Frequency Forecasting using Time Series ARIMA model

Manish Kumar Tikariha

DGM(O) NSPCL Bhilai

Abstract

In view of stringent regulatory stance and recent tariff guidelines, Deviation Settlement mechanism (DSM) provides an opportunity to increase profitability for power generating IPPs'. DSM signifies the importance of grid frequency prediction as deviation charges of energy supplied in deviation from pre committed schedule depends on the block grid frequency. This paper attempts to forecast block ahead grid frequency based on Auto Regressive Integrated Moving Average (ARIMA) time series model. Performance of the proposed model is compared with other models such as moving average, holte winters, exponential moving average using error indices such as Mean Absolute Percentage Error(MAPE), Absolute Percentage Error(APE). Results show that the proposed model has been able to outperform in forecasting grid frequency for the sample period over other techniques.

1. Introduction

Frequency forecasting plays an important role in power system operations. Frequency forecasting becomes crucial as Electricity as a commodity cannot be stored, it has to be generated and consumed in real time. Accurate forecasting of power demand and frequency in the grid is beneficial to generators and beneficiaries for proper scheduling, trading and resource management. Under Deviation Settlement mechanism (DSM), future frequency forecasting becomes crucial to both generators and beneficiaries as it provides enormous potential of profit through under /over injection/drawl based on block frequency.

Power frequency deviation from nominal 50 Hz depends on the instantaneous imbalance between the demand and generation of active power in the grid. Severe impact may occur to power system in terms of grid failures if the grid frequency is allowed to deviate heavily from its nominal range. Thus frequency prediction also becomes crucial for power system operations from system stability point of view. Power system frequency is very stochastic in nature as it depends on various independent random variables such as power demand fluctuations, climatic conditions, change in generating capabilities, transmission capabilities and system outages etc. With higher proportion of renewable energy sources like solar and wind in the grid, variability in demand and frequency is bound to increase. Most of the previous studies and literature on the short term energy demand forecasting were mainly based on time series and Non time series techniques like machine learning techniques, Artificial neural network, fuzzy logic etc. Artificial intelligence methods like artificial neural network (ANN), fuzzy logic etc are based on the learning from experiences. These techniques are useful to model the non-linear relationship between frequency and other variables based on their historical data.

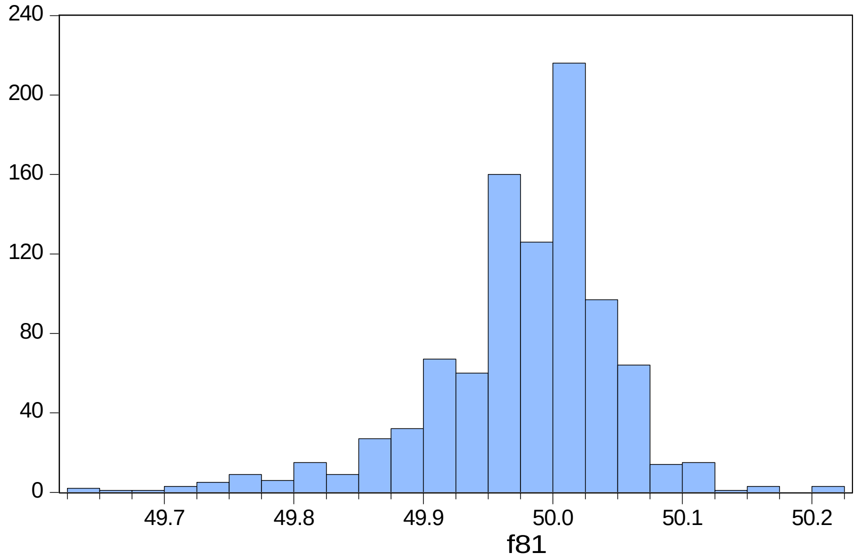
Time series forecasting methods can further be classified into Unvariate time series and multivariate time series regression based methods. Univariate time series analysis is based on the historical values of frequency to predict future frequency like Auto Regressive Integrated moving Average

(ARIMA) models, Exponential smoothing, Holte Winters etc. In multivariate analysis other independent variables like weather parameters are also taken into account.

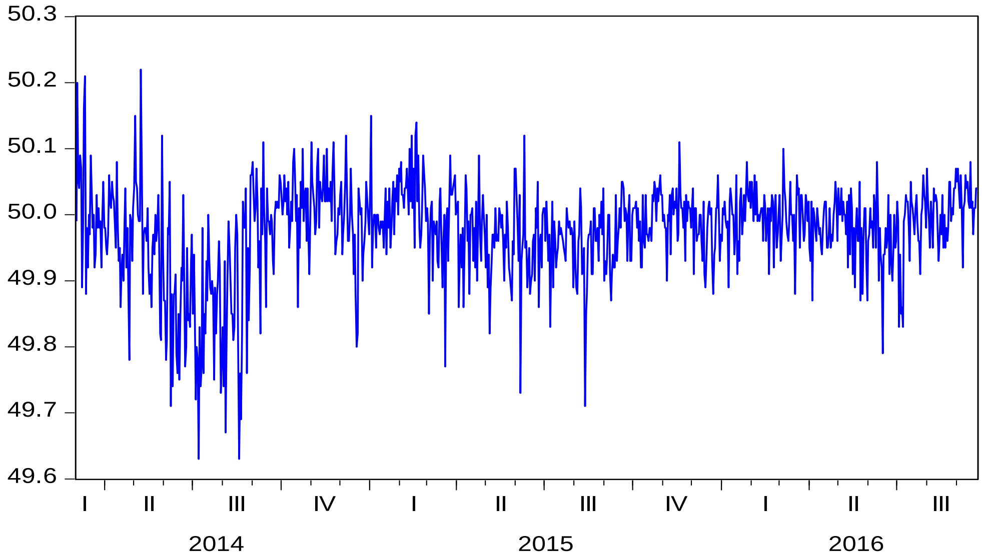
Weather is a key variable effective on the fluctuation of electricity demand and thus on frequency. However online real time weather forecast as an independent variable for multivariate modelling **is** usually considered impractical (Taylor 2003). Therefore univariate models are preferred for short lead time as weather variations in short time will be captured in load demand and frequency variations.

2. Data Description

Data for average frequency per block (15 mins) have been collected from WRLDC for the period March'14 to Aug'16. Forecasting based on ARIMA model can be done for each of the ninety six blocks of the day. In this paper data description and analysis is shown for randomly selected five blocks to save space. Out of the total 936 samples for each block frequency, first 822 samples are used for estimation of model and rest 114 samples are used as test samples to validate the model and checking purpose. Statistical software E-views is used for modelling, estimation and forecasting purpose. Summary statistics of the 81ST block frequency data is given below for illustration purpose-.



Series: F81	
Sample 3/02/2014 9/22/2016	
Observations 936	
Mean	49.97497
Median	49.99000
Maximum	50.22000
Minimum	49.63000
Std. Dev.	0.072787
Skewness	-1.055149
Kurtosis	5.667497
Jarque-Bera	451.1871
Probability	0.000000



3. Empirical Analysis

Box-Jenkins methodology has been employed to identify the most suitable ARIMA model. Box-Jenkins considers model building as iterative processes which can be divided into four stages: identification, estimation, diagnostic checking and forecasting.

The Augmented Dickey fuller test is used to test sample series stationarity . If the series is non-stationary it is first transformed into covariance stationary series and then the lag order of autoregressive and moving average part is identified. The sample Auto correlation function (ACF) and Partial autocorrelation function (PACF) have been used to identify the lag length of the ARIMA model. Now this ARIMA model can be estimated by maximum likelihood. The residuals are then inspected for any remaining autocorrelation of the residual series.

Step wise empirical analysis is presented below for 81st block frequency for illustration purpose-

3.1 Test of Stationarity

Augmentnted Dickey Fuller Test

Null Hypothesis: F81 has a unit root
Exogenous: Constant
Lag Length: 6 (Automatic - based on SIC, maxlag=20)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.536499	0.0000
Test critical values:	1% level	-3.437167	
	5% level	-2.864439	
	10% level	-2.568366	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(F81)
Method: Least Squares
Date: 01/10/17 Time: 13:09
Sample (adjusted): 3/09/2014 9/22/2016
Included observations: 929 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
F81(-1)	-0.222827	0.040247	-5.536499	0.0000
D(F81(-1))	-0.472956	0.045938	-10.29563	0.0000
D(F81(-2))	-0.327888	0.045694	-7.175677	0.0000
D(F81(-3))	-0.290749	0.043773	-6.642270	0.0000
D(F81(-4))	-0.299602	0.041531	-7.213946	0.0000
D(F81(-5))	-0.262273	0.038575	-6.798981	0.0000
D(F81(-6))	-0.146627	0.032280	-4.542318	0.0000
C	11.13566	2.011320	5.536492	0.0000

R-squared	0.345253	Mean dependent var	0.000161
Adjusted R-squared	0.340276	S.D. dependent var	0.074045
S.E. of regression	0.060142	Akaike info criterion	-2.775654
Sum squared resid	3.331273	Schwarz criterion	-2.734026
Log likelihood	1297.291	Hannan-Quinn criter.	-2.759776
F-statistic	69.37849	Durbin-Watson stat	2.025145
Prob(F-statistic)	0.000000		

* ADF test statistic p- value is 0.00, so the Null hypothesis can be rejected.

3.2 ACF and PACF

Date: 01/10/17 Time: 12:49
Sample: 3/02/2014 9/22/2016
Included observations: 936

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
|*** |	|*** |	1	0.476	0.476	212.49	0.000
|*** |	|* |	2	0.385	0.206	352.12	0.000
|** |	|* |	3	0.310	0.090	442.73	0.000
|** |	| |	4	0.267	0.063	509.70	0.000
|** |	|* |	5	0.278	0.107	582.74	0.000
|** |	|* |	6	0.333	0.165	687.60	0.000
|*** |	|* |	7	0.364	0.147	812.80	0.000
|** |	|* |	8	0.348	0.081	927.39	0.000
|** |	| |	9	0.305	0.026	1015.3	0.000
|** |	| |	10	0.276	0.028	1087.7	0.000
|** |	| |	11	0.262	0.036	1153.1	0.000
|** |	| |	12	0.239	0.004	1207.6	0.000
|** |	| |	13	0.241	0.010	1262.8	0.000
|** |	| |	14	0.272	0.055	1333.4	0.000
|** |	| |	15	0.235	-0.019	1386.0	0.000
|** |	| |	16	0.259	0.049	1450.2	0.000
|** |	|* |	17	0.282	0.075	1526.0	0.000
|** |	| |	18	0.239	-0.006	1580.8	0.000
|** |	|* |	19	0.299	0.111	1666.4	0.000
|** |	| |	20	0.279	0.039	1741.2	0.000
|** |	| |	21	0.266	0.023	1809.3	0.000
|** |	| |	22	0.288	0.067	1889.0	0.000
|** |	| |	23	0.245	-0.016	1946.6	0.000
|** |	| |	24	0.226	-0.018	1995.9	0.000
|** |	| |	25	0.238	0.017	2050.5	0.000
|** |	| |	26	0.259	0.040	2115.3	0.000
|** |	| |	27	0.256	0.009	2178.4	0.000
|** |	| |	28	0.261	0.015	2244.2	0.000
|** |	| |	29	0.263	0.033	2311.4	0.000
|* |	| |	30	0.201	-0.061	2350.5	0.000
|* |	| |	31	0.169	-0.048	2378.3	0.000
|* |	| |	32	0.164	-0.013	2404.3	0.000
|* |	| |	33	0.176	-0.015	2434.4	0.000
|* |	| |	34	0.189	0.008	2469.3	0.000
|* |	| |	35	0.179	-0.026	2500.5	0.000
|** |	| |	36	0.227	0.046	2550.6	0.000

Partial Auto correlation factors indicate inclusion of 1st, 7th 8th and 7th auto regressive terms in estimation. Seasonal Auto regressive term with 7th lag order is included for weekly seasonal effects in ACF and PACF.

3.3 **Estimation**

Four dummy variables q1, q2, q3 and q4 are included in the specification as there is quarterly effect on the observed data. Previous block frequencies of 73rd, 76th, 79th and 80th block have predominant effect on 81st block frequency as their t-stat and associated p values are significant.

Estimation Equation:
=========================
F81 = C(1)*F80 + C(2)*F79 + C(3)*F76 + C(4)*F73 + C(5)*Q1 + C(6)*Q2 + C(7)*Q3 + C(8)*Q4 + [AR(1)=C(9),AR(7)=C(10),AR(8)=C(11),AR(17)=C(12),SAR(7)=C(13),UNCOND]

Substituted Coefficients:
=========================
F81 = 0.486727209225*F80 + 0.090888294887*F79 + 0.101233777497*F76 + 0.0540560998447*F73 + 13.3644089579*Q1 + 13.3493075075*Q2 + 13.3515565544*Q3 + 13.3489901214*Q4 + [AR(1)=0.0768156756932,AR(7)=-0.184386923437,AR(8)=0.106984643857,AR(17)=0.0876064054728,SAR(7)=0.311774486381,UNCOND]

Dependent Variable: F81
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 01/10/17 Time: 13:07
Sample: 3/02/2014 5/31/2016
Included observations: 822
Convergence achieved after 20 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
F80	0.486727	0.030332	16.04652	0.0000
F79	0.090888	0.029111	3.122092	0.0019
F76	0.101234	0.025495	3.970683	0.0001
F73	0.054056	0.022681	2.383308	0.0174
Q1	13.36441	1.418502	9.421494	0.0000
Q2	13.34931	1.417883	9.414954	0.0000
Q3	13.35156	1.416065	9.428634	0.0000
Q4	13.34899	1.418322	9.411820	0.0000
AR(1)	0.076816	0.030290	2.535998	0.0114
AR(7)	-0.184387	0.066859	-2.757859	0.0059
AR(8)	0.106985	0.032023	3.340890	0.0009
AR(17)	0.087606	0.034315	2.553014	0.0109
SAR(7)	0.311774	0.066484	4.689492	0.0000
SIGMASQ	0.002296	9.34E-05	24.58776	0.0000

R-squared	0.590045	Mean dependent var	49.97300
Adjusted R-squared	0.583449	S.D. dependent var	0.074891
S.E. of regression	0.048335	Akaike info criterion	-3.203951
Sum squared resid	1.887712	Schwarz criterion	-3.123703
Log likelihood	1330.824	Hannan-Quinn criter.	-3.173163
Durbin-Watson stat	1.993868		

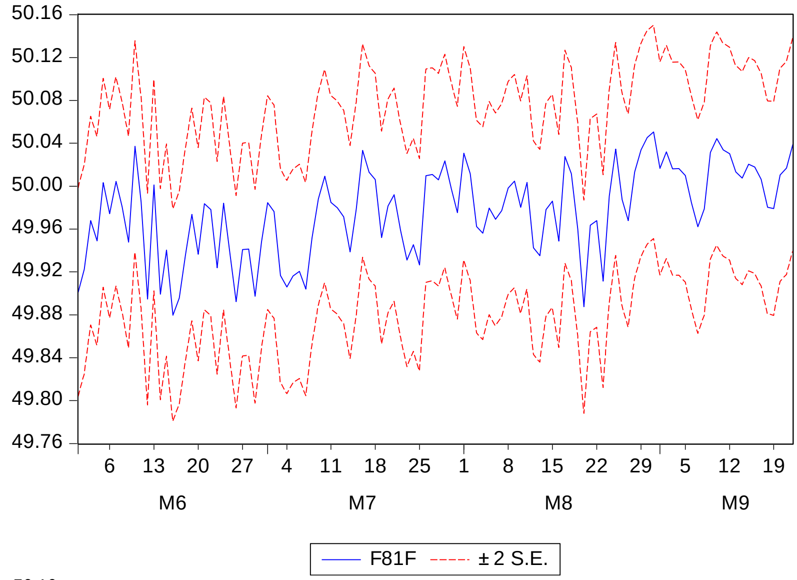
Inverted AR Roots	.86	.85	.81-.33i	.81+.33i
	.66-.57i	.66+.57i	.53+.66i	.53-.66i
	.37+.77i	.37-.77i	.10+.89i	.10-.89i
	-.19+.83i	-.19-.83i	-.22-.80i	-.22+.80i
	-.55+.71i	-.55-.71i	-.70-.47i	-.70+.47i
	-.76-.37i	-.76+.37i	-.86+.12i	-.86-.12i

3.4 **Residual Correlation test**

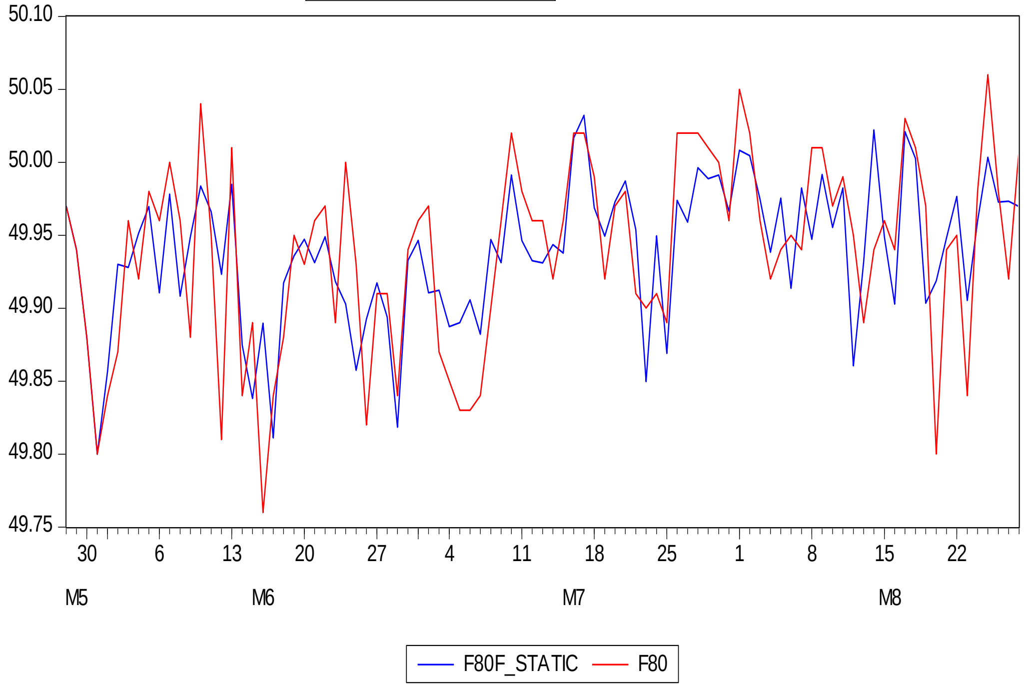
Residuals of the estimated equation are tested for correlation in terms of ACF and PACF. Based on Q-stat and associated p-values in the correlogram, auto correlation among residuals can be checked

3.5 **Forecasting**

Based on the estimated model, forecasting is done for the rest of the 114 observations from 1st June'16 to 22nd Sept'16. Results of the forecasting is presented below-



Forecast: F81F	
Actual: F81	
Forecast sample: 6/01/2016 9/22/2016	
Included observations: 114	
Root Mean Squared Error	0.033248
Mean Absolute Error	0.026260
Mean Abs. Percent Error	0.052540
Theil Inequality Coefficient	0.000333
Bias Proportion	0.145203
Variance Proportion	0.128217
Covariance Proportion	0.726580



Performance of forecasted frequency F80f_STATIC against actual frequency F80

4. Forecasting performance

Estimated ARIMA model is used to forecast block ahead grid frequency and forecasts are then evaluated using standard performance criterion such as root mean square error(RMSE), mean absolute error(MAE) and mean absolute percentage error(MAPE). The smaller the error, the better

is the forecasting performance for the series .Forecasted series performance is also compared against traditional forecasting model erstwhile used such moving average, weighted moving average and exponential moving average.
Forecasting performance is evaluated for block frequencies from 73rd to 88th and it can extended for all the 96 block frequencies.

	MAE				
	ARIMA	3MA	5MA	WMA	EMA
F73	***0.030855***	0.039825	0.037912	0.038912	0.036604
F74	***0.025926***	0.03462	0.031333	0.033763	0.031502
F75	***0.027001***	0.036901	0.035193	0.038202	0.035235
F76	***0.028035***	0.041579	0.039579	0.043018	0.039027
F77	***0.027534***	0.038567	0.037123	0.039991	0.036614
F78	***0.043775***	0.055234	0.055667	0.054026	0.052888
F79	***0.037977***	0.05424	0.054614	0.052272	0.051022
F80	***0.033008***	0.046813	0.050982	0.045316	0.045495
F81	***0.02626***	0.034351	0.035066	0.033829	0.030774
F82	***0.029401***	0.042661	0.041877	0.042175	0.039841
F83	***0.029505***	0.04117	0.037982	0.040439	0.037372
F84	***0.025882***	0.037865	0.039	0.036737	0.036244
F85	***0.03074***	0.048509	0.04993	0.04643	0.045203
F86	***0.024301***	0.04614	0.04393	0.04514	0.041447
F87	***0.028196***	0.039971	0.040228	0.039886	0.037419
F88	***0.023318***	0.035175	0.033035	0.034605	0.031734

	MAPE				
	ARIMA	3MA	5MA	WMA	EMA
F73	***0.06168***	0.079605	0.075782	0.077783	0.073167
F74	***0.051862***	0.069258	0.062686	0.067544	0.063022
F75	***0.054011***	0.073829	0.070413	0.076432	0.070496
F76	**0.056117**	0.083231	0.079235	0.086112	0.07813
F77	***0.05512***	0.077212	0.074322	0.080061	0.073302
F78	***0.087681***	0.11063	0.111505	0.108215	0.105937
F79	***0.076068***	0.108633	0.109383	0.104689	0.102188
F80	***0.066101***	0.093728	0.102074	0.090733	0.09109
F81	***0.052557***	0.068749	0.070181	0.067704	0.061592
F82	***0.058856***	0.085397	0.083831	0.084426	0.079755
F83	***0.059054***	0.082389	0.076013	0.080928	0.074793
F84	***0.051802***	0.075789	0.078059	0.07353	0.072544
F85	***0.061563***	0.097163	0.100004	0.092999	0.090541
F86	***0.048658***	0.092385	0.087959	0.090384	0.082988
F87	***0.056414***	0.079972	0.080486	0.079801	0.074867
F88	***0.046623***	0.070334	0.066056	0.069194	0.063454

	RSME				
	ARIMA	3MA	5MA	WMA	EMA
F73	***0.037851***	0.049143	0.048489	0.048219	0.04588
F74	***0.033338***	0.042665	0.03914	0.042121	0.038974
F75	***0.03556***	0.048062	0.046178	0.049531	0.046161
F76	***0.036013***	0.055646	0.053159	0.05641	0.052373
F77	***0.037128***	0.050052	0.04862	0.05026	0.047542
F78	***0.052416***	0.068321	0.069266	0.068595	0.065797
F79	***0.049654***	0.068008	0.068053	0.066523	0.063735
F80	***0.041737***	0.060515	0.063352	0.058974	0.057966
F81	***0.033248***	0.042744	0.04302	0.041626	0.0385
F82	***0.039845***	0.055496	0.054562	0.05538	0.052384
F83	***0.037612***	0.050829	0.04818	0.05041	0.047114
F84	***0.032649***	0.047915	0.048616	0.047191	0.045578
F85	***0.038094***	0.061079	0.06257	0.058714	0.057134
F86	***0.034488***	0.060356	0.059232	0.059149	0.055506
F87	***0.034866***	0.053028	0.052404	0.053326	0.049719
F88	***0.030794***	0.049513	0.046515	0.048436	0.044639

5. Conclusion

Forecasting power system frequency in real time is an arduous task as system frequency is random and stochastic in nature. It depends on various independent variables like generating capability, transmission capability, load demand, weather conditions, renewables source injection etc. Apart from above mentioned exogenous variables, it also depends upon its own lagged values, previous blocks frequencies and has quarterly and weekly periodic effects as shown in this paper.
Forecasting performance evaluation clearly indicates that the estimated ARIMA model outperforms all the other forecasting techniques in terms of all the evaluating criteria for all the block frequencies. The technique presented in the paper can be implemented for online block ahead real time frequency.
Under present regulatory scenario when other marginal contributions have been squeezed out, net gain maximisation under DSM stands out as the way to pursue. There is further scope of study in this area to include Artificial intelligence based techniques like artificial neural network based model and artificial intelligence based hybrid time series ARIMA models to further refine the forecasting results.

6.0 References

Deshmukh Surekha, Nerkar Y., 2008.Optimal Generation Scheduling under ABT using Forecasted load and Frequency. IEEE xplore 2008 conference paper.

Ghosh S., Kanjlal K.,2014.Modelling and Forecasting day ahead electricity price in indian Energy Exchange: Evidence from MSARIMA-EGARCH model. International journal of indian Culture and Business Management.

Kaur S., Agrawal S., Verma Y.P., 2013. Power Grid frequency Prediction using ANN considering the Stochasticity of Wind Power. International conference on Computational intelligence and communication Networks.

Taylor J.W., 2003, Short Term Electricity Demand forecasting using double seasonal exponential Smoothing, Journal of Operational research Society.