CHAPTER - VI

COMPARISON OF THE FORECASTING PERFORMANCE AND FINAL MODEL SELECTION

Contents **Page No.**

CHAPTER -VI

COMPARISON OF THE FORECASTING PERFORMANCES AND FINAL MODEL SELECTION

The main purpose of this study is to select the best predictive model out of selected three types of models, so that the energy consumption in Bangladesh can be predicted efficiently. In this chapter, analysis toward the comparison of predictive performance of selected three types of models is performed. Thus, section 6.1 focuses review of some works on comparative aspect of predictive performance, the criteria used for comparison of predictive performance in this study both pair wise and individual are discussed in section 6.2, discussion of the results obtained for comparison for all types of data are given in section 6.3 and finally, a summary of the findings of this chapter is provided in section 6.4. The results are reported in tables in the appendix.

6.1. INTRODUCTION

Some empirical studies on the comparative aspect of predictive performance of econometric models have been performed by some researchers.

For example, Ferrer *et. al.*(1997) investigated the forecasting ability of unobserved components models compared to the univariate ARIMA approach for five types of monthly automobile sales in Spain. Accuracy of forecasting ability was assessed by comparing some measures of forecasting performance based on out-of-sample predictions. They showed that the performance of ARIMA approach was fair by most of the performance criteria. Jamal and Abdullah (1996) developed an econometric model of sinusoidal type to forecast the electricity consumption and to study the impact of ambient temperature on consumption in the Eastern Province of Saudi Arabia. They compared the results of this model with that of simple linear regression and quadratic models. They showed that simple regression model gave better results followed by sinusoidal and the quadratic model. Kohzadi *et. al.* (1996) compared the predictive performance of ANN technique with ARIMA type of model using data of live cattle and wheat price. They used AME, MAPE and MSE criteria for comparison of prediction accuracy. They also performed turning point test suggested by Cumby and Modest (1981) and showed that ANN model is the better predictive model. Nizami and Garni (1995) performed a comparative study on the predictive performance of ANN technique with that of regression technique. They used chi-sqaure (goodness of fit) criterion for comparison of prediction performance and showed that ANN outperformed regression technique. Chiang *et. al.* (1996) defined and developed a back-propagation neural network and compare its predictive performance with that of linear and nonlinear regression techniques. They used net asset value of mutual fund data for their study. For comparison they used MAPE Criterion only. They showed that the neural network significantly out performed regression models in the situation with limited data availability. Leshno and Spector (1996) verified the predictive performance of ANN technique in comparison with that of classical multivariate discriminant analysis (MDA) using data of bank bankruptcy in New York. For comparison purposes they used type I, type II and total error of the selected models and showed that performance of ANN model is better. Chakrabarty *et. al.*(1992) compared the performance of multivariate neural network technique with that of ARMA model using the data of monthly flour prices in the Buffalo, Minneapolis and Kansas cities. They used MSE and coefficient of variation criteria for comparison. They showed that performance of MNN model superseded the ARMA model by most of the criteria. Tang *et. al.* (1991) compared neural networks and Box-Jenkins models, using international airline passenger traffic, domestic car sales and foreign car sales in the USA. They concluded that Box-Jenkins models outperformed the neural net models in short-term forecasting. On the other hand, neural net models outperformed the Box-Jenkins models in the long run.

In this study, an attempt is made to compare the predictive performance of three types of models, *viz.* growth types of models, ARIMA types of models and the ANN techniques to the two types of energy consumption data. The best models of each of the three types are analytically selected in last three chapters. Predictive performances of these three types of models to the log-transformed data are also compared. A brief description of the criteria used for comparison of predictive performance are given in next section.

6.2. CRITERIA USED FOR COMPARISON OF PREDICTIVE PERFORMANCES

In order to select the model of best fit for efficient prediction of energy consumption figures, the predictive performances of the selected models are compared using the following criteria:

6.2.1. *Pair wise accuracy comparison*

At first predictive performance of the three competing models considering pair wise performance is undertaken. The following methodology is employed to total series for this purpose.

Diebold and Mariano (1995 p.254) proposed and evaluated explicit tests of a null hypothesis of no difference in the accuracy of two competing forecasts for finite samples which is given below. Allowing the time-t loss associated with a forecast to be an arbitrary function of the realization and prediction, say, $g(c_t, c_{it})$, where c_{it} is the forecasted value of the ith model. They consider the loss function as a direct function of forecast error i,e, $g(c_t, c_{it}) = g(e_{it})$. Then the null hypothesis to be tested is the hypothesis of equal forecast accuracy for two forecasts is $E[g(e_{it})] = E[g(e_{jt})]$, where i and j stand for the ith model and the jth model respectively, which can be written as $E(d_t) = 0$, where $d_t \equiv [g(e_{it})\text{-}g(e_{jt})]$, the linear loss differential. Thus, tests of equal accuracy null hypothesis is equivalent to the tests of null hypothesis that the population mean or median of the loss differential series is zero.

The proposed two powerful tests, the Sign Test based on observed loss differential and the Wilcoxon's Signed rank test based on the ranks of loss differentials are discussed below.

i) ***Sign Test***

In this case the null hypothesis is considered to be a zero median loss differential Med.[$g(e_{it})$ - $g(e_{jt})$] = 0. If the loss differential is symmetrically distributed, then the $g(e_{it})$ and $g(e_{jt})$ is the same. This symmetry of loss differential may be obtained if the distribution of $g(e_{it})$ and $g(e_{jt})$ is the same up to location shifts.

Assuming the loss differential series is identically and independently distributed (iid), and the number of positive loss differentials in a sample of size N has the Binomial distribution with parameter N and 1/2 under the null hypothesis. Thus, the test statistic is

$$S = \text{Sum of } I_{+}(d_t)$$

where, $I_{+}(d_t) = 1$ if $d_t > 0$

$= 0$, otherwise.

In a large sample, the standardized version of the sign-test statistic is standard normal,

$$S_a = \frac{S - \frac{1}{2}N}{\sqrt{\frac{N}{4}}} \overset{a}{\sim} N(0,1)$$

given by

ii) ***Wilcoxon's Signed Rank Test***

A relatively powerful test than the sign test is Wilcoxon's signed-rank test. In this case too, it is assumed that the loss differential series is iid. Thus, the test statistic is defined as

$$\text{WSR} = \text{Sum of } [I_+(d_t) \text{ x Rank of} \mid d_t \mid]$$

i,e, WSR is equal to the sum of the ranks of absolute values of positive observations. In this case the standardised version is also standard normal, given by

$$\text{WSR}_a = \frac{\text{WSR} - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{2N}}} \overset{a}{-} N(0,1)$$

In the study assuming sample size is large, the asymptotic versions of tests are used. Lehman (1975) also showed that usual Sign test and Wilcoxon's signed rank tests were standard. In all of the above tests a squared error loss differential or absolute loss differential may be used. However, in this study linear loss differential is considered.

iii) *Morgan-Granger-Newbold (MGN) Test*

Granger and Newbold (1986) introduced an orthogonal transformation of forecasting error components due to Morgan to the error in forecasted values which enabled relaxation of contemporaneously non-correlation assumption of forecast errors. This test introduced by Granger and Newbold termed as MGN, due to Morgan, Granger and Newbold, and advocated by Diebold and Mariano (1995, p.256) is as follows:

Let $x_t = (e\hat{}_{it} + e\hat{}_{jt})$ and $z_t = (e\hat{}_{it} - e\hat{}_{jt})$, then testing the null hypothesis of equal forecast accuracy is equivalent to zero correlation between x and z , thus the test statistic for testing the null hypothesis of equal accuracy given by

$$MGN = \frac{\hat{\rho}_{xz}}{\sqrt{\dfrac{1 - \hat{\rho}^2_{xz}}{N - 1}}}$$

is distributed as student's t with N-1 degrees of freedom.

iv) *Test for encompassing-in-forecasts*

The test for encompassing-in-forecasts introduced by Chong and Henry (1986) was applied by Hallman and Kamstra (1989) and recently by Donaldson and Kamstra (1996). In each study they used the test for encompassing-in-combined-forecasts to compare the combined forecasts' accuracy. However, in this study this test is performed following Donaldson and Kamstra (1996) for testing the null hypothesis that the forecasts accuracy of pair wise model is equal.

The test formalises the intuition that model i should be preferred to model j if model i can explain what model j cannot, while model j is preferred to model i if model j being able to explain what model i cannot explain. As such, the test provides a useful method for ranking out-of-sample forecasts.

Let e_{it} be the model i's forecasts' error in time t as defined earlier and f_{jt} be model j's forecasts, thus, the test for encompassing involves testing for significance of the β_1 parameter in the regression $e_{it} = \beta_0 + \beta_1 f_{jt} + \epsilon_t$.

Given forecasts for two models i and j, the null hypothesis considered is that neither model encompasses the other. Then it is first required to regress the forecast error from model i on the forecast for model j, as in above equation, call this regression 'ij' and the resulting estimates of β_1 coefficient β_{ij}. Then regress the forecast error from model j on forecasts from model i, call this regression 'ji', and the resulting estimates of β_1 coefficient β_{ji}. If β_{ij} is not significant at some predetermined level, but β_{ji} is significant, then reject the null hypothesis that neither model encompasses other in favor of the alternative hypothesis that model i encompasses model j. Conversely, if β_{ji} is not significant at some predetermined level, but β_{ij} is significant, then it is said that model j encompasses model i. However, if both ß's are significant or if both are insignificant, then one fails to reject the hypothesis that neither model encompasses the other. Multicollinearity can lead to both estimated coefficients being insignificant, while sufficiently non-overlapping information sets can lead to both estimated coefficients being significant.

It is clear from the above discussion that the null hypothesis to be considered in case of two tailed tests i), ii) and iii) is

H_0: Forecasts' accuracy of considered two types of models are same;

against the alternative

H_A: They are different.

Again the null hypothesis to be considered in case of test iv) is

H_0: Neither of the considered two models (i and j) encompasses the other.

against two alternatives

H_A: Model i encompasses model j

H_A: Model j encompasses model i.

6.2.2 *Individual predictive performance comparison*

Apart from testing the pair wise equal forecast accuracy of the selected models, the predictive performances of the models are also compared against the values of following criteria for all of the three periods, *viz.* estimation period, validation period and total period.

i) R^2, *The coefficient of determination*: The value of R^2 is considered here instead of $\overline{R}^2$, the adjusted value of R^2, because, $\overline{R}^2$ does not exist for prediction period of electricity consumption. The reason is that in case of growth and ARIMA models, k is greater than N for prediction period.

ii) **AME,** the mean absolute error, iii) **RMSE,** the root mean square error and iv) **MAPE,** the mean absolute percent error as defined in section 3.3 of chapter III.

Miller (1978; p.580) stated that there was a drawback of using this RMSE as a measure of forecast accuracy. For a model which fails test for stability over time, the value of RMSE Statistic will be sensitive to the period over which it is computed. Even if the models are stable over time, the use of RMSE still does not permit a valid comparison of some models. However, this criterion is used by almost all researchers who deal with the problem of predictive performance comparison and in addition to the RMSE, values of other criteria are also compared.

v) SAPE, *the 'smoothed' absolute percent error* : The smoothed absolute percent error is defined as

$$SAPE = \sum_{t=1}^{N} \frac{|c_t(pred) - c_t(obs)|}{[c_t(pred) + c_t(obs)]/2} \times 100$$

The advantage of this measure of forecasts accuracy over other measures is that it not only views accuracy as a relative measure, widely used and easily understood (Armstrong and Collopy 1992), but it is also less susceptible to small values in the denominator (Makridakis 1993).

This type of the measure of forecasts accuracy is used by O.Connor *et. al.* (1997) for comparing the forecasts accuracy of different types of time series data such as- downward series, flat series and upward series.

vi) PC, *the prediction criterion*: Amamiya (1972; 1976, p.7) suggested the use of a criterion named, prediction criterion, defiend as

$$PC = \sigma^{\wedge 2}(1 + \frac{k}{N})$$

Although he suggested this criterion for the selection of subset regressors of regression model, here this criterion is used for comparing the predictive performance.

vii) *Chi-square criterion*: A chi-sqaure statistic is also used as a criterion to measure the goodness of fit , defined as

$$\chi^2 = \sum_{t=1}^{N} \frac{[c_t(\text{obs}) - c_t(\text{pred})]^2}{c_t(\text{pred})}$$

Minimum values of the above-mentioned criteria are desirable for a comparatively adequately good model.

viii) *Theil's U criterion*

The familiar AME and RMSE criteria can be misleading in certain cases. They are not comparable, especially, when the units of measurements of two or more variables to be forecasted are different. Theil (1961, p.32) introduced a measure of coefficient of inequality U which was free from this problem and used as testing criterion for forecasting performance, defined as

$$U = \frac{\sqrt{\frac{1}{N}\sum_{j=1}^{k}[p_{t+j} - c_{t+j}]^2}}{\sqrt{\frac{1}{N}\sum_{j=1}^{k}p_{t+j}^2} + \sqrt{\frac{1}{N}\sum_{j=1}^{k}c_{t+j}^2}}$$

Leitch and Tanner (1991, p.581) viewed this U as RMSE of a forecast divided by the RMSE

of naïve forecasts of no change and defined as

$$U = \sqrt{\frac{\sum_{j=1}^{k}(\Delta p_{t+j} - \Delta c_{t+j})^2}{\sum_{j=1}^{k}\Delta c_{t+j}^2}}$$

where, p_{t+j} is the forecast value j period ahead, c_{t+j} is the actual value of the variable.

U = 0, when the prediction is perfect

= 1, when RMSE of the prediction change equals the accuracy of the baseline forecast of no change and

> 1, indicates that the forecasts have higher RMSE than the no change forecast.

This U statistic is one of the most popular criteria in the literature of forecasting. In this study the formula given by Theil is used. Intrilligator (1978, p.524) called this statistic as the inequality coefficient.

In all of the above formula, the notations possess usual meanings. It is to be noted here that in all cases of computations of the values of criteria, the degree of freedom for ANN model is considered to be one following Nizami and Garni (1995;p.1102) and for growth and ARIMA models it is usual.

The result obtained for the values of the criteria for different competing models are discussed below.

6.3. RESULT AND DISCUSSION

For the purpose of comparison, the predicted values obtained by three competing models of the types growth, ARIMA and ANN are arranged in tables (each for same type of data). The values of the criteria of both types (for individual and pair wise comparison) are computed and reported in tables. The results obtained for predictive performance comparison for each type of data are discussed below.

6.3.1. *For untransformed gas consumption*

The values of the criteria discussed in section 6.2 are computed and reported in different tables. In case of pair wise comparison all the test statistics discussed in section 6.2.1. are computed for total period. The sample size is assumed here large and as a result, the distributions of the statistics become normal. The values of the criteria selected for individual performance comparison are computed separately for three periods *viz*. estimation period, prediction period and total period. As mentioned earlier, the d.f. for ANN model is considered as one.

In case of pair wise equal forecasts' accuracy comparison, the values of the criteria of non-parametric tests for untransformed gas consumption are reported in table 6.1.1(a) and those of the encompassing-in-forecasts are reported in table 6.1.1(b). From the values of statistics in table 6.1.1(a), it is evident that there is no significant difference between the forecasts' accuracy of growth model and ARIMA Model, and between that of growth and ANN models. Only WSR and MGN tests show a significant difference between the forecasts accuracy of ARIMA and ANN models. Again, the regression coefficients in table 6.1.1(b) show that neither of the three competing models encompasses either model. This may happen due to the fact that as far as the criteria are concerned (in table 6.1.2), performance of ANN and growth models are almost same and better. On the other hand, performance of growth and ARIMA models are almost same and a bit worse than that of the first pair. So, for selecting the most adequate predictive model, it is necessary to look at table 6.1.2. for individual predictive performance comparison.

The values of the criteria for the selected models for individual predictive performance comparison for untransformed gas consumption are reported in table 6.1.2. It is clear from the table that values of R^2 for estimation and prediction periods are maximum for selected ANN model. The values of all other criteria like AME, RMSE, MAPE, SAPE, chi-square, PC for the three periods are in favour of the best predictive performance of the same model. The above result shows that the predictive performance of ANN model is the best. Again, the values of Theil's U criterion are nearer to zero for ANN model than other two models which imply that predictive performance of ANN model is the most perfect. So although ARIMA model outperforms other two models only by R^2 (prediction), ANN model outperforms ARIMA and growth models by almost all of the other selected criteria.

It is also clear that so far as the predictive accuracy is concerned, growth model is more competing to ANN model, while ARIMA model is far from ANN but closer to growth model. That is why the tests of equal performance show no significant difference between predictive performance of ANN and growth models, and growth and ARIMA models. Figure 6.1.1. presented in the appendix also reveals that the predictive performances of the ANN model is the best.

From above discussion it is clear that out of the three competing models, the ANN model is the best predictive model of untransformed gas consumption followed by quadratic model and ARIMA model.

6.3.2. *For log-transformed gas consumption*

As in the case of untransformed series the computed values of test statistics for pair wise forecasts accuracy comparisons are presented in table 6.1.3(a) and those of the regression coefficients for the test of encompassing-in-forecasts are presented in table 6.1.3(b). It is evident from table 6.1.3(a) that there is no significant difference between the forecasts' accuracy of growth and ANN models, and that of growth and ARIMA models. However, only MGN test shows a significant difference between the forecasts' accuracy of ARIMA and ANN models. Again, as in the case of untransformed series regression coefficients for the tests of encompassing-in-forecasts show that neither model encompasses either. This may happen due to the same fact as in the case of untransformed series. So, from here it is rather difficult to select the best predictive model. Hence, for this purpose let us have a look at the individual predictive performance comparison.

The values of the selected criteria computed for individual predictive performance comparison are reported in table 6.1.4. The table shows that performance of ARIMA model is the worst among the three competing models. In this case too, growth model is only the competing model to ANN. ANN model outperforms the growth model by R^2, AME , SAPE and RMSE each for prediction and total periods, by MAPE and chi-sqaure each for prediction period only, by PC for all periods. The values of Theil's U for three periods show that predictive performance of ANN is more perfect than that of growth model. Growth model fits the transformed consumption pattern well in estimation period by the criteria R^2, AME, SAPE, RMSE, MAPE and chi-square. It is obvious that sometimes if a model fits well in estimation period, its prediction performance is poor, and the converse is also true. As the study is oriented to the good predictive model and it is clear that no model uniquely supersedes either. Whereas, predictive performance of ANN model supersedes growth model by maximum criteria, unlike other two. Again, figure 6.1.2. presented in the appendix shows that the predicted values obtained by ANN model are more nearer to those of observed values, which implies that performance of the ANN model is the best. So ANN model is considered as the best predictive model for log-transformed series too.

6.3.3. *For untransformed electricity consumption*

The values of the criteria described in section 6.2 are also computed for the three types of competing models of untransformed electricity consumption. The values of the criteria for pairwise accuracy comparison tests, such as Sign test, WSR test and MGN tests are reported in table 6.2.1(a). All the statistics of this table show that there is no significant difference between the predictive accuracy of growth model and ARIMA model, while predictive accuracy of ANN model differs significantly from that of growth model and ARIMA model. Again it is evident from regression coefficients in table 6.2.1(b) that none of the three models encompasses any other. From this table it is rather difficult to take idea about the best predictive model. For selection of the best predictive model it is required to compare the values of the selected criteria for individual models' predictive performance accuracy which are reported in table 6.2.2.

The values of the criteria in table 6.2.2 show that predictive accuracy of growth model and ARIMA model are better than that of ANN model. Again in all the cases except Theil's U statistic, prediction performance of ARIMA model is superior to growth model so far as the accuracy in prediction period is concerned. While growth model seems to be more accurate than ARIMA model by the values of the criteria in estimation period and total period of the study. Only Theil's U statistic is in favour of the superiority of growth model over ARIMA model in prediction period. So it is clearly understood that none of the growth model or ARIMA model supersedes either uniquely. It is known that if a model outperforms in estimation period, it may not predict the future data good, the converse is also true. Same things happened in this case too. Again since most of the values of criteria confirm that performance of growth model is better than ARIMA model and values of Theil's U criteria for growth model in all the three periods are nearer to zero in comparison with those of ARIMA model. Moreover, figure 6.2.1. presented in the appendix also reveals that the predictive performances of growth model is the best, so growth model is selected as the best predictive model of untransformed electricity consumption.

6.3.4. *For log-transformed electricity consumption*

The values of the statistics of pair wise models' prediction accuracy comparison reported in table 6.2.3(a,b) support that there is no significant difference between the prediction accuracy of pairwise competing three models. In order to select the better performed model, let us look at the nature of the values of the criteria of individual model's predictive performance.

It is a difficult job to select a model out of these three considering their individual prediction performance. Because, it is clear from table 6.2.4 that prediction performances of different models are different periods. For example, ANN model performs better for estimation period, growth model performs better for total period, while ARIMA model performs better for prediction period. So it can be inferred that prediction performances of all the three models are almost the same. However, out of 21 values of the criteria growth model outperforms other two by nine values, whereas other two models outperform growth model by six values each. In this case too the figure 6.2.1. presented in the appendix exhibits that the predictive performances of the growth model is better than other two. So, growth model can be considered as comparatively the best model for log-transformed electricity consumption.

6.4. SUMMARY

The final result regarding the selection of the best model from the three competing models after prediction accuracy comparison can be summarized as below.

i) Tests of encompassing-in-forecasts for untransformed gas consumption show that none of the three competing models encompasses any other. MGN and WSR tests for pair wise equal accuracy comparison indicate that there is significant difference between forecasts accuracy of ARIMA and ANN models. On the other hand, no significant difference between the prediction accuracy of the growth and ANN models, and growth and ARIMA models are found by pair wise equal accuracy tests. Again, prediction performance of ANN model supersedes other two models by maximum criteria of individual predictive performance comparison. So, ANN model is selected as the best predictive model for untransformed gas consumption followed by quadratic model and ARIMA model.

ii) Interpretation of the results of pair wise forecasts accuracy tests and individual prediction performance comparison for log-transformed gas consumption is the same as that of untransformed series. Thus, in this case too, ANN model is selected as the best predictive model for log-transformed gas consumption followed by quadratic model and ARIMA model.

iii) For untransformed electricity consumption, tests of encompassing-in-forecasts show no significant difference between the predictive accuracy of competing three models. The sign test, WSR test and MGN test reveal that predictive performance of

ANN model significantly differs from that of growth and ARIMA models. While no significant difference is found in predictive accuracy of growth and ARIMA models. Neither of the growth and ARIMA models uniquely outperforms either by individual predictive performance accuracy tests. However, the values of the most of the criteria are in favour of better performance of growth model. So growth model is considered as the best predictive model for untransformed electricity consumption.

iv) It is found by the tests of encompassing-in-forecasts and pairwise accuracy comparison tests that there are no significant difference between predictive performances of three types of selected models for log-transformed electricity consumption. Individual predictive performance accuracy also reveals that predictive accuracy of all the three models are almost the same. In this case too predictive performance of growth model is a little bit better so far as the values of the criteria are concerned. So, growth model is selected as the best predictive model for log-transformed electricity consumption.

Finally, the predicted series of gas and electricity obtained by the selected models are shown in figures.

APPENDIX VI

Table 6.1.1. Statistics for pair wise prediction accuracy comparison for untransformed gas consumption

(a). Absolute values of z statistics for sign test and WSR test, and t-statistic for MGN test

Type of model	Growth			ARIMA		
	Sign	WSR	MGN	Sign	WSR	MGN
ARIMA	1.091	0.817	0.927			
ANN	0.218	0.298	1.453	1.527	2.092*	3.193*

* Statistical significance at $P \leq 0.025$

(b). Regression coefficient for test of encompassing in-forecasts

Regression coefficient of growth error on ARIMA forecasts is $\beta = -0.0035$ and $t = -0.1624$

Regression coefficient of ARIMA error on Growth forecasts is $\beta = 0.0039$ and $t = 0.0021$

Regression coefficient of growth error on ANN forecasts is $\beta = -0.0037$ and $t = -0.1736$

Regression coefficient of ANN error on Growth forecasts is $\beta = -0.0014$ and $t = 0.0011$

Regression coefficient of ARIMA error on ANN forecasts is $\beta = 0.0053$ and $t = 0.2136$

Regression coefficient of ANN error on ARIMA forecasts is $\beta = 0.0001$ and $t = 0.0072$

Note: None of the regression coefficients is significant at standard level.

Table 6.1.2. The values of the predictive performance comparison criteria for different types of selected models of untransformed gas consumption.

Criteria	Period	MODEL		
		ANN	Growth	ARIMA
R^2	Estimation	0.9879*	0.9851	0.9785
	Prediction	0.9422	0.9382	0.9490*
	Total	0.9948*	0.9919	0.9918
AME	Estimation	3312.0942*	4066.2927	5310.8522
	Prediction	3145.2500*	4134.6883	4831.4556
	Total	3283.0770*	4078.1876	5219.5369
SAPE	Estimation	171.4644*	204.2936	287.1909
	Prediction	7.6890*	9.9787	10.6939
	Total	179.1534*	214.2724	297.8849
RMSE	Estimation	4335.7238*	6013.1105	7216.7655
	Prediction	4466.8274*	11145.0721	10837.3823
	Total	5254.5732*	5605.0194	6858.0550
MAPE	Estimation	13.6285*	14.3148	18.2428
	Prediction	1.9036*	2.4487	2.7156
	Total	11.5895*	12.2511	15.2853
Chi-Square	Estimation	13716.1821*	15576.3501	26143.6423
	Prediction	370.9011*	759.8635	648.7853
	Total	14087.7142*	16336.2131	26792.4303
Theil's U	Estimation	0.0324*	0.0396	0.0477
	Prediction	0.0111*	0.0159	0.0155
	Total	0.0222*	0.0279	0.0326
PC	Estimation	21766685.49*	36481868.98	61272594.48
	Prediction	34916957.70*	217372391.1	205535441.6
	Total	20462445.30*	33882925.95	53751917.34

Note: The value of the criterion for a model with starlet means the performance of that model is superior to other two models with respect to that criterion.

Table 6.1.3. Statistics for pair wise prediction accuracy comparison for log- transformed gas consumption

(a). Absolute values of z statistics for sign test and WSR test, and t-statistic for MGN test

Type of model	Growth			ARIMA		
	Sign	WSR	MGN	Sign	WSR	MGN
ARIMA	0.218	0.643	2.115			
ANN	0.218	0.115	1.115	0.218	0.253	2.767*

* Statistical significance at $P \leq 0.025$

(b). Regression coefficient for test of encompassing in-forecasts

Regression coefficient of growth error on ARIMA forecasts is $\beta = -0.0149$ and $t = -0.6937$

Regression coefficient of ARIMA error on Growth forecasts is $\beta = -0.0037$ and $t = 0.0009$

Regression coefficient of growth error on ANN forecasts is $\beta = 0.0014$ and $t = -0.0683$

Regression coefficient of ANN error on Growth forecasts is $\beta = -0.0164$ and $t = -0.9953$

Regression coefficient of ARIMA error on ANN forecasts is $\beta = 0.0338$ and $t = -1.1499$

Regression coefficient of ANN error on ARIMA forecasts is $\beta = 0.0148$ and $t = 0.82686$

Note: None of the regression coefficients is significant at standard level.

Table 6.1.4. The values of the predictive performance comparison criteria for different types of selected models of log-transformed gas consumption.

Criteria	Period	MODEL		
		ANN	Growth	ARIMA
R^2	Estimation	0.9854	0.9854	0.9652
	Prediction	0.9089*	0.8192	0.8452
	Total	0.9938*	0.9920	0.9815
AME	Estimation	3711.6227	3269.2929*	4790.4799
	Prediction	3749.6950*	5890.7761	9262.3819
	Total	3718.2431*	3725.2028	5642.2707
SAPE	Estimation	192.8109*	178.9051	261.2332
	Prediction	8.9617*	13.8038	20.3505
	Total	201.7723*	192.7190	281.5837
RMSE	Estimation	4761.4763	4710.5899*	7586.1916
	Prediction	4777.5243*	9235.1433	21413.7997
	Total	4654.2433*	4692.1355	8380.7155
MAPE	Estimation	12.7085	10.1431*	17.8713
	Prediction	2.2273*	3.4100	5.2699
	Total	10.8857	8.9721*	15.4711
Chi-Square	Estimation	12616.1620	9766.9191*	20547.0221
	Prediction	416.5779*	1234.0931	1648.7853
	Total	13032.7423	11001.0184*	22195.8074
Theil's U	Estimation	0.0356*	0.0357	0.0501
	Prediction	0.0118*	0.0207	0.0307
	Total	0.0243*	0.0276	0.0395
PC	Estimation	23864874.63*	29633147.08	67706239.42
	Prediction	28530923.24*	366239152.2	802463932.8
	Total	22603805.78*	33307761.85	80270163.03

Note: The value of the criterion for a model with starlet means the performance of that model is superior to other two models with respect to that criterion.

Table 6.2.1. Statistics for pair wise prediction accuracy comparison for untransformed electricity consumption

(a). Absolute values of z statistics for sign test and WSR test, and t-statistic for MGN test

Type of model	Growth			ARIMA		
	Sign	WSR	MGN	Sign	WSR	MGN
ARIMA	0.259	0.762	0.072			
ANN	2.840*	2.222*	5.321*	2.840*	1.968*	5.193*

* Statistical significance at $P \leq 0.05$

(b). Regression coefficients for test of encompassing-in-forecasts

Regression coefficient of growth error on ARIMA forecasts is ß = 0.0287 and t =0.083

Regression coefficient of ARIMA error on Growth forecasts is ß = -0.0157 and t = -0.457

Regression coefficient of growth error on ANN forecasts is ß = 0.00038 and t = 0.013

Regression coefficient of ANN error on Growth forecasts is ß = -0.2114 and t =-1.926

Regression coefficient of ARIMA error on ANN forecasts is ß = -0.0138 and t = -0.490

Regression coefficient of ANN error on ARIMA forecasts is ß = -0.2061 and t = -1.8256

Note: No regression coefficient is significant at standard level.

Table 6.2.2. The values of the predictive performance comparison criteria for different types of selected models of untransformed electricity consumption.

Criteria	Period	MODEL		
		ANN	Growth	ARIMA
R^2	Estimation	0.9361	0.9884*	0.9851
	Prediction	0.5248	0.7915	0.8076*
	Total	0.9438	0.9875*	0.9871
AME	Estimation	313.3704	102.6636*	143.3811
	Prediction	781.8690	334.8150	252.4369*
	Total	368.4879	129.9755*	157.9219
SAPE	Estimation	127.2776	57.4052*	71.3826
	Prediction	27.7871	12.4546	9.4108*
	Total	155.0648	69.8599*	80.7935
RMSE	Estimation	486.4364	145.9366*	186.2591
	Prediction	876.7573	479.9085	254.8609*
	Total	559.1353	186.2950*	199.3390
MAPE	Estimation	10.8505	4.1606*	5.4436
	Prediction	15.2988	6.3264	4.7569*
	Total	11.3739	4.4154*	5.3520
Chi-Square	Estimation	653.9875	86.6212*	121.0636
	Prediction	252.8073	42.5393	24.1682*
	Total	906.7949	129.1605*	137.8372
Theil's U	Estimation	0.0070	0.0016*	0.0329
	Prediction	0.0051	0.0011*	0.0341
	Total	0.0054	0.0016*	0.0330
PC	Estimation	283944.5	25557.0015*	42698.4022
	Prediction	1921758.8	287890.2210	162385.2045*
	Total	6780211.7	33882925.95*	53751917.34

Note: The value of the criterion for a model with starlet means the performance of that model is superior to other two models with respect to that criterion.

Table 6.2.3. Statistics for pair wise prediction accuracy comparison for log- transformed electricity consumption

(a). Absolute values of z statistics for sign test and WSR test, and t- statistic for MGN test

Type of model	Growth			ARIMA		
	Sign	WSR	MGN	Sign	WSR	MGN
ARIMA	0	1.015	0.565			
ANN	0	0.537	1.489	0.500	1.373	1.352

* Statistical significance at P ≤ 0.025

(b). Regression coefficient for test of encompassing forecasts for log-transformed electricity consumption

Regression coefficient of growth error on ARIMA forecasts is $\beta = 0.0052$ and $t = 0.1758$

Regression coefficient of ARIMA error on Growth forecasts is $\beta = -0.0719$ and $t = -1.471$

Regression coefficient of growth error on ANN forecasts is $\beta = 0.00074$ and $t = 0.0219$

Regression coefficient of ANN error on Growth forecasts is $\beta = 0.0515$ and $t = 1.3552$

Regression coefficient of ARIMA error on ANN forecasts is $\beta = -0.0733$ and $t = -1.3718$

Regression coefficient of ANN error on ARIMA forecasts is $\beta = 0.05379$ and $t = 1.5495$

Note: No regression coefficient is significant at standard level.

Table 6.2.4. The values of the predictive performance comparison criteria for different types of selected models of log-transformed electricity consumption.

Criteria	Period	MODEL		
		ANN	Growth	ARIMA
R^2	Estimation	0.9883	0.9885*	0.9810
	Prediction	0.4314	0.7801	0.9898*
	Total	0.9796	0.9883*	0.9848
AME	Estimation	100.9510*	103.2713	137.1802
	Prediction	433.9585	325.5142*	305.8175*
	Total	140.1283	120.4173	158.2598
SAPE	Estimation	59.9441	58.6182*	71.3504
	Prediction	15.7015	12.1597	11.2346*
	Total	75.6492	70.7778*	82.5850
RMSE	Estimation	149.8713*	149.8785	199.7811
	Prediction	738.6789	475.7247	368.6351*
	Total	241.2024	188.2014*	220.6352
MAPE	Estimation	4.0295*	4.7885	5.1118
	Prediction	7.4885	6.2061	5.7933*
	Total	4.4365*	4.4565	5.1970
Chi-Square	Estimation	85.2909*	93.0154	132.8633
	Prediction	103.2420	41.7302	35.0897*
	Total	188.5329	134.7443*	167.9530
Theil's U	Estimation	0.0240*	0.0241	0.0361
	Prediction	0.0491	0.0305*	0.0400
	Total	0.0343	0.0265*	0.0369
PC	Estimation	26885.4006*	26956.2714	48466.4420
	Prediction	1364116.21	281204.4510	169724.9100*
	Total	68445.4311	41590.4562*	57807.3122

Note: The value of the criterion for a model with starlet means the performance of that model is superior to other two models with respect to that criterion.

Fig 6.1.1 Fitted Values and Observed Series of Gas Consumption. (Untransformed)

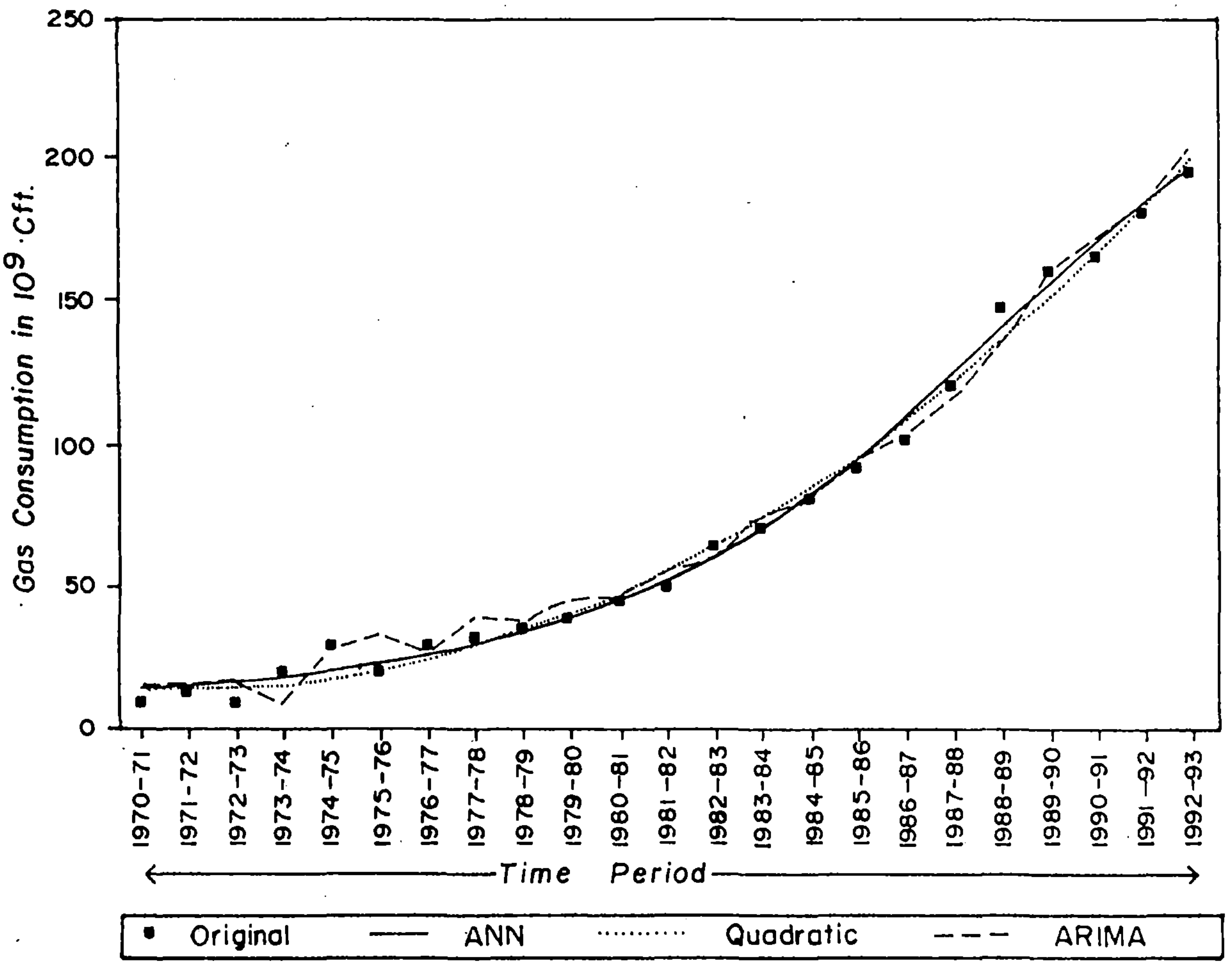

Fig.6.1.2 Fitted Values and Observed Series of Gas Consumption (Log-transformed)

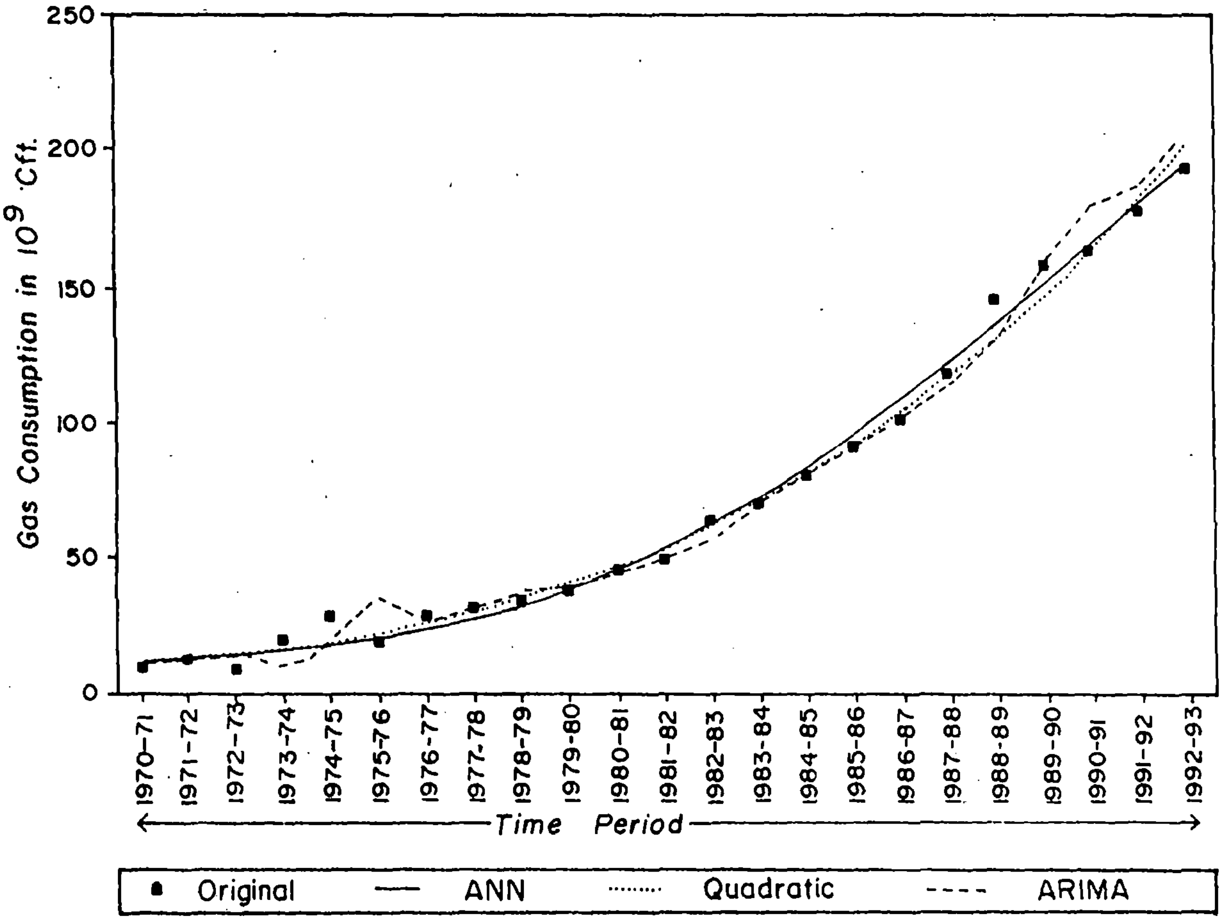

Fig.6.2.1 Fitted Values and Observed Series of Electricity Consumption (*Untransformed*)

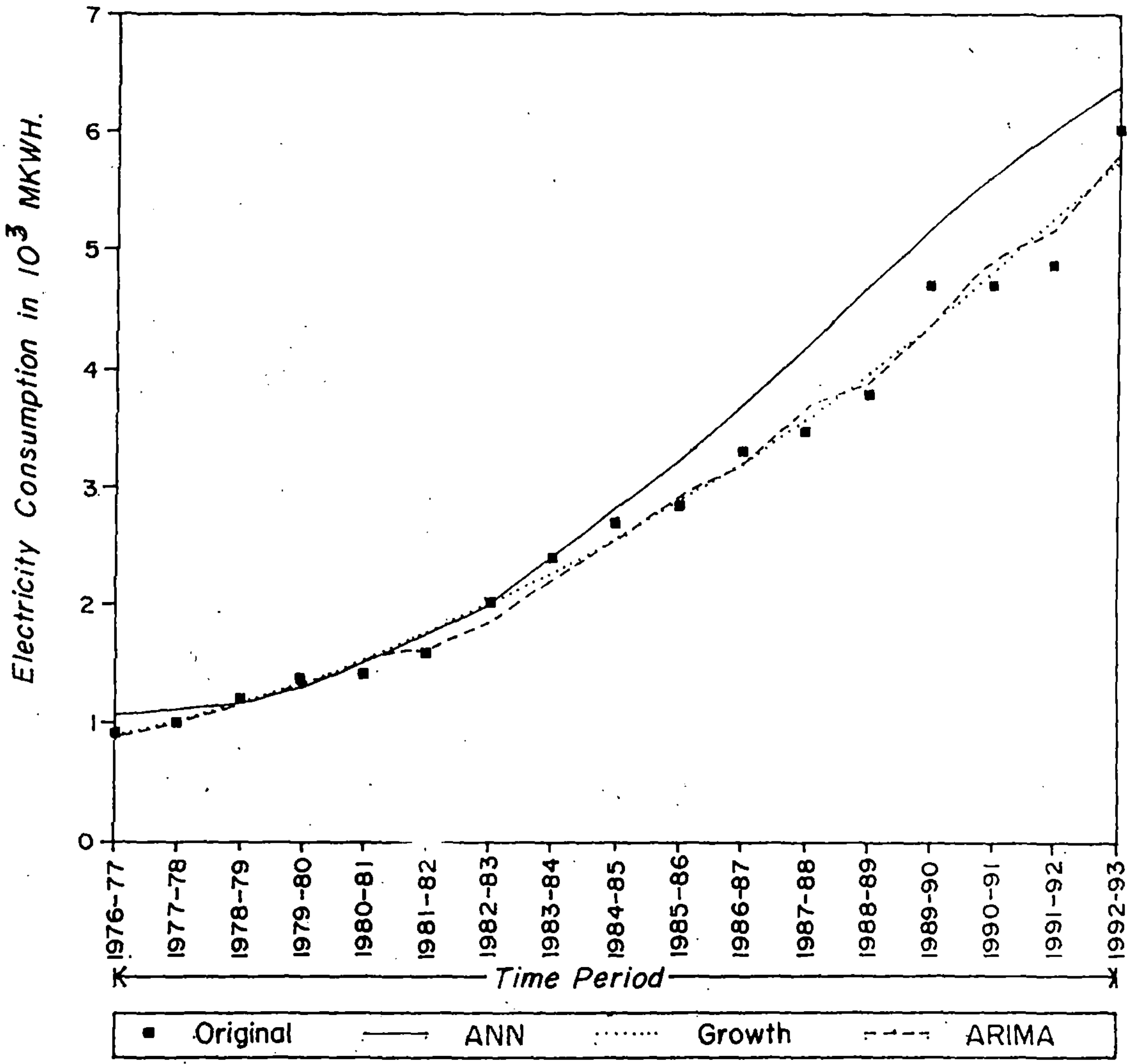

Fig. 6.2.2 Fitted Values and Observed Series of Electricity Consumption (*Log-transformed*)