Short Time forecasting of U.S. unemployment rate: A Comparison

Amit Manchanda Anuma Rathore
14116013 14116018
Indian Institute of Technology, Roorkee
{amitmanchanda1995, anumarathore}@gmail.com

ABSTRACT

Some major economic crisis like The Great Depression-1929 and Global Recession 2008, make the economic forecasting a necessity to predict the behaviour early when structural changes occur haphazardly. The paper presents a comparison, in terms of root mean square error (RMSE), of different linear and nonlinear time series models designed for short term forecasting of U.S. unemployment rate, and demonstrate the correlation of future values with past data.

1. INTRODUCTION

Unemployment is an important measure of market (labour) performance. It is an important macroeconomic factor and its forecasting can help policymakers in devising proper decisions regarding the future state of the economy and country. One of the approaches used for future unemployment rate prediction is based on past time-series properties of the same.

This paper presents a study of the forecasts for the monthly U.S. unemployment rate between 1948-1981 using various time series models. Comparisons between these forecasting methods are made to further enhance our understanding of the strengths and deficiencies of these methods. These comparisons provide several insights about using the various methods in forecasting the unemployment rate that can be exploited by future researchers.

We have concentrated the study on understanding the behaviour of natural unemployment rates disaggregated by age and gender [2.6]. The unemployment rates dataset follows asymmetric business cycle exhibiting nonlinearity. It rises quickly in economic downturns and steeps through in expansion. Hence, we can conclude that nonlinear forecasting methods outperform conventional linear methods. These improvements are, however, sensitive to stationary series transformations.

Using monthly observations to forecast quarterly unemployment rate improves performance as compared to using only quarterly data because of the more number of points between them which help in generalizing the data.

The code is made open source along with the open source dataset.*

In the forthcoming section, we describe our Datasets, followed by the comprehensive theoretical aspects in Models Review in Section 3. The experimental setup, observed results and their discussion are presented under Section 4. Section 5 closes with the conclusion.

2. DATASETS

The past datasets contain an enormous amount of information that comes on a continual time basis which classical econometrics seldom tap into that can help us to predict a certain range of values in future. It is particularly welcomed when there are risks of an economic crisis, vitals fluctuate and sound economic decisions have to be made. It also helps to estimate how unemployment is affected after the recovery period. In this paper, we re-access the data using ARIMA model, SVR mode, ELM model and Neural Network model parallely to observe and forecast relevant new information from

*https://github.com/amitmanchanda1995/ANN-Course-Project previous projections. A forecasting competition between the above three models is done on the basis of Root-Mean-Squared Error (RMSE). The dataset contains monthly US unemployment figures [1] between 1948 - 1981 in 4 categories:

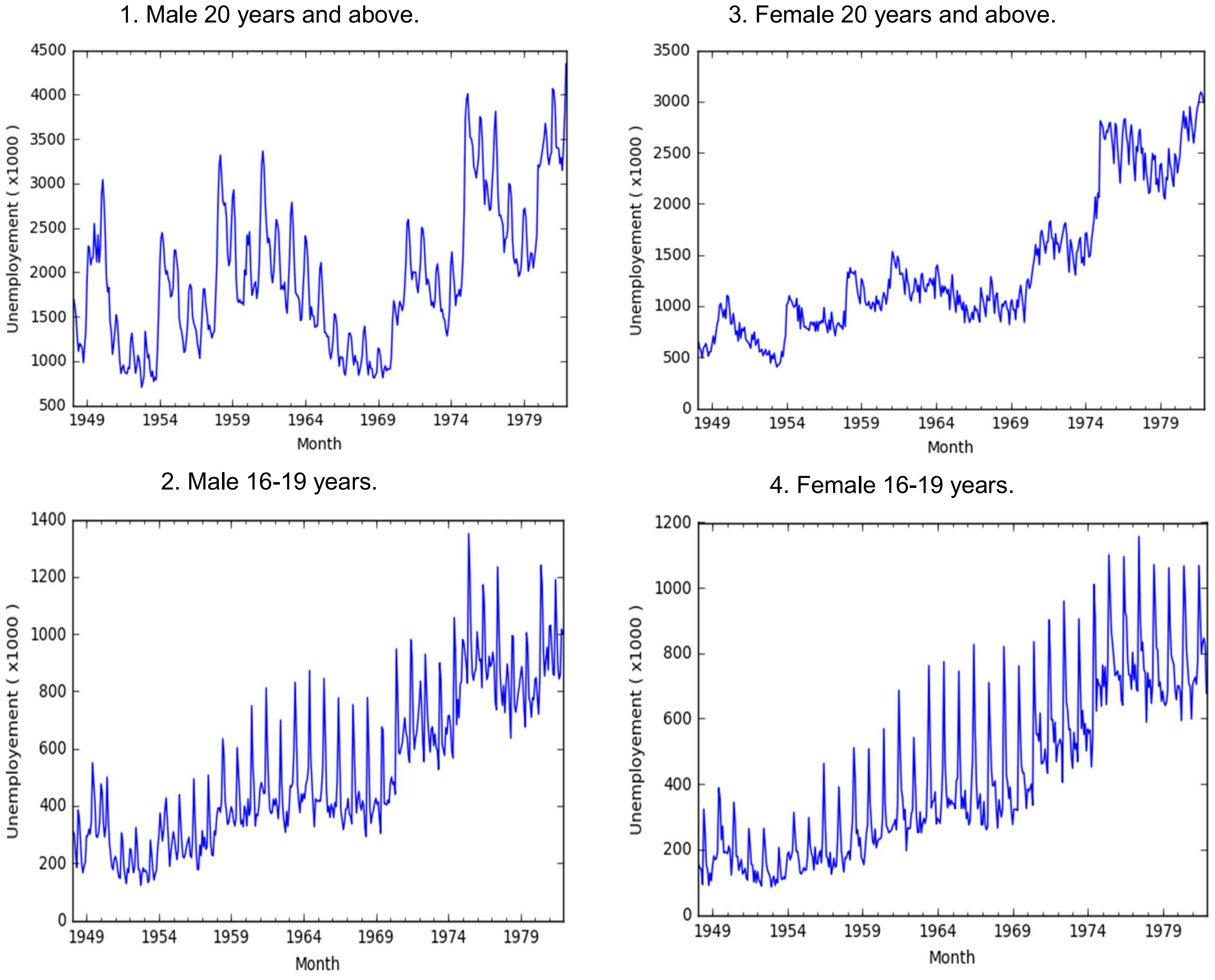

Fig. 1 Represent unemployment plots of different datasets in a fixed internal

3. MODELS REVIEW:

3.1. ARIMA

Auto Regressive Integrated Moving Average is the most used and generalized stochastic model for forecasting a time series. The model is applied linearly over the time series made stationary by differencing in conjugation with non-linear transformations [2.7]. The forecasting equation derives from the weighted sum of the lags of the dependent variable. In ARIMA, lag defines autoregressive terms (p), lag of the forecast errors define moving average terms (q) and 'd' defines the number of nonseasonal differences required to obtain stationarity. The forecasting equation can be written as:

$$\hat{y} = \mu + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

Where, y_d denotes the d^{th} difference of Y, θ moving average parameters (as introduced in Box and Jenkins). The stationarized series require p ≥ 1 and q ≥ 1 to compensate auto correlated errors and avoid random walk.

3.2. SUPPORT VECTOR REGRESSION

Support Vector Machines (SVM) are supervised learning machines which implement structural risk minimization(SRM) inductive principle to obtain good generalization on sparse patterns. SRM involves simultaneous attempt to minimize the empirical risk and the VC (Vapnik– Chervonenkis) dimension [2.4]. SVR is the most common application form of SVM. They are useful for nonlinear input-output knowledge prediction by taking in account regularization and capacity control aspects. The formulation can be done as a convex optimization problem with slack variables ξi, *ξi.

$$Minimize\ \frac{1}{2}\|w\|^2 + C\sum(\xi_i + \xi_i^*)$$

$$subject\ to \begin{cases} y_i - \langle\omega, x_i\rangle - b \le \varepsilon + \xi_i \\ \langle\omega, x_i\rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \text{--- (1)}$$

Where, ||ω|| is the weight function, x(i) inputs, y(i) corresponding outputs with ε precision and b offset. The constant C > 0 determines the trade-off between the flatness of Euclidean norm and the toleration for error greater than ε. Using Lagrangian Multipliers and formulating the dual problem, we obtain:

$$\begin{aligned} Maximize\ -\frac{1}{2}\sum(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j\rangle \\ - \varepsilon\sum(\alpha_i + \alpha_i^*) \\ + \sum y_i(\alpha_i - \alpha_i^*) \end{aligned}$$

$$subject\ to\ \sum(\alpha_i - \alpha_i^*) = 0\ and\ \alpha_i, \alpha \in [0, C] \text{ --- (2)}$$

Where, αi *αi >0 are Lagrangian Multipliers following Karush-Kuhn Tucker (KKT) condition:

$$f(x) = \sum(\alpha_i - \alpha_i^*)\langle x_i, x\rangle + b \text{ --- (3)}$$

The non-linear transformation can be made using kernel function, which maps inputs to some feature space.

$$k(x_i, x) = \emptyset(x_i)\emptyset(x) \text{ --- (4)}$$

Thus, regression estimate becomes:

$$f(x) = \sum(\alpha_i - \alpha_i^*)k\langle x_i, x\rangle + b \text{ --- (5)}$$

It is applied over various fields like disaster mapping, image segmentation, blind identifications and filter smoothing techniques.

3.3. EXTREME LEARNING MACHINES

ELM contains randomly generated hidden neurons that are application independent and unconstrained to feed forward learning and kernel learning [2.3]. Hence, are fast, easily implemented with minimum human intervention. For single layer feed-forward network, the input data is mapped to L-dimensional ELM random feature space, and the network output is given by:

$$f_l(x) = \sum\beta_i h_i(x) = h(x)\beta, \text{ --- (6)}$$

$$H\beta = T$$

Which can be formulated as where, β is the weight matrix between hidden nodes and output nodes, h(x) represents random hidden features and T target levels. Solving, we obtain:

β = H†T

where H† is the Moore-Penrose generalized inverse of matrix H. Generalization can be improved by adding regularization term. It is applied in fields where computation time needs to reduced and routing is required.

3.4. ARTIFICIAL NEURAL NETWORKS

ANN are robust, non-stochastic, non-parametric self-adaptive non-linear model that uses backpropagation to learn the relation between inputs-outputs and predict new values [2.1]. Backpropagation involves feed-forward computation, backpropagation to output layers and hidden layers and finally weight updation. The Levenberg and Marquardt algorithm (LMA) which interpolates between the Gauss-Newton algorithm (GNA) and the steepest descent method is usually employed in training. The main drawbacks of the model are computation time increment, uncertainty of overtraining and under-fitting, and lack of its evident internal working knowledge [2.2].

$$y_t = \alpha_0 + \sum\alpha_j g\left(\sum\beta_{ij}x_i + \beta_{0j}\right) + \varepsilon_t \text{ --- (7)}$$

4. EXPERIMENT SETUP

The Correlation Matrices are calculated for all four datasets between the present value and the lag values to determine how many past instances should be taken into account for making a new prediction. The magnitude closer to 1 indicates a strong correlation between values and should be taken into account while the value close to zero suggests a weak correlation, and thus, can be discarded.

.

A. Male 20 years and above.

	t-6	t-5	t-4	t-3	t-2	t-1	t
t-6	1.000000	0.957844	0.878812	0.795827	0.737062	0.703317	0.67522
t-5	0.957844	1.000000	0.958147	0.879589	0.797403	0.739236	0.70548
t-4	0.878812	0.958147	1.000000	0.958387	0.880570	0.799143	0.74121
t-3	0.795827	0.879589	0.958387	1.000000	0.958557	0.881004	0.79939
t-2	0.737062	0.797403	0.880570	0.958557	1.000000	0.958796	0.88134
t-1	0.703317	0.739236	0.799143	0.881004	0.958796	1.000000	0.95895
t	0.675223	0.705488	0.741213	0.799392	0.881341	0.958955	1.00000

C. Female 20 years and above

	t-6	t-5	t-4	t-3	t-2	t-1	t
t-6	1.000000	0.980753	0.969042	0.957302	0.955533	0.956419	0.948554
t-5	0.980753	1.000000	0.981029	0.969483	0.957918	0.956179	0.957027
t-4	0.969042	0.981029	1.000000	0.981303	0.969931	0.958542	0.956793
t-3	0.957302	0.969483	0.981303	1.000000	0.981602	0.970396	0.959116
t-2	0.955533	0.957918	0.969931	0.981602	1.000000	0.981884	0.970791
t-1	0.956419	0.956179	0.958542	0.970396	0.981884	1.000000	0.982127
t	0.948554	0.957027	0.956793	0.959116	0.970791	0.982127	1.000000

B. Male 16-19 years

	t-6	t-5	t-4	t-3	t-2	t-1	t
t-6	1.000000	0.884845	0.788617	0.773618	0.793181	0.815182	0.807121
t-5	0.884845	1.000000	0.885255	0.789366	0.774594	0.795142	0.816856
t-4	0.788617	0.885255	1.000000	0.885733	0.790318	0.775874	0.796290
t-3	0.773618	0.789366	0.885733	1.000000	0.886195	0.791151	0.776831
t-2	0.793181	0.774594	0.790318	0.886195	1.000000	0.886501	0.792112
t-1	0.815182	0.795142	0.775874	0.791151	0.886501	1.000000	0.887483
t	0.807121	0.816856	0.796290	0.776831	0.792112	0.887483	1.000000

D. Female 16-19 years

	t-6	t-5	t-4	t-3	t-2	t-1	t
t-6	1.000000	0.878231	0.794981	0.771221	0.759312	0.746859	0.712204
t-5	0.878231	1.000000	0.878849	0.796519	0.773051	0.761049	0.747111
t-4	0.794981	0.878849	1.000000	0.879578	0.797779	0.774400	0.761700
t-3	0.771221	0.796519	0.879578	1.000000	0.880400	0.799080	0.774932
t-2	0.759312	0.773051	0.797779	0.880400	1.000000	0.881182	0.799465
t-1	0.746859	0.761049	0.774400	0.799080	0.881182	1.000000	0.881282
t	0.712204	0.747111	0.761700	0.774932	0.799465	0.881282	1.000000

Fig. 2 Represent the correlation factor of a time series value with is correspondent lag value

From the above observations, we concluded to take previous 6 values to calculate the next value. We trained the data on first 75% values and assigned the last 25% (101) observations as "unseen" data for testing purpose. We followed walk-forward validation model for prediction so that we could persist the most recent observations for the next day. This means we did not make one 101 days forecast but rather one hundred and one 1 day forecasts.

We compared the values of these 4 datasets based on the models explained above. The parameters of each model are constant over datasets for comparative analysis.

5. Comparisons between Unemployment Rate Forecasts Made by Different Forecaster:

Training:

Model Used | Dataset	Male 20 years and above	Male 16-19 years	Female 20 years and above	Female 16-19 years
ARIMA (6,1,0)	-	-	-	-
Support Vector Regression	173.48159564	104.304436665	102.226009283	110.579242719
Extreme Learning Machine	270.154809697	121.052994259	149.820781571	116.46770666
Neural Network (Backpropagation)	154.700533881	91.4330108232	102.457294303	99.9797213018

Table 1. Shows the Training Root Mean Square error for different models and different datasets

Testing:

Model Used | Dataset	Male 20 years and above	Male 16-19 years	Female 20 years and above	Female 16-19 years
ARIMA (6,1,0)	227.188	117.852	158.725	122.039
Support Vector Regression	239.696483757	141.790012924	169.595491674	144.733925174
Extreme Learning Machine	554.255127922	265.198183893	699.62632301	239.932014761
Neural Network (Backpropagation)	252.655382869	118.320224331	163.449556267	125.674253055

Table 2. Shows the Testing Root Mean Square error for different models and different datasets

From Fig.1, it is visible that there is a huge amount of variations in Unemployment data of Male who are 20 years and above. It varies from being 500 (x 1000) to 4500 (x 1000), as compared to other datasets with fewer variations. This could be one of the reasons of the maximum RMSE value of the predicted data on any model which compared across datasets

5.1 ARIMA

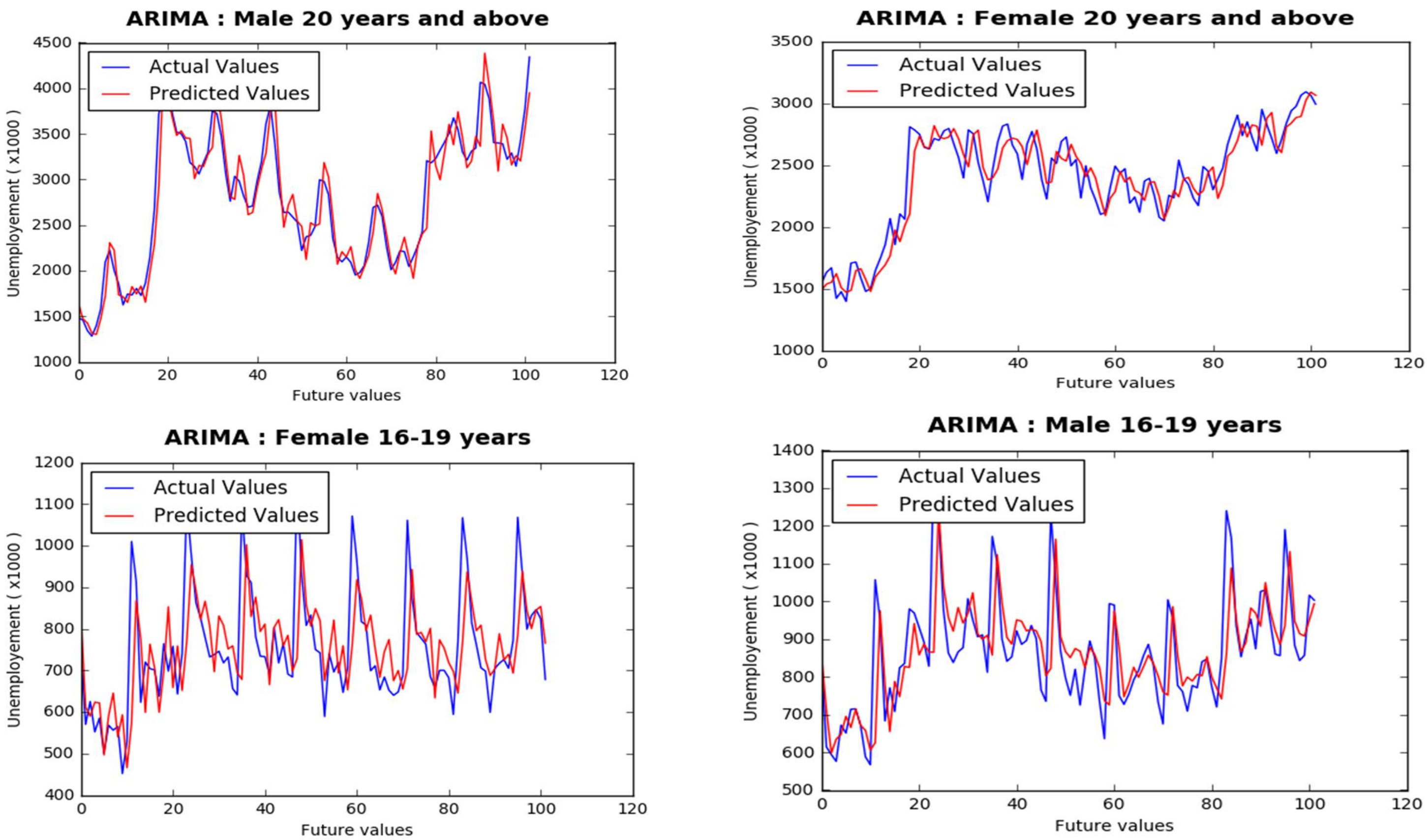

Fig. 3 shows the variation in RMSE in ARIMA for different datasets

Optimum value of number of past value taken for prediction is 6. It was observed that with increasing ‘p’ from beyond a point decrease the general trend. It is evident that among all the models, ARIMA (6,1,0) has shown the best results in all datasets

5.2 ARTIFICIAL NEURAL NETWORK

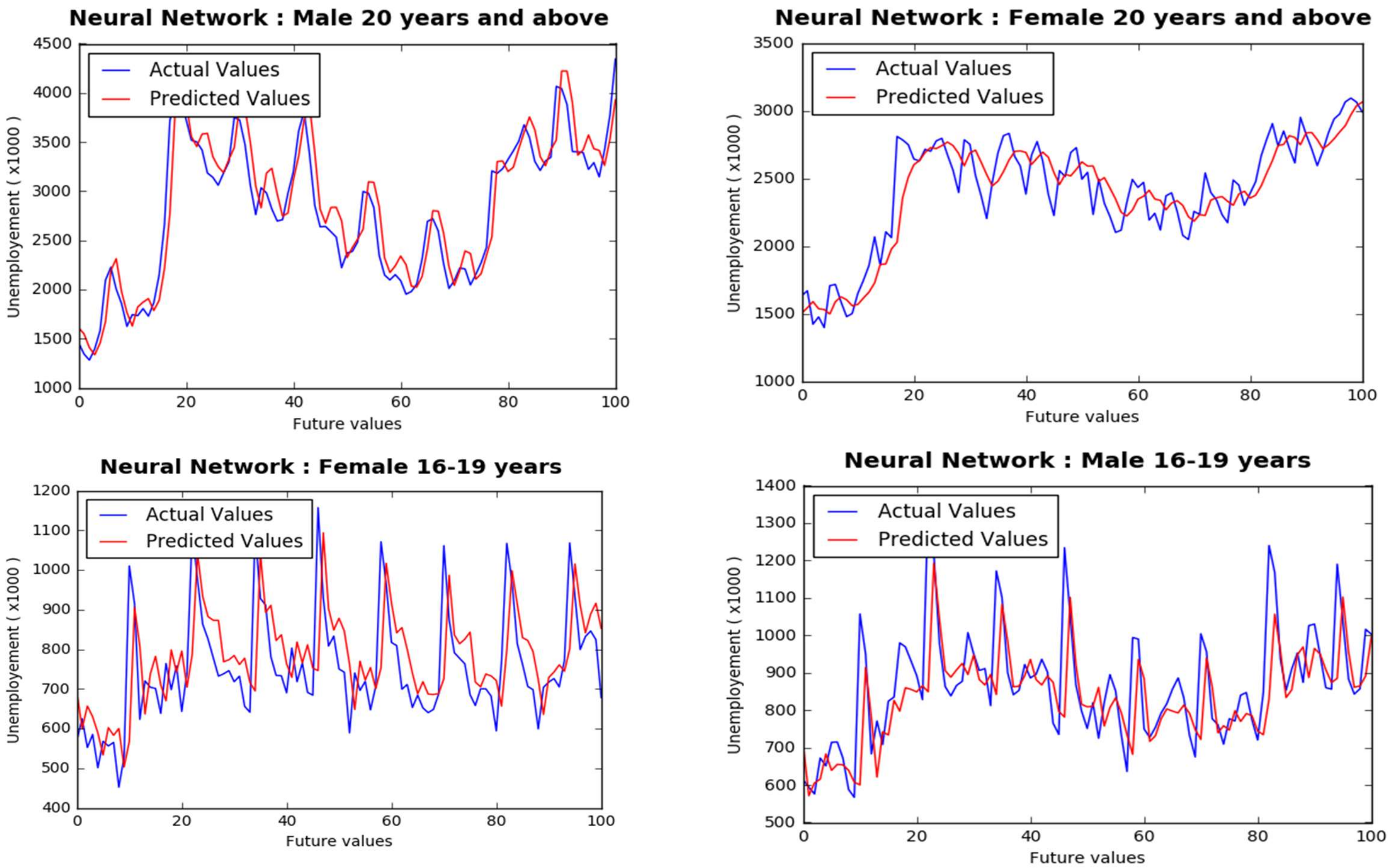

Fig. 4 shows the variation in RMSE in Neural Network for different datasets

Neural Network with backpropagation using rectified linear units (ReLU) in the two hidden layers followed by Linear activation function is also able to generalise the data because of its nonlinearity and has results comparable with ARIMA (6,1,0). Features like Dropout and batch training has also been implemented. Training of Neural Network required GPU and Cuda.

5.3 SVR

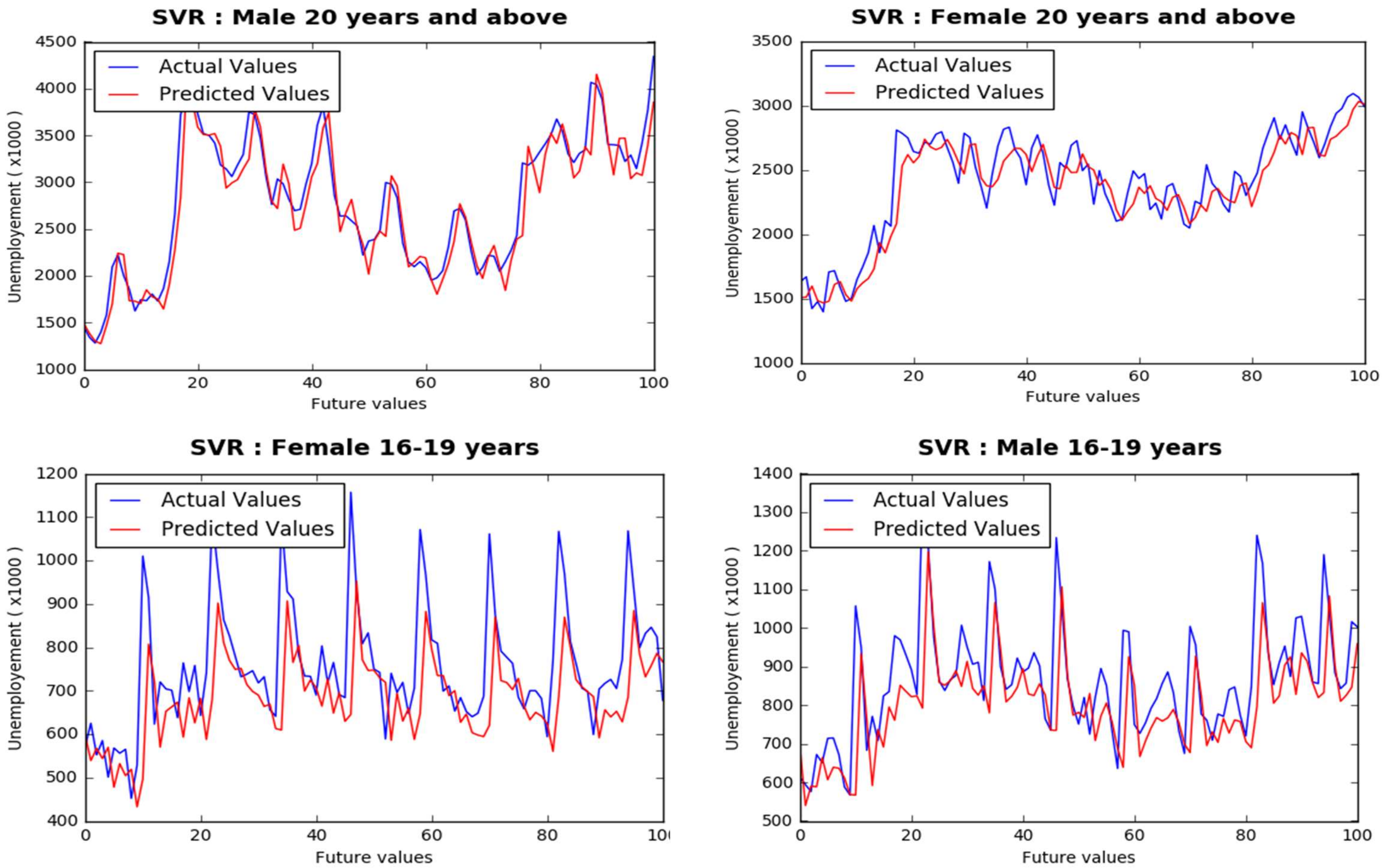

Fig. 5 shows the variation in RMSE in SVR for different datasets

The linear support vector regression with C=10.0 did not perform well as compared to the best model of neural network and ARIMA. Increasing ε leads to increase in RMSE.

5.4 ELM

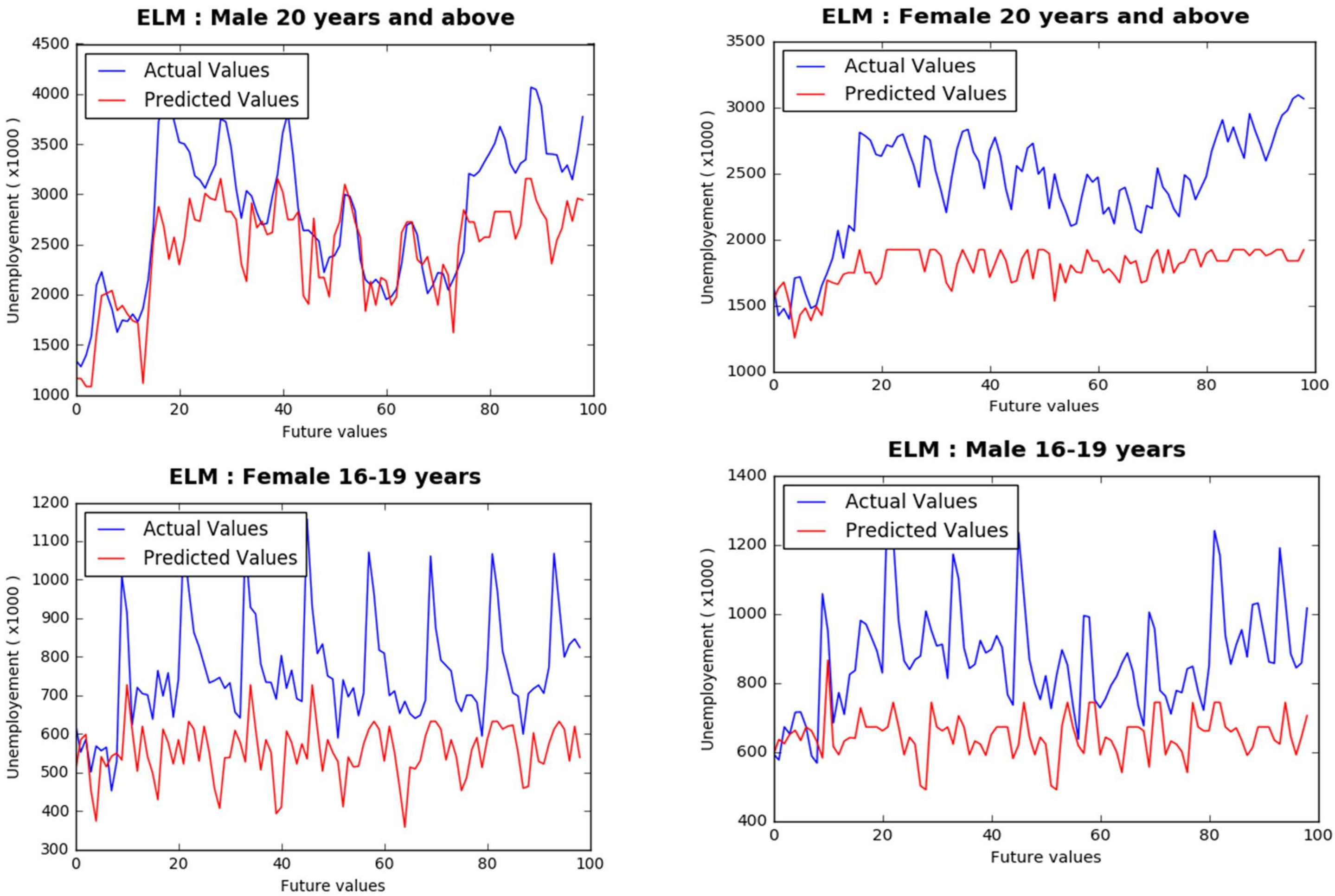

Fig. 6 shows the variation in RMSE in ELM for different datasets

Calculations in ARIMA and learning of weights in neural network takes more computation time. Whereas, in ELM matrix inverse is calculated which takes much less computation time and converges to global minima. However, in our case, even though after tuning the number of parameters (number of neurons), and having less computation time, ELM did not produce any good results

6. CONCLUSION

This paper discussed the various time-series analysis techniques, linear and non-linear, for modelling the monthly U.S. unemployment rate changes of males and females over the period of 1948-1981. Unemployment is an important factor for a country and needs to be captured before any policy making. The results of the paper showed ARIMA to be the most suitable technique to model the unemployment rate along with Neural Network. It is also believed that variations in the architecture of neural network may be able to produce better results. ELM being fast was not able to generalise properly.

7. REFERENCES

[1] Datasets:

[1.1] "*Monthly U.S. male (16-19 years) unemployment figures (thousands) 1948-1981*", https://datamarket.com/data/set/22q4/monthly-us-male-16-19-years-unemployment-figures-thousands-1948-1981#!ds=22q4&display=line

[1.2] "*Monthly U.S. male (20 years and over) unemployment figures (thousands) 1948-1981*", https://datamarket.com/data/set/22xk/monthly-us-male-20-years-and-over-unemployment-figures-103-1948-1981#!ds=22xk&display=line

[1.3] "*Monthly U.S. female (20 years and over) unemployment figures (thousands) 1948-1981*",https://datamarket.com/data/set/22lq/monthly-us-female-20-years-and-over-unemployment-figures-103-1948-1981#!ds=22lq&display=line

[1.4] "*Monthly U.S. female (16-19 years) unemployment figures (thousands) 1948-1981*", https://datamarket.com/data/set/22s8/monthly-us-female-16-19-years-unemployment-figures-thousands-1948-1981#!ds=22s8&display=line

[2] Theory:

[2.1] Box, G.E.P., Jenkins, G.M., 1976. "*Time Series Analysis: Forecasting and Control*." Holden-Day, San Francisco, CA.

[2.2] Chakraborty, K., Mehrotra, K., Mohan, C.K., Ranka, S., 1992 "*Forecasting the behavior of multivariate time series using neural networks*" Neural Networks 5, 961–970.

[2.3] G.-B. Huang et al., "*Extreme Learning Machine for Regression and Multiclass Classification*," IEEE Trans. Systems, Man, and Cybernetics, vol. 42, no. 2, 2011, pp. 513–529.

[2.4] V. Vapnik, S. Golowich and A. Smola, 1997. "*Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing*", in M. Mozer, M. Jordan, and T. Petsche (eds.), Neural Information Processing Systems, Vol. 9. MIT Press, Cambridge, MA.

[2.5]A.J. Smola, and B. Schölkopf, 1998. "*A Tutorial on Support Vector Regression*", NeuroCOLT, Technical Report NC-TR-98-030, Royal Holloway College, University of London.

[2.6] "*The female-male unemployment differential : effects of changes in industry employment*", https://stats.bls.gov/opub/mlr/1984/11/art2full.pdf

[2.7] Fischer, B. and Planas, C. 1998. "*Large Scale Fitting of ARIMA Models and Stylized Facts of Economic Time Series*" Eurostat Working Paper 9/1998/A/8. To appear in The Journal of Ocial Statistics.

[2.8] MATLAB and Signal Processing Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States. http://www.mathworks.com/