is the set division. The values from 1.970 to 2.386 are grouped into the Low state and the values from 2.386 to 2.635 are grouped into the High state. The results in table 5 show that SVM using Radial kernel is the best model when we predict 2 states of inflation rate (H or L) with the test data in-sample-accuracy of 80.39 percent. The second-best model is Logistic regression with the test data in-sample-accuracy of 80.39 percent.

Table 5: Comparison of 2-State Models Based on Accuracy-Inflation Rate

	Model	**Train**	**Test**
1	**Logistic Regression**	**72.22**	**80.39**
2	Decision Trees	71.11	60.78
3	SVM- Linear	70.66	71.21
4	**SVM- Radial**	**73.88**	**80.39**
5	KNN	73.33	76.47
6	L-1 Regularization	76.66	78.43
7	Random Forest	84.44	78.43

Source: Authors' Calculations.

Three-State Modelling

The three states are High, Medium, and Low where 3 states [(1.970, 2.312] < (2.312, 2.460] < (2.460, 2.635]] is the set division. The values from 1.970 to 2.312 are grouped into the Low state, the values from 2.312 to 2.460 are grouped into the Medium state and the values from 2.460 to 2.635 are grouped into the High state. The results in table 6 show that Random Forest is the best model when we predict 3 states of inflation rate (H, M or L) with the test data in-sample-accuracy of 82.55 percent. The second-best model is Linear SVM with the test data in-sample-accuracy of 72.59 percent.

Table 6: Comparison of 3-State Models Based on Accuracy-Inflation Rate

	Model	**Train**	**Test**
1	Logistic Regression	86.11	72.54
2	Decision Trees	80.55	72.54
3	**SVM- Linear**	**80.56**	**72.59**
4	SVM- Radial	88.88	72.51
5	L-1 Regularization	87.77	72.54
6	**Random Forest**	**91.2**	**82.55**

Source: Authors' Calculations.

IMPROVISING MODELS

We find that the baseline models could indeed be improved to obtain better levels of accuracy by appropriately tuning the hyperparameters and performing the previously mentioned data transformations. We approach toward getting more stable and consistent results. We identify that as we consider the lags of the data, there is correlation among the independent variables in the model. The Variance Inflation Factor (VIF) is very high and violating the assumption of multi-collinearity in the model. We also observe that there is a pattern in seasonality that is making independent variables difficult to explain the dependent variable accurately. Due to this, the algorithms have not been able to capture characteristic information encoded in the features, resulting in low accuracy level. To correct this, the first observation we make is that the inherent seasonality pattern must be removed, which is making the independent variables be poor explanators of the dependent variables.

The following steps were taken on both the datasets (Inflation/Unemployment Univariate Time Series) to avoid the inconsistency in our models:

(i) Initially Seasonal adjustment was made to the data using Weighted Moving Average Smoothing. This ensures the important patterns in the data are preserved, while leaving out noise.

(ii) Applying a log transformation to the data and make it conform to normality assumptions of various models. Also, enabling us better separation of the states.

(iii) Annotating the class labels for each quarter. The threshold for 2-state model is the mean and 3-state model are Mean ± standard deviation.

(iv) Once the states are labelled, then we take 3 previous time lagged values for each observation. These are essentially our 'features' or independent variables based on which, we will predict the state of the dependent/ target variable.

(v) After getting the features, a basis expansion is performed to convert the dataset into a high dimensional space. This is done mainly to add more information to the dataset when we only take the Univariate series. A 3-degree polynomial basis expansion is used giving us (n x 3 x 3) i.e., we now have 9 features.

(vi) Once the basis expansion is performed, we wish to extract a linear combination of features such that these new complex features are uncorrelated with each other. Basically, we want to eliminate the problem of multicollinearity. Hence, we need to extract uncorrelated features to ensure that this problem does not disrupt performance of the models.

(vii) We perform a principal component analysis and extract the first 3 principal components which explain the maximum variance in the data. This orthogonal transformation of the data makes the features linearly uncorrelated. Finally, the principal components will be our features of the model to fit the classification problem.

UNEMPLOYMENT DATA

Two-State Modelling

The two states are taken as High and Low based on the mean (1.814) of the log transformed data where [(1.312, 1.814] < (1.814, 2.341]] is the set division. The values from 1.312 to 1.814 are grouped into the Low state and the values from 1.814 to 2.341 are grouped into the High state. The results in table 7 show that KNN (k=3) is the best model when we predict 2 states of inflation rate (H or L) with test data in-sample-accuracy of 95.31 percent. The second-best model is Random Forest giving test data in-sample-accuracy of 94.74 percent.

Table 7: Comparison of Improved 2-State Models Based on Accuracy- UR

	Model	Train	Test
1	Logistic Regression	96.24	92.18
2	L-1 Regularization	93.98	92.18
3	Decision Trees	94.73	92.17
4	SVM- Linear	96.24	92.18
5	SVM- Radial	96.24	93.75
6	SVM- Sigmoid	96.24	92.17
7	**KNN**	**98.49**	**95.31**
8	GBM	99.24	95.31
9	**Random Forest**	**98.57**	**94.74**

Source: Authors' Calculations.

Three-State Modelling

The three states are High, Medium, and Low where [(1.312, 1.693] < (1.693, 1.936] < (1.936, 2.341]] is the set division. The values from 1.312 to 1.693 are grouped into the Low state, the values from 1.693 to 1.936 are grouped into the Medium state and the values from 1.936 to 2.341 are grouped into the High state. The results in table 8 show that KNN (k=3) is the best model when we predict 3 states of unemployment rate (H, M or L) with test data in-sample-accuracy of 96.875 percent. The second-best model is GBM with test data in-sample-accuracy of 95.31 percent.

Table 8: Comparison of Improved 3-State Models Based on Accuracy-UR

	Model	Train	Test
1	Logistic Regression	85.71	81.25
2	L-1 Regularization	85.71	84.37
3	Decision Trees	96.99	93.75
4	SVM- Linear	87.96	82.81
5	SVM- Radial	89.47	87.5
6	SVM- Sigmoid	88.72	84.37
7	**KNN**	**96.99**	**96.875**
8	**GBM**	**98.49**	**95.31**
9	Random Forest	95.49	90.62

Source: Authors' Calculations.

INFLATION DATA

Two-State Modelling

The two states are taken as High and Low based on the mean (2.388) of the log transformed data where [(2.267, 2.388] < (2.388, 2.597]] is the set division. The values from 2.267 to 2.388 are grouped into the Low state and the values from 2.388 to 2.597 are grouped into the High state. The results in table 9 show that Linear SVM is the best model when we predict 2 states of inflation rate (H or L) with test data in-sample-accuracy of 94.36 percent. The second-best model is KNN (k=3) with test data in-sample-accuracy of 91.54 percent.

Table 9: Comparison of Improved 2-State Models Based on Accuracy- Inflation Rate

	Model	**Train**	**Test**
1	Logistic Regression	92.12	90.14
2	L-1 Regularization	92.72	88.73
3	Decision Trees	92.72	88.73
4	**SVM- Linear**	**96.36**	**94.36**
5	SVM- Radial	95.75	92.95
6	SVM- Sigmoid	95.15	91.54
7	**KNN**	**95.75**	**91.54**
8	GBM	98.78	84.50
9	Random Forest	96.36	83.10

Source: Authors' Calculations**.**

Three-State Modelling

The three states are High, Medium, and Low where [(2.267, 2.358] < (2.358, 2.417] < (2.417, 2.597]] is the set division. The values from 2.267 to 2.358 are grouped into the Low state, the values from 2.358 to 2.417 are grouped into the Medium state and the values from 2.417 to 2.597 are grouped into the High state. The results in table 10 show that Random Forest is the best model when we predict 3 states of inflation rate (H, M or L) is Random Forest giving test data in-sample-accuracy of 88.73 percent. The second-best model is KNN (k=3) with test data in-sample-accuracy of 88.73 percent.

Table 10: Comparison of Improved 3-State Models Based on Accuracy- Inflation Rate

	Model	Train	Test
1	Logistic Regression	88.48	78.73
2	L-1 Regularization	88.48	84.5
3	Decision Trees	93.93	81.69
4	SVM- Linear	92.12	85.91
5	SVM- Radial	91.51	87.32
6	SVM- Sigmoid	90.3	87.32
7	**KNN**	**92.72**	**85.91**
8	GBM	98.18	81.69
9	**Random Forest**	**93.33**	**88.73**

Source: Authors' Calculations**.**

ECONOMETRIC vs ML MODELS

The table 11 shows that the ML models are more accurate in prediction than the Econometric models. The best model for Unemployment data was EGARCH (1,1,1) in TS with accuracy of 74.62 percent and 89.55 percent in 2-State and 3-State, respectively. K- Nearest Neighbors in ML with accuracy 95.31 percent and 96.87 percent in 2-State and 3-State, respectively. The best model for Inflation data was EGARCH (1,1,1) in TS with accuracy of 71.62 percent and 67.56 percent in 2-State and 3-State, respectively. In ML, Linear SVM with accuracy 94.36 percent in 2-State and Random Forest with accuracy 88.73 percent in 3-State.

Table 11: Comparison of Econometric and ML model

Models		**States**	**Train**	**Test**
	Unemployment Data			
ML	KNN (K=3)	2	98.49	95.31
	KNN (K=3)	3	96.99	96.875
TS	EGARCH (1, 1, 1)	2	63.90	74.62
	EGARCH (1, 1, 1)	3	94.73	89.55
	Inflation Data			
ML	Linear SVM	2	96.36	94.36
	Random Forest	3	93.33	88.73
TS	EGARCH (1, 1, 1)	2	87.27	71.62
	EGARCH (1, 1, 1)	3	60.0	67.56

Source: Authors' Calculations **.**

CONCLUSION and SUGGESTIONS

This study was highly motivated to investigate how Machine learning could be incorporated to a time series data and improve the efficiency of the predictions. Along the study, much interest was invoked towards converting the time series data into a classification problem and predict the states of the data with higher confidence. Various data transformations were done to improvise the models like log transformation, basis expansion and principal component analysis which ensured the models to be stable and consistent.

The models performed in this study showed that Machine learning models are superior compared to the traditional Econometric models for time series data. The best model for Unemployment data was EGARCH (1,1,1) in TS and K- Nearest Neighbors to predict both 2 states and 3 states in ML. The best model for Inflation data was EGARCH (1,1,1) in TS and Linear SVM, Random forest to predict 2 states and 3 states respectively in ML. Even though the ML models lack the interpretability and clarity in the exact internal process, these models have resulted exceptional in terms of accuracy in predictions. If our main focus is to only understand the effect and interpret the casual effect of the data, then Econometric modelling would be more suitable.
A potential extension of the states could be done in future studies based on the economic cycles of these macroeconomic variables. As we know that ML models perform better on high dimensional data and large datasets, we could consider studying how states will change on a monthly basis and daily basis. Inclusion of several macroeconomic indicators that affect these univariate time series would enable higher explainability of models. There are also recent developments to the field of econometric modelling and ML modelling. The performance of these new techniques could be explored. Incorporation of ML methods in the field of econometrics would be an efficient addition in terms of highly accurate and faster computation models.

REFERENCES

Bajracharya, D. (2011), "Econometric Modeling vs Artificial Neural Networks: A Sales Forecasting Comparison".

Beechey, M. and P.Osterholm (2018), "The Rise and Fall of US Inflation Persistence", 30th issue (September 2012), *International Journal of Central Banking.*

Bishop, C. M. (2016), "Pattern Recognition and Machine Learning", *Information Science and Statistics*, Softcover reprint of the original 1st ed. 2006 ed., Springer.

Brooks, C. (2014), "Introductory Econometrics for Finance", 3rd ed. *Cambridge University Press.*

Burman, J. P. (1965), "Moving Seasonal Adjustment of Economic Time Series", *Journal of the Royal Statistical Society*, Series A (General), 534-558.

Cerqueira, V., L. Torgo and C. Soares (2019), "Machine Learning vs Statistical Methods for Time Series Forecasting: Size Matters", arXiv preprint arXiv:1909.13316.

Enders, W. (2014), "Applied Econometric Time Series", Wiley Series in Probability and Statistics 4th ed., *Wiley*.

Floros, C. (2005), "Forecasting the UK Unemployment Rate: Model Comparisons", *International Journal of Applied Econometrics and Quantitative Studies*, 2(4), 57-72.

Gardner Jr, E. S. (1985), "Exponential Smoothing: The State of the Art", Journal of forecasting, 4(1), 1-28.

Gujarati, D., D. Porter and S. Gunasekar (2017), "Basic Econometrics 5th ed.", *Mcgraw*.

Gutierrez, P. A., C. Hervas-Martinez, F. J. Martinez-Estudillo and J. C. Fernandez (2009, June), "MultiLogistic Regression Using Initial and Radial Basis Function Covariates", *International Joint Conference on Neural Networks* (pp. 1067-1074), IEEE.

James, G., D.Witten, T.Hastie and R.Tibshirani (2013), "An Introduction to Statistical Learning: With Applications in R", Springer Texts in Statistics 1st ed. 2013, Corr. 7th printing 2017 ed., *Springer*.

Johnson, R. and D.Wichern (2018), "Applied Multivariate Statistical Analysis", Classic Version Pearson Modern Classics for Advanced Statistics Series 6th ed., *Pearson*.

Johnston, J. and J.Dinardo (1996), "Econometric Methods 4th ed.", *McGraw-Hill*/Irwin. Karanasos, M. E. N. E. L. A. O. S., Karanassou, M. A. R. I. K. A., *Fountas*, S. T. I.

L. I. A. N. O. S. (2004), "Analyzing US inflation by a GARCH Model with Simultaneous Feedback", *WSEAS Transactions on Information Science and Applications*, 1(2), 767-772.

Kreiner, A. and J. Duca (2020), "Can Machine Learning on Economic Data Better Forecast the Unemployment Rate?", *Applied Economics Letters*, 27(17), 1434-1437.

Maehashi, K. and M. Shintani (2020), "Macroeconomic Forecasting Using Factor Models and Machine Learning: An application to Japan", *Journal of the Japanese and International Economies*, 58, 101104.

Moshiri, S. and L. Brown (2004), "Unemployment Variation Over the Business Cycles: A Comparison of Forecasting Models", *Journal of Forecasting*, 23(7), 497-511.

Müller, A. C. and S.Guido (2016), "Introduction to Machine Learning with Python: A Guide for Data Scientists 1st ed.", *O'Reilly Media*.

OECD (2021), Inflation (CPI) (indicator). doi: 10.1787/eee82e6e-en OECD (2021), Unemployment rate (indicator). doi: 10.1787/52570002-en

Olmedo, E. (2014), "Forecasting Spanish Unemployment Using Near Neighbour and Neural Net Techniques", *Computational Economics*, 43(2), 183-197.

Sermpinis, Georgios, Charalampos Stasinakis, Konstantinos Theofilatos, and Andreas Karathanasopoulos (2014), "Inflation and Unemployment Forecasting with Genetic Support Vector Regression", *Journal of Forecasting* 33, no. 6 :471-487.

Ülke, V., A.Sahin and A.Subasi, (2018), "A Comparison of Time Series And Machine Learning Models For Inflation Forecasting: Empirical Evidence From The USA", *Neural Computing and Applications*, 30(5), 1519-1527.

VanderPlas, J. (2016), "Python Data Science Handbook: Essential Tools for Working with Data 1st ed.", *O'Reilly Media.*

W. W., Chen and R. S. Deo (2004), "Power Transformations to Induce Normality and Their Applications", *Journal of the Royal Statistical Society*, Series B (Statistical Methodology), 66(1), 117-130.

APPENDIX

Additional Tables- Model Results

Model Results- Inflation data

ARIMA in Range

Table A.1: Range of ARIMA Model Results- IR

Model	MSE	AIC
ARIMA(0, 1, 0)	0.42	197.095
ARIMA(0, 1, 1)	0.42	185.785
ARIMA(0, 1, 2)	0.42	176.312
ARIMA(0, 1, 3)	0.42	163.187
ARIMA(0, 1, 4)	0.41	164.903
ARIMA(0, 2, 0)	0.41	307.948
ARIMA(0, 2, 1)	0.41	203.350
ARIMA(0, 2, 2)	0.40	192.861
ARIMA(0, 2, 3)	0.40	183.591
ARIMA(0, 2, 4)	0.41	170.124
ARIMA(1, 1, 0)	0.42	194.106
ARIMA(1, 1, 1)	0.42	185.877
ARIMA(1, 1, 2)	0.43	169.931
ARIMA(1, 1, 3)	0.41	164.967
ARIMA(1, 1, 4)	0.42	165.136
ARIMA(1, 2, 0)	0.44	285.689
ARIMA(1, 2, 1)	0.41	200.682
ARIMA(1, 2, 3)	0.41	177.142
ARIMA(2, 1, 0)	0.40	167.782
ARIMA(2, 1, 1)	0.40	167.300
ARIMA(2, 1, 2)	0.42	167.755
ARIMA(2, 1, 3)	0.42	166.712
ARIMA(2, 1, 4)	0.42	163.165
ARIMA(2, 2, 0)	0.40	215.687
ARIMA(2, 2, 1)	0.39	175.026
ARIMA(2, 2, 2)	0.39	174.422
ARIMA(2, 2, 3)	0.40	174.937
ARIMA(3, 1, 0)	0.40	167.839
ARIMA(3, 1, 1)	0.42	166.588
ARIMA(3, 1, 2)	0.42	168.586
ARIMA(3, 1, 3)	0.42	168.699
ARIMA(3, 1, 4)	0.43	164.824
ARIMA(3, 2, 0)	0.40	205.224
ARIMA(3, 2, 1)	0.39	174.971
ARIMA(3, 2, 2)	0.40	173.778
ARIMA(3, 2, 3)	0.40	175.774
ARIMA(3, 2, 4)	0.40	175.633

Source: Authors' Calculations.

EGARCH

Table A.2: EGARCH (1,1,1) Model Results- IR

Dep. Variable:	Inflation_diff		**R-squared:**	0.218	
Mean Model:	AR		**Adj. R-squared:**	0.205	
Vol Model:	EGARCH		**Log-Likelihood:**	-62.6456	
Distribution:	Normal		**AIC:**	139.291	
Method:	Maximum Likelihood		**BIC:**	159.255	
Df Residuals:	197		**No. Observations:**	180	
Df Model:	3				
Mean Model					
	Coefficient	**Std err**	**t**	**P>|t|**	**95.0 percent Conf. Int.**
Constant	0.0247	1.10E-04	223.84	0	[2.445e-02,2.488e-02]
AR.1	-0.3222	3.94E-03	-81.7	0	[-0.330, -0.315]
AR.2	-0.366	8.16E-03	-44.839	0	[-0.382, -0.350]
Volatility Model					
	Coefficient	**Std err**	**t**	**P>|t|**	**95.0 percent Conf. Int.**
Omega	-0.0103	1.04E-11	-9.84E+08	0	[-1.027e-02,-1.027e-02]
Alpha[1]	-0.2147	2.89E-03	-74.339	0	[-0.220, -0.209]
Gamma[1]	0.0867	1.57E-03	55.361	0	[8.364e-02,8.977e-02]
Beta[1]	0.9999	1.75E-09	5.71E+08	0	[1.000, 1.000]

Source: Authors' Calculations.

Model Results- Unemployment Rate

ARIMA in Range

Table A.3: Range of ARIMA Model Results- UR

Model	MSE	AIC
ARIMA(0, 0, 0)	4.39	352.5704310
ARIMA(0, 0, 1)	3.77	238.3801743
ARIMA(1, 0, 0)	3.55	99.80020027
ARIMA(1, 0, 1)	3.54	60.73036788
ARIMA(1, 0, 2)	3.55	48.10853041
ARIMA(1, 0, 3)	3.54	47.37341208
ARIMA(1, 0, 4)	3.56	41.34758614
ARIMA(2, 0, 0)	3.54	39.76390202
ARIMA(2, 0, 1)	3.54	41.72994692
ARIMA(2, 0, 2)	3.54	43.61671105
ARIMA(2, 0, 3)	3.54	45.47011416
ARIMA(2, 0, 4)	3.56	42.45962296

Source: Authors' Calculations.

EGARCH

Table A.4: EGARCH (1,1,1) Model Results- UR

Dep. Variable:	Unemployment_rate	**R-squared:**	0.956
Mean Model:	AR	**Adj. R-squared:**	0.955
Vol Model:	EGARCH	**Log-Likelihood:**	5.17084
Distribution:	Normal	**AIC:**	3.65831
Method:	Maximum Likelihood	**BIC:**	21.8945
Df Residuals:	147	**No. Observations:**	150
Df Model:	3		

Mean Model

	Coefficient	**Std err**	**t**	**P>|t|**	**95.0 percent Conf. Int.**
Constant	0.3783	9.97E-02	3.796	1.47E-04	[0.183, 0.574]
AR.1	1.5777	5.59E-02	28.224	3.00E-175	[1.468, 1.687]
AR.2	-0.6338	4.46E-02	-14.206	8.37E-46	[-0.721, -0.546]

Volatility Model

	Coefficient	**Std err**	**t**	**P>|t|**	**95.0 percent Conf. Int.**
Omega	-1.4596	0.56	-2.606	9.17E-03	[-2.557, -0.362]
Alpha[1]	-0.0387	0.191	-0.203	0.839	[-0.413, 0.336]
Gamma[1]	0.6558	0.176	3.717	2.02E-04	[0.310, 1.002]
Beta[1]	0.503	0.2	2.51	1.21E-02	[0.110, 0.896]

Source: Authors' Calculations.

MSE Monographs

* Monograph 34/2015
Farm Production Diversity, Household Dietary Diversity and Women’s BMI: A Study of Rural Indian Farm Households
Brinda Viswanathan

* Monograph 35/2016
Valuation of Coastal and Marine Ecosystem Services in India: Macro Assessment
K. S. Kavi Kumar, Lavanya Ravikanth Anneboina, Ramachandra Bhatta, P. Naren, Megha Nath, Abhijit Sharan, Pranab Mukhopadhyay, Santadas Ghosh, Vanessa da Costa and Sulochana Pednekar

* Monograph 36/2017
Underlying Drivers of India’s Potential Growth
C.Rangarajan and D.K. Srivastava

* Monograph 37/2018
India: The Need for Good Macro Policies *(4th Dr. Raja J. Chelliah Memorial Lecture)*
Ashok K. Lahiri

* Monograph 38/2018
Finances of Tamil Nadu Government
K R Shanmugam

* Monograph 39/2018
Growth Dynamics of Tamil Nadu Economy
K R Shanmugam

* Monograph 40/2018
Goods and Services Tax: Revenue Implications and RNR for Tamil Nadu
D.K. Srivastava, K.R. Shanmugam

* Monograph 41/2018
Medium Term Macro Econometric Model of the Indian Economy
D.K. Srivastava, K.R. Shanmugam

* Monograph 42/2018
A Macro-Econometric Model of the Indian Economy Based on Quarterly Data
D.K. Srivastava

* Monograph 43/2019
The Evolving GST
Indira Rajaraman

MSE Working Papers
Recent Issues

* Working Paper 198/2020
  Inventory Cycles and Business Cycles – Has the relationship lost its importance over the years: A Time-Varying Parameter Approach using U.S. Data
  Parijat Maitra and Naveen Srinivasan
* Working Paper 199/2020
  From Income to Household Welfare : Lessons from Refrigerator Ownership in India
  Sowmya Dhanaraj, Vidya Mahambare and Poonam Munjal
* Working Paper 200/2020
  Double Burden of Malnutrition in India : Decadal Changes among Adult Men and Women
  Brinda Viswanathan and Archana Agnihotri
* Working Paper 201/2020
  Understanding the Water Crisis in India: Application of Causal Loop Modelling to Examine the Environment-Economy Interlinkage across Sectors
  Ashwin Ram Sridharan and Zareena Begum Irfan
* Working Paper 202/2020
  Effect of Farming Systems for Nutrition on Nutritional Intakes: A Study of Two Regions in India
  Nithya D.J., S Raju, R V Bhavani, Akshaya Kumar Panda, Rupal D.Wagh and Brinda Viswanathan
* Working Paper 203/2020
  Going Beyond Gold: Can Equities be Safe-Haven?
  Parthajit Kayal and Janani Sri SG
* Working Paper 204/2020
  Impact of Climate Change on Economic Growth: A Case Study of India
  Medhavi Sandhani, Anubhab Pattanayak, K.S. Kavi Kumar
* Working Paper 205/2021
  Crop Diversity and Resilience to Droughts: Evidence from Indian Agriculture
  S. Madhumitha, K.S. Kavi Kumar, Anubhab Pattanayak

---

* Working papers are downloadable from MSE website http://www.mse.ac.in
$ Restricted circulation