Combining Statistical and Machine Learning Methods for Unemployment Rate Forecasting

by

Dr. Tanujit Chakraborty

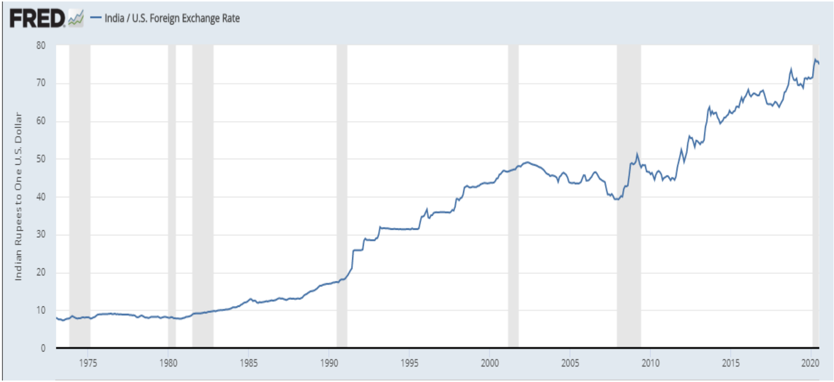
Ph.D from Indian Statistical Institute, Kolkata, India.
Currently a Postdoctoral Fellow, IIIT Delhi, India.

Webpage : www.ctanujit.org/

- Time series is a set of observations, each one being recorded at a specific time. (e.g., Annual GDP of a country, Sales figure, etc)
- Discrete time series is one in which the set of time points at which observations are made is a discrete set. (e.g., All above including irregularly spaced data)
- Continuous time series are obtained when observations are made continuously over some time intervals. (e.g., ECG graph)
- Forecasting is estimating how the sequence of observations will continue into the future. (e.g., Forecasting of major economic variables like GDP, Unemployment, Inflation, Exchange rates, Production and Consumption)
- Forecasting is very difficult, since it's about the future ! (e.g., forecasts of daily cases of COVID-19)

A Forecasting Problem: India / U.S. Foreign Exchange Rate (EXINUS)

- Source : FRED ECONOMICS DATA (Shaded-areas indicate US recessions)
- Units : Indian Rupees to One U.S. Dollar, Not Seasonally Adjusted
- Frequency : Monthly (Averages of daily figures)



Time Series Components

- Trend (T_t) : pattern exists when there is a long-term increase or decrease in the data.
- Seasonal (S_t) : pattern exists when a series is influenced by seasonal factors (e.g., the quarter of the year, the month, or day of the week).
- Cyclic (C_t) : pattern exists when data exhibit rises and falls that are not of fixed period (duration usually of at least 2 years).
- Decomposition : $Y_t = f(T_t, S_t, C_t, I_t)$, where Y_t is data at period t and I_t is irregular component at period t.
- Additive decomposition : $Y_t = T_t + S_t + C_t + I_t$.
- Multiplicative decomposition : $Y_t = T_t \times S_t \times C_t \times I_t$.
- A stationary series is: roughly horizontal, constant variance and no patterns predictable in the long-term.

- Additive model is more appropriate if magnitude of seasonal fluctuations does not vary with level.
- Multiplicative model is more appropriate if seasonal fluctuations are proportional to level of series.
- In Economic Series, multiplicative decomposition is more prevalent.
- Alternative : Use a Box-Cox transformation and then use additive decomposition.
- Logs turn multiplicative relationship into an additive relationship :

$$Y_t = T_t \times S_t \times C_t \times I_t \implies \log Y_t = \log T_t + \log S_t + \log C_t + \log I_t$$

Unemployment Rate Forecasting : A Hybrid Approach

Related Publication:

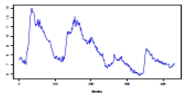
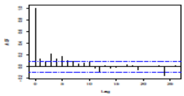
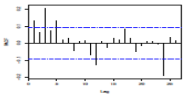
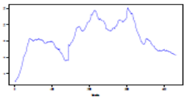
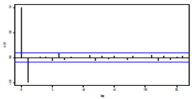
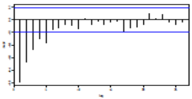
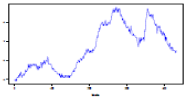
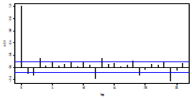
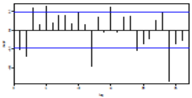
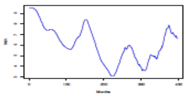
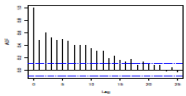
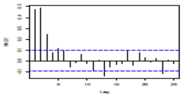
Tanujit Chakraborty et al. "Unemployment Rate Forecasting : A Hybrid Approach". *Computational Economics*, Vol. 57, Pg. 183–201 (2021). **Read Online**

Unemployment Rate Forecasting Problem

- Unemployment has always been a very focused issue causing a nation as a whole to lose its economic and financial contribution.
- Unemployment rate prediction of a country is a crucial factor for the country's economic and financial growth planning and a challenging job for policymakers.
- The unemployment rate represents the number of unemployed as a percentage of the labor force.
- Forecasting unemployment rate can be defined as the projected value for the number of unemployed people as a percentage of the labor force.
- Traditional stochastic time series models (ARIMA), as well as modern nonlinear time series techniques (Neural Net), were employed for unemployment rate forecasting previously.
- These macroeconomic data sets are mostly nonstationary and nonlinear in nature.
- Thus, it is atypical to assume that an individual time series forecasting model can generate a white noise error.

Unemployment Rate data sets

- Source : FRED Economic Data sets and OECD data repository
- Graphical analysis of training unemployment rate data sets (seasonally adjusted monthly data sets) for different countries and its corresponding ACF and PACF plots.

Country	Training data	ACF plot	PACF plot
Canada			
Germany			
Japan			
Netherlands			

Previous works on Unemployment rate forecasting

- The autoregressive integrated moving average (ARIMA) is extensively utilized in constructing a forecasting model for unemployment data (Funke, Journal of Forecasting, 1992). Read online
- ARIMA cannot be utilized to produce an accurate model for forecasting nonlinear time series. The threshold autoregressive (TAR) model, a classical nonlinear time series model, outperformed the linear time series models for forecasting the USA unemployment rate data set (Montgomery et al., Journal of American Statistical Association, 1998). Read online
- ANN is found to be the most accurate in forecasting unemployment over the asymmetric business cycle for the USA, Canada, UK, France, and Japan (Pelaez, Business Economics, 2006). Read online
- For short term forecasting of seasonally adjusted monthly USA unemployment data sets, nonlinear models outperform the linear models. But there is an asymmetry in unemployment rate forecasting and its elimination is bound to be challenging (Nagao et al., Finance Research Letters, 2019). Read online
- Fractional ARIMA (FARIMA or ARFIMA) is a suitable model when long memory exists in a time series and has been applied successfully for predicting unemployment (Kartis, Computational Economics, 2020). Read online

AutoRegressive Integrated Moving Average (ARIMA) Model

- The ARIMA model, introduced by Box and Jenkins (1976) (Read online), is a linear regression model indulged in tracking linear tendencies in stationary time series data.
- AR: autoregressive (lagged observations as inputs) I: integrated (differencing to make series stationary) MA: moving average (lagged errors as inputs).
- The model is expressed as ARIMA(p, d, q) where p, d, and q are integer parameter values that decide the structure of the model.
- More precisely, p and q are the order of the AR model and the MA model respectively, and parameter d is the level of differencing applied to the data.
- The mathematical expression of the ARIMA model is as follows:

  $$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q},$$

  where y_t is the actual value, ε_t is the random error at time t, ϕ_i and θ_j are the coefficients of the model.
- It is assumed that ε_{t-l} $(\varepsilon_{t-l} = y_{t-l} - \hat{y}_{t-l})$ has zero mean with constant variance, and satisfies the i.i.d condition.
- Three basic Steps: Model identification, Parameter Estimation, and Diagnostic Checking.

Artificial Neural Networks (ANN) Model

- Neural nets are based on simple mathematical models of the brain, used for sophisticated nonlinear forecasting.
- ANN is composed of several perceptron-like units arranged in multiple layers (input layer, one or more hidden layer, and an output layer).
- Each hidden node computes a nonlinear transformation of its incoming inputs: Weighted linear combination followed by a nonlinear activation function.
- Learn the parameters by minimizing some loss function.
- Gradient descent backpropagation is commonly used to do this efficiently (Rumelhart, D. E., Hinton, G.E., and Williams, R. J., 1986) (Read online).

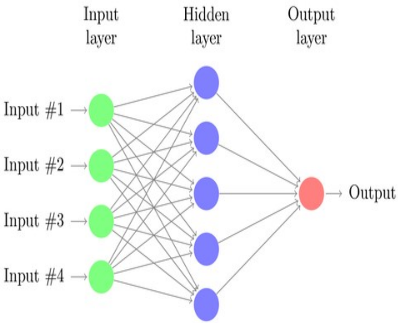


Fig: Also called a Feedforward Neural Network (FFNN) Model with one hidden layer with 5 hidden units (since there is no backward connections between layers, viz., no loops).

AutoRegressive Neural Network (ARNN) Model

- ARNN overcomes the problems of fitting ANN for time series data sets like the choice on the number of hidden neurons, and its black box nature (Faraway and Chatfield, Journal of Royal Statistical Society - C, 1998) (Read online).
- ARNN model is a nonlinear time series model which uses lagged values of the time series as inputs to the neural network.
- ARNN(p, k) is a feed-forward neural network having one hidden layer with p lagged inputs and k nodes in the hidden layer.
- Thus, ARNN model with one hidden layer with the following mathematical form:

$$\hat{x}_t = \phi_0 \left\{ w_{c_0} + \sum_h w_{h_0} \phi_h \left(w_{c_h} + \sum_i w_{i_h} x_{t-j_i} \right) \right\}$$

  where $\{w_{c_h}\}$ denotes the the connecting weights and ϕ_i is the activation function.
- An ARNN(p, k) model uses p as the optimal number of lags (calculated based on the AIC value) for an AR(p) model and k is set to $k = \left[\frac{(p+1)}{2}\right]$ for non-seasonal data sets.

Motivation for Hybrid Techniques

- A Forecaster wants the ARIMA model error series to be composed by i.i.d. random chocks or unpredictable or unsystematic terms with zero mean and constant variance, reflecting the piece of variability for which no reduction is possible (Proietti, Computational Statistics & Data Analysis, 2002) (Read online).

- However, due to model mis-specification or to disturbances introduced in the stochastic process after forecasters elaboration, this (white noise) assumption may be violated during application phase.

- If the information underlying the error series is modeled, the performance of the original forecaster can be improved (Zhang, Neurocomputing, 2002) (Read online).

- Also, various hybrid models were developed for stock price forecasting, see (Pai and Lin, Omega, 2005) (Read online) and (Hajirahimi and Khashei, Neural Processing Letters, 2020) (Read online).

- Forecast combinations: Clemen (International journal of forecasting, 1989) stated "The results have been virtually unanimous: combining multiple forecasts leads to increased forecast accuracy... In many cases one can make dramatic performance improvements by simply averaging the forecasts" (Read online).

Proposed Hybrid Model

- An integrated approach based on linear and non-linear models is proposed that can predict unemployment more accurately.
- The hybrid forecasting method is asymptotically stationary and applied to Canada, Germany, Japan, Netherlands, New Zealand, Sweden and Switzerland unemployment data sets.
- The complexity in the time series is composed of a linear component (L) and a nonlinear component (N). We assume two models to analyse such a time series, an additive model ($L + N$) and a multiplicative model ($L \times N$).
- Step 1: The ARIMA model is used to model the linear component of time series.
- Step 2: The residuals (errors) from the ARIMA model are obtained. The residuals are modeled by the ARNN model.
- Step 3: the forecasts of the hybrid model are obtained by adding (multiplying) the forecasted values of linear (ARIMA) and nonlinear (ARNN) models.

Proposed Additive Hybrid Model

- The proposed hybrid ARIMA-ARNN model (Y_t) can be represented as $Y_t = L_t + N_t$, where L_t is the linear part and N_t is the nonlinear part of the hybrid model.
- Both L_t and N_t are estimated from the data set.
- Let, $\hat{L_t}$ be the forecast value of the ARIMA model at time t and ε_t represent the residual at time t as obtained from the ARIMA model, i.e., $\varepsilon_t = Z_t - \hat{Y_t}$.
- The residuals are modeled by the ARNN model and can be represented as follows $\varepsilon_t = f(\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-n}) + \varsigma_t$, where f is a nonlinear function modeled by the ARNN approach and ς_t is the random error.
- Therefore, the combined forecast is $\hat{Y_t} = \hat{L_t} + \hat{N_t}$, where, $\hat{N_t}$ is the forecast value of the ARNN model.

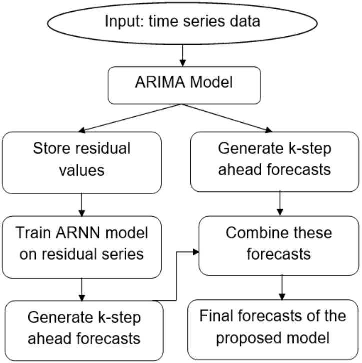


Fig: Graphical Representation of Hybrid ARIMA + ARNN Model

Proposed Multiplicative Hybrid Model

- $Z_t = L_t \times N_t$, where L_t is the linear part and N_t is the nonlinear part of the hybrid model.
- Both L_t and N_t are estimated from the data set.
- Let, $\hat{L_t}$ be the forecast value of the ARIMA model at time t and ε_t represent the residual at time t as obtained from the ARIMA model, i.e., $\varepsilon_t = Z_t / \hat{L_t}$.
- The residuals are modeled by the ARNN model and can be represented as follows $\varepsilon_t = f(\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-n}) + \varsigma_t$, where f is a nonlinear function modeled by the ARNN approach and ς_t is the random error.
- Therefore, the combined forecast is $\hat{Z_t} = \hat{L_t} \times \hat{N_t}$, where, $\hat{N_t}$ is the forecast value of the ARNN model.

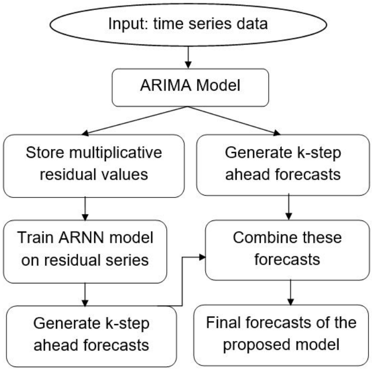


Fig: Graphical Representation of Hybrid ARIMA $\times$ ARNN Model

Asymptotic Stationarity of the Proposed Additive Model

ARIMA model has the in-built mechanism to transform a nonstationary time series into a stationary one and then it models the remainder by a stationary process. This is done by simple differencing to transform nonstationary ARIMA into stationary.

Consider the stochastic difference equation:

$$\varepsilon_t \quad = \quad f(\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-p}, \theta) + \varsigma_t, \tag{0.1}$$

where ς_t is an i.i.d. white noise and $f(., \theta)$ is a feedforward neural network with weight parameter θ. This is called an ARNN process of order p and has k hidden nodes in its one hidden layer. Thus, we refer the model as ARNN(p, k) model.

We consider the following architecture:

$$f(\underline{\varepsilon}) \quad = \quad c_0 + \sum_{i=1}^{k} w_i \sigma\left(a_i + b_i' \underline{\varepsilon}\right) \tag{0.2}$$

Let ε_t denote a time series generated by a nonlinear autoregressive process as defined in (0.1). Let $E(\varepsilon_t) = 0$, then f equals to the conditional expectation $E\left(\varepsilon_t | \varepsilon_{t-1}, ..., \varepsilon_{t-p}\right)$ is the best prediction for ε_t in the L_2-minimization sense.

We use the following notation:

$$z_{t-1} = \left(\varepsilon_{t-1}, ..., \varepsilon_{t-p}\right)' ; F(z_{t-1}) = \left(f(z_{t-1}), \varepsilon_{t-1}, ..., \varepsilon_{t-p+1}\right)' ; \hat{\varsigma}_t = \left(\varsigma_t, 0, ..., 0\right)'$$

Then we can write scalar AR(p) model in (0.1) as a first-order vector model,

$$z_t = F(z_{t-1}) + \hat{\varsigma}_t \tag{0.3}$$

with $z_t, \hat{\varsigma}_t \in \mathbb{R}^p$.

Definition (Geometric ergodicity, Chan & Tong, 1985, AAP)

Let $\{z_t\}$, a markov chain, is said to be geometrically ergodic if there exists a probability measure $\Pi(A) = \lim_{t\to\infty} P(\varepsilon_t \in A)$ on the state space $(\mathbb{R}^p, \mathbb{B}, \mathbb{P})$, where $\mathbb{B}$ are Borel set on $\mathbb{R}^p$ and $\mathbb{P}$ be the Lebesgue measure, and for $\rho > 1$ and for all $z \in \mathbb{R}^p$,

$$\lim_{n\to\infty} \rho^n \| P\{z_{t+n} \in A | z_t = z\} - \Pi(A) \| = 0$$

where $\|.\|$ denotes the total variation and $P\{z_{t+n} \in A | z_t = z\}$ denote the probability of going from point z to set $A \in \mathbb{B}$ in n steps.

If the markov chain is geometrically ergodic then its distribution will converge to Π and the corresponding time series will be called asymptotically stationary (Chan & Tong, 1985, Advances in Applied Probability) (Read online).

It is also important to note that all neural network activation functions (like logistics or tan-hyperbolic) are continuous and compact functions and must have a bounded range.

Lemma (Chakraborty et al., Computational Economics, 2020)

Suppose $\{z_t\}$ is defined as in (0.1) and (0.3), F be a compact set can be decomposed as $F = F_m + F_n$, and the following conditions hold:
(i) $F_m(.)$ is continuous and homogeneous and $F_n(.)$ is bounded;
(ii) $E|\varsigma_t| < \infty$ and probability distribution function of ς_t is positive everywhere in $\mathbb{R}$;
then $\{z_t\}$ is geometrically ergodic.

Theorem (Chakraborty et al., Computational Economics, 2020)

Let $E|\varsigma_t|^{1+\delta} < \infty$ for all $\delta > 1$ and the probability density function of ς_t is positive everywhere in $\mathbb{R}$ and $\{\varepsilon_t\}$ and $\{z_t\}$ are defined as in (0.1) and (0.3). Then if f is a nonlinear neural network as defined in (0.2), then $\{z_t\}$ is geometrically ergodic and $\{\varepsilon_t\}$ is asymptotically stationary.

Theoretical results on asymptotic stationarity is important for predictions over larger intervals of time, for example, one might train the network on an available sample and then use the trained network to generate new data with similar properties.

The asymptotic stationarity guarantees that the proposed model cannot have growing variance with time.

Performance metrics such as mean absolute error (MAE), root mean square error (RMSE), and Mean Absolute Percent Error (MAPE) are used to evaluate the performances of different forecasting models for the unemployment rate data sets:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}; \; MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|; \; MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|,$$

where y_i is the actual output, $\hat{y}_i$ is the predicted output, and n denotes the number of data points.

By definition, the lower the value of these performance metrics, the better is the performance of the concerned forecasting model.

Table: Performance metrics for different forecasting models on the Canadian unemployment rate (monthly) data

Model	1-Year ahead forecast			3-Year ahead forecast		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
ARIMA	0.133	0.115	1.623	0.847	0.685	9.708
ANN	0.137	0.117	1.095	0.837	0.614	9.365
ARNN	0.126	0.113	1.084	0.801	0.613	9.247
SVM	0.273	0.248	1.915	0.998	0.740	10.92
Hybrid ARIMA-SVM	0.145	0.135	1.135	0.835	0.711	9.595
Hybrid ARIMA-ANN	0.118	0.108	1.017	0.638	0.615	8.387
Proposed Hybrid ARIMA-ARNN	**0.106**	**0.098**	**0.838**	**0.627**	**0.601**	**8.017**

Table: Performance metrics for different forecasting models on the Germany unemployment rate (monthly) data

Model	1-Year ahead forecast			3-Year ahead forecast		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
ARIMA	0.098	0.092	1.490	0.361	0.303	5.177
ANN	0.127	0.120	4.295	0.564	0.505	7.394
ARNN	0.104	0.099	6.783	0.569	0.533	6.365
SVM	0.101	0.099	1.594	0.566	0.509	6.272
Hybrid ARIMA-SVM	0.090	0.089	1.537	0.360	0.305	5.120
Hybrid ARIMA-ANN	0.082	0.096	1.558	0.306	0.297	4.243
Proposed Hybrid ARIMA-ARNN	**0.077**	**0.071**	**1.068**	**0.300**	**0.291**	**4.156**

Table: Performance metrics for different forecasting models on the Japan unemployment rate (monthly) data

Model	1-Year ahead forecast			3-Year ahead forecast		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
ARIMA	0.300	0.272	6.985	0.586	0.502	14.621
ANN	0.260	0.229	6.518	0.469	0.413	11.627
ARNN	0.269	0.236	6.713	0.483	0.426	11.934
SVM	0.221	0.192	5.902	0.469	0.421	10.334
Hybrid ARIMA-SVM	0.226	0.187	5.377	0.386	0.336	9.722
Hybrid ARIMA-ANN	**0.191**	**0.172**	**4.987**	**0.370**	**0.321**	**9.317**
Proposed Hybrid ARIMA-ARNN	0.222	0.186	5.348	0.371	0.326	9.468

Table: Performance metrics for different forecasting models on the Netherlands unemployment rate (monthly) data

Model	1-Year ahead forecast			3-Year ahead forecast		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
ARIMA	0.144	0.119	2.003	0.306	0.270	5.671
ANN	0.228	0.174	2.906	1.304	1.048	17.935
ARNN	0.249	0.192	3.177	0.938	0.784	14.518
SVM	0.228	0.174	2.906	1.304	1.048	17.935
Hybrid ARIMA-SVM	0.145	0.120	2.023	0.308	0.272	5.706
Hybrid ARIMA-ANN	0.143	0.118	2.002	0.306	0.270	5.668
Proposed Hybrid ARIMA-ARNN	**0.140**	**0.114**	**1.192**	**0.300**	**0.264**	**5.529**

Forecast Curves

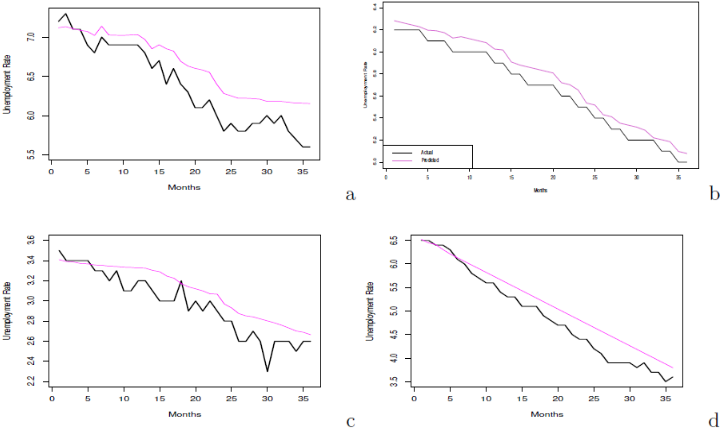


Fig: Actual vs Predicted (based on hybrid ARIMA-ARNN model) forecasts for the test data sets of the Canada (a), Germany (b), Japan (c), Netherlands (d)

- In practice, it is often challenging to determine whether a time series under study is generated from a linear or nonlinear underlying process.
- The additive and multiplicative hybridization approach studies the relationship between linear and nonlinear components of the econometric time series.
- The multiplicative method is appropriate for explaining variations of economic and business data where there are interactions between linear and nonlinear time series.
- But both the models suppose that the residuals from the linear model will contain only the nonlinear relationship. However, one may not always guarantee that the residuals of the linear component may comprise valid nonlinear patterns.
- Also, these models supposed that the linear and nonlinear patterns of a time series can be separately modeled by different models and then the forecasts can be combined together and this may degrade performance, if it is not true.

Forecasting models for different frequencies

- Models for annual data: Exponential Smoothing State Space (ETS), ARIMA, Dynamic regression, ARNN and LSTM model.

- Models for quarterly, monthly data: ETS, ARIMA/SARIMA, Dynamic regression, Seasonal and Trend decomposition using Loess (STL), ARNN and LSTM model.

- Models for weekly, daily, hourly data: ARIMA/SARIMA, Dynamic regression, STL.

- Other alternatives to the proposed Hybrid Models :

  - Bagging and Ensemble time series approaches (Hands-On Ensemble Learning with R: Chapter 11) (Read online).

  - Bayesian Model Averaging Methods for Forecasting (Bencivelli et al., Empirical Economics, 2017) (Read online).

  - Nowcasting of COVID-19 confirmed cases: Foundations, trends, and challenges (Chakraborty et al., Modelling, Control and Drug Development for COVID-19 Outbreak Prevention, 2021) (Read online).

Future Challenges

- Over the last four decades, the unemployment rates for most of the countries considered in this study had no consistent trend at all and has asymmetrical cyclical movements.
- On average, the 'best' short-term and long-term forecasts of monthly unemployment rate data sets are obtained using the proposed hybrid model in our paper.
- Future works :

  (a) An immediate extension of this work is to see the application of the model for seasonal unemployment rate data sets.

  (b) Another possible extension of the work would be to expand this hybrid approach for multivariate time series forecasting problems that arise in various macroeconomic problems.

  (c) Designing some new hybrid / ensemble / Bayesian Model Averaging techniques for Macroeconomic Forecasting.


Thank You