CHAPTER - II

SALIENT FEATURES OF ENERGY CONSUMPTION AND STRUCTURAL RELATIONSHIP

Contents **Page No.**

CHAPTER-II

SALIENT FEATURE OF ENERGY CONSUMPTION AND STRUCTURAL RELATIONSHIP

In order to get some idea about the nature of data used in the study, some statistical/ econometric properties of the energy consumption data are demonstrated in this chapter. Thus, in section 2.1 statistical properties of gas and electricity consumption data are given, in section 2.2 correlation of gas and electricity consumption with some selected factors are discussed, in section 2.3 structural relationship (multivariate regression models) of both types of energy consumption data are obtained using two types of methods for selecting the subset regressors and finally, in section 2.4 findings of this chapter are presented. Numerical results obtained for the analysis are presented in the appendix in last part of the chapter.

2.1. STATISTICAL PROPERTIES OF ENERGY CONSUMPTION

Yearly average consumption, standard deviation and consumption growth rates of gas and electricity in Bangladesh are given below:

2.1.1. For gas consumption

In the study period the average annual gas consumption (in 10^6 cft.) in Bangladesh is 73646.261 and standard deviation of gas consumption is 59338.952. Both are statistically significant at $P<0.001$. Over the period from 1970-71 to 1992-93 total gas consumption increases 1964.89%. The highest average annual gas consumption growth (46.15%) occurs in the period 1975-76 to 1976-77, while the lowest (-

30.23%) occurs in just preceding period 1974-75 to 1975-76. The average annual growth rate over the period is 17.82%. Negative growths occur in two periods. Average gas consumption efficiency is 91.68%, while the remaining 8.32% is due to system loss. Note that this energy efficiency is defined as the percentage of energy used per unit of output (Wilson *et.al.*1994; p.287).

2.1.2. For electricity consumption

The per year electricity consumption (in MKWH) in Bangladesh is 2844.47 and standard deviation is 1559.70. These two statistics are also statistically significant at 0.01 level. Over the period from 1976-77 to 1992-93 total electricity consumption increases 546.03%. The highest average annual electricity consumption growth (27.23%) occurs in the period 1981-82 to 1982-83, while the lowest (0.23%) occurs in 1989-90 to 1990-91. The average annual growth rate over the period is 12.66%. Average electricity consumption efficiency could not be computed due to unavailability of annual total electricity production in Bangladesh.

2.2. CORRELATION WITH OTHER FACTORS

Linear correlation coefficient of total gas consumption (GC) and electricity consumption (EC), in Bangladesh with some available and apparently correlated factors are shown the table 1.1. The factors considered in the study are annual production of gas (GP), annual consumption of petroleum and petroleum products (PETC), gross domestic products in factor costs (GDP) and annual population (POP). It would be better if the factors like average temperature, per unit price of gas and electricity and areas under gas and electricity supply were consider in the study, but due to unavailability these factors could not be considered.

For computing the correlation matrix, data from 1976-77 to 1992-93 have been used for all the factors, because data for electricity consumption are available only for that period. In addition to see the strength of relationship between pair wise factors, this correlation matrix is constructed with a view to selecting the regressors for appropriate regression model by factor analysis given in section 2.3.2.

It is clear from correlation matrix (table 2.1 in the appendix) that gas consumption is significantly strongly correlated with all the factors considered in the study such as gas production, petroleum consumption, electricity consumption, gross domestic product and population per annum. All of the above factors have positive correlation with gas consumption in Bangladesh. Electricity consumption is also strongly linearly correlated with gas consumption, petroleum consumption, gross domestic product and population.

2.3. STRUCTURAL RELATIONSHIP

In following subsections attempt is made to observe the structural relationship of gas and electricity consumption to identify the factors responsible for changes in energy consumption. The analysis is performed considering the factors found to be correlated in table 2.1. The tentatively postulated form of regression model for gas consumption with full set of regressors is

$$GC_t = \beta_0 + \beta_1 GP_t + \beta_2 EC_t + \beta_3 PETC_t + \beta_4 GDP_t + \beta_5 POP_t + \epsilon_t \quad \text{.............................(2.1)}$$

Similarly, for electricity consumption is

$$EC_t = \beta_0 + \beta_1 GC_t + \beta_2 PETC_t + \beta_3 DP_t + \beta_4 POP_t + \epsilon_t \quad \text{..............................(2.2)}$$

where, t stands for time and ϵ_t is the residual term assumed to be identically and independently distributed as normal with mean zero and constant variance σ^2. The appropriate regression model may contain full or subset of the above-mentioned 5 regressors for gas consumption and 4 regressors for electricity consumption. Thus in order to select the most outfitted subset of regressors for gas and electricity consumption the following methodological tools are adopted.

2.3.1. Step-wise regression

To find the best subset regressors for regression model to explain gas and electricity consumption in Bangladesh, in this section regression models of the form (2.1) for gas consumption and form (2.2) for electricity consumption using all possible subset regressors are fitted. There are $2^5 = 32$ equations with 32 subset regressors for gas consumption and $2^4 = 16$ models with 16 subset regressors for electricity consumption. This type of step-wise regression technique is advocated by Amemya (1976), Drapper and Smith (1966), Montogomery and Peck (1993) among others.

2.3.1.1. Criteria for selecting appropriate subset regressors

The following selection criteria are used for selecting the appropriate subset regressors for regression model.

MSE, the mean square error: The mean square error is defined as

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

where, N is the number of observations and k is the number of regressors used in the model.

$\overline{R}^2$, ***the adjusted coefficient of determination.***: It might be objected that R^2, the coefficient of multiple correlation or coefficient of determination defined as $R^2 = 1 - SS(ERROR)/SS(TOTAL)$ can overstate the value of a regression fit since the quantity SS(ERROR) can be reduced simply by adding further explanatory variables, even if they are not relevant to explaining the explained variable. Amemiya (1976, p.1) also restates the same drawback in different language that R^2 has an obvious weakness *i.e.* it can be maximised by maximising the number of regressors. Therefore, some kind of correction that accounts for the degree of freedom is necessary. Thus the alternative of the R^2 denoted by $\overline{R}^2$ that corrects for degree of freedom in estimating the error variance and variance of explained variables is defined as

$$\overline{R}^2 = 1 - \frac{N-1}{N-k}(1-R^2)$$

where, N is the number of observations used in model fitting and k is the number of parameters estimated. It is to note that in case of particularly bad fit, $\overline{R}^2$ can be negative and it does not exist when $n<k$. Theil (1961; p.177), Granger and Newbold (1986,p.191), Johnston (1991, p.177) and other econometricians used and recommended this alternative.

***C_p, the Mallows criterion*:** Mallows (1964;1973, p.662) and Montgomery and peck (1992, p.271) proposed a criterion named as C_p-plot, a graphical display device, which helped the researcher for selection of subset regressors, defined as

$$C_p = \frac{1}{\hat{\sigma}^2} RSS_p - N + 2p$$

where, p is the number of variables used in a model, N is the sample size, RSS_p is the corresponding residual sum of squares of a model and $\hat{\sigma}^2$ is the unbiased estimate of residual variances. Frequently, the MSS of the full equation is used as $\hat{\sigma}^2$. In this study too, MSS(error) of full equation (MSS of last row of respective table) is used as $\hat{\sigma}^2$. Mallows (1973, p.671) also stated that using the minimum C_p-plot rule to select a subset of terms for least squares fitting could not be recommended universally. However, that by examining this C_p plot one can see whether or not a single best subset is uniquely indicated, the ambiguous cases, where the 'minimum C_p ' rule will give bad results are exactly those where a large number of subsets are close competitors for the honour. With such data no selection rule can be expected to perform reliably.

However, in order to take decision regarding number of regressors in a model, it is necessary to construct a plot of C_p as a function of p, for each regression equation. Regression model with little bias will have values of C_p that fall near the line $C_p = p$, while those equations with substantial bias will deviate more from respective p values. Although generally small values of C_p are more desirable, it may be preferable to accept some bias in the equation to reduce the average error of prediction.

Amemiya (1976) and Montgomery and Peck (1993) advocated the use of this C_p criterion for selection of subset regression model and this criterion was also used by

Hurvich and Tsai(1989) for selection of order of autoregression model. In chapter IV, this C_p criterion is used for selecting the order of autoregression too.

PC, the prediction criterion: Amemiya (1972; 1976, p.7) suggested the use of a criterion named, prediction Criterion, defined as

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

for the selection of subset regressors of regression model, where the notations have the usual meaning.

PRESS Criterion: The prediction error sum of squares (PRESS) initially proposed by Allen (1971 p.47; 1974 p.126) which provided a useful residual scaling. The computational formula of the statistic given by Montogomery and Peck (1992 p.275) is

$$PRESS = \Sigma (\frac{e_t}{1 - h_{tt}})^2$$

where, e_t is the deviation of actual value from predicted value at time t and h_{tt} is the (t,t)th element of Hat matrix $X'(X'X)^{-1}X$. Montgomery and Peck also stated that this statistic was potentially useful for discriminating between alternative models, and hence very much useful for measuring the prediction accuracy of a model.

2.3.1.2. Results of step-wise regression analysis for gas consumption

Regression models of type (2.1) are fitted to the gas consumption data using all the 2^5

= 32 subsets of regressors. The parameters are estimated applying most popularly used least square methods. Outputs of all the models are not reported in the study, some of the models whose fitness is better are reported. Note that this fitness is judged on the values of R^2. The output obtained for step-wise regression analysis for a selected subset of regressors for gas consumption are reported in tables 2.2.1 and 2.2.2.

From tables 2.2.1 and 2.2.2, it is evident that all the coefficients for all the considered models are significant at 0.01 level. MSE is minimum for the model with GP and EC as the regressors, $\overline{R}^2$ is maximum for the same models, value of C_p is minimum and at the same time closer to the value of p for that model, while C_p for all other models deviate much more from the value of respective p's, and finally, value of the PC is also minimum for the model with GP and EC as the regressors. So by step-wise regression methodology model with GP and PC as the regressors are considered to be the best model for gas consumption as per the above-mentioned criteria are concerned. It is also clear from the values of D-W statistics that there exists a positive autocorrelation problem in all the selected equations.

2.3.1.3. Result of step-wise regression analysis for electricity consumption

Like gas consumption data regression models of type (2.2) are fitted to the electricity consumption data using all the 2^4 = 16 subsets of regressors. In this case too, the parameters are estimated applying most popularly used least square methods. Outputs of all the models are not reported in the study, some of the models whose fitness is better are reported in tables 2.3.1 and 2.3.2.

From tables 2.3.1 and 2.3.2, it is evident that all the coefficients for all the considered models are significant at 0.01 level. MSE is minimum for the model with GC and PETC as the regressors, $\overline{R}^2$ is maximum for same models, value of C_p is minimum and at the same time closer to the value of p for model with GC and PETC as the regressors, while C_p for all other models deviate much more from the values of respective p's, and finally, value of the PC is also minimum for the model with GC and PETC as the regressors. So, by step-wise regression methodology as per the selected criteria are concerned the model of type (2.2) with GC and PETC as the regressors are considered to be the best model for electricity consumption. In this case too, it is clear from the values of D-W statistics that there exists a positive autocorrelation in all the selected equations. There is also a multicollinearity problem.

2.3.2. Selection of subset regressors by factor analysis

Factor analysis, a branch of statistical science, is a method of data reduction that assumes a specific underlying model. Factor analysis reduces data by extracting the more significant factor(s) from a set of factors in such a way that the extracted factors can specify a model adequately. The extraction of the factors can be performed by two methods such as maximum likelihood method and principal component method. In the study the former one is adopted. For details of the methodology see Srivastava and Carter (1983). Factor analysis for selecting the appropriate subset regressors out of a set of regressors is used and recommended by Engel and Watson (1981) and Elston and Proe (1995). Thus, for selecting the more influential factors for formulating regression model for both the gas and electricity consumption in Bangladesh, a factor analysis is performed. The factors considered for gas consumption are GP, EC, PETC, POP and GDP. While for electricity consumption, the factors are GC, PETC, POP and GDP. The output and discussion of factor analysis are given below.

2.3.2.1. Result of factor analysis for gas consumption

From table 2.4.1 it is evident that maximum likelihood method extracts two factors such as GP and EC as significant among the considered five factors. These two factors explain 95.64% of total variation together. So it can be said that the subset regressors {GP, EC} are enough for explaining the gas consumption. Step-wise regression analysis also gives the same result. Hence model

$$GC = -3400.31 + 0.89\ GP + 23.22\ EC \qquad \text{(2.3)}$$

with the criteria given in tables 2.2.1 and 2.2.2, is the finally selected regression model for gas consumption in Bangladesh.

2.3.2.2. Result of factor analysis for electricity consumption

Table 2.4.2 shows that maximum likelihood method extracts two factors such as GC and PETC, for which again values are greater than unity, as significant among the considered four factors. These two factors explain 93.27% of total variation together. So it can be said that the subset regressors {GC, PETC} are enough for explaining the electricity consumption. Step-wise regression analysis also gives the same result, although a multicollinearity problem exists there. However, model like

$$EC = 1672.84 + 0.359\ GC - 1.251\ PETC \qquad \text{.............................(2.4)}$$

with the criteria given in tables 2.3.1 and 2.3.2, is the finally selected regression model for electricity consumption in Bangladesh.

Forecasts from the selected models may be attributed to specification errors. The forecast efficiency can be improved by incorporating the variables like per unit price, temperature and area under energy consumption.

2.4. SUMMARY

It is observed from the analysis in this chapter that

i) Average gas consumption (GC) in Bangladesh is 73646.261 X 10^6 cft. per annum and that of electricity consumption (EC) is 2844.47 MKWH, gas consumption efficiency in Bangladesh is 91.68%.

ii) There exists strong correlation of gas consumption with GP, EC, PETC, POP and GDP. Similarly, electricity consumption is strongly correlated with GC, PETC, POP and GDP.

iii) Step-wise regression analysis for selecting the subset of regressors using the selection criteria like MSE, $\overline{R}^2$, C_p, PC, PRESS shows that {GP, EC} are so far the best subset regressors for expressing the gas consumption by linear regression model. The same type of analysis shows that {GC, PETC} are the best subset regressors for electricity consumption.

iv) Factor analysis by maximum likelihood method extracts the same two factors GP and EC as in step-wise regression analysis for gas consumption which together explain

95.64% of total variation. Similarly, in case of electricity consumption factor analysis extracts the same two factors GC and PETC as in step-wise regression which together explain 93.27% of total variation.

v) Finally, regression model of the type (2.3) is selected for gas consumption and regression model of the type (2.4) is selected for electricity consumption in Bangladesh. Autocorrelation problem exists in both the models. Montogomery and Peck (1992, p.366) stated that there were several causes of autocorrelation, perhaps the primary cause of autocorrelation in regression problems involving time-series data was failure to include one or more regressors in the model. In addition, a multicollinearity problem exists in model (2.4) which may cause over specification of the regression models. Due to limitation of scope (lack of data and factors) the above-mentioned problems could not be removed.

However, emphasis will not be given in structural model for forecasting purposes due to the reason mentioned in chapter 3. So in subsequent chapters investigation toward the fitness of classical growth models, univariate Box-Jenkins' type of model and artificial neural network type of models are undertaken.

APPENDIX-II

Table 2.1 Pair wise correlation coefficient between different factors

Facotrs	GP	GC	EC	PETC	POP	GDP
GP	1.000					
GC	0.998*	1.000				
EC	0.491	0.998*	1.000			
PETC	0.626	0.850*	-0.837*	1.000		
POP	0.358	0.861*	0.884*	0.605*	1.000	
GDP	0.992*	0.782*	0.631*	0.938*	0.584	1.000

* Significant at 0.05 level.

Table 2.2.1 Step-wise regression output for gas consumption

Regressors in the model	β_0	coefficients of				
		GP	EC	PETC	POP	GDP
GP	-2734.8	1.11* (92.62)				
EC	-7701.5		35.69* (27.30)			
GP EC	-3400.3	0.98* (18.78)	23.22* (11.23)			
GP PETC EC	-2932.7	0.75* (15.93)	7.18* (49.54)	3.33* (23.51)		
EC PETC POP	-1803.1		23.87* (19.31)	10.59* (30.25)	3.66* (27.62)	
GP PETC EC POP	-4235.4	0.93* (62.91)	8.59* (33.61)	1.38* (18.73)	1.14* (26.90)	
GP PETC EC GDP POP	-358.29	0.98* (62.91)	12.65* (8.95)	6.15* (21.54)	2.32* (31.52)	0.75* (14.70)

* significant at 0.05 level.
Values in the parentheses are the absolute values of corresponding t-statistics.

Table 2.2.2. Values of diagnostic criteria of considered regressions for gas consumption

Regressors	M S E	$\overline{R}^2$	C_p	PRESS	PC	D-W
GP	275401.6	0.984	3.88	35703.7	287375.6	0.395
EC	305612.5	0.973	5.27	41230.4	318899.5	0.226
GP EC	251255.5	0.989	2.71	27125.5	273101.4	0.147
GP PETC EC	372152.6	0.965	13.8	45111.3	420693.6	0.159
EC PETC POP	581543.2	0.933	28.9	71234.5	657396.4	0.323
GP PETC EC POP	402314.0	0.955	14.6	53352.4	472281.6	0.493
GP PETC EC GDP POP	276991.9	0.981	7.00	31892.2	337207.6	0.115

Table 2.3.1. Step-wise regression output for electricity consumption

Regressors in the model	β_0	Coefficient of			
		GC	PETC	POP	GDP
GC	267.51	0.027* (27.30)			
PETC	1958.22		-0.079* (31.51)		
POP	5463.23			3.721* (51.29)	
GC PETC	1672.84	0.359* (11.35)	-1.251* (12.21)		
GC PETC POP	7140.55	0.172* (19.33)	-0.471* (23.12)	10.323* (12.22)	
GC PETC POP GDP	6289.94	0.893* (22.51)	-0.338* (19.57)	2.915* (35.61)	15.88* (8.99)

* significant at 0.05 level.
Values in parentheses are the absolute values of corresponding t-statistics.

Table 2.3.2. Values of diagnostic criteria of considered regressions for electricity consumption

Regressors	MSE	$\overline{R}^2$	C_p	PRESS	PC	D-W
GC	45091.15	0.974	9.77	20511.1	47743.57	0.422
PETC	51362.30	0.938	12.9	29402.5	54385.61	0.356
POP	110356.5	0.645	42.7	81539.6	11848.06	0.237
GC PETC	29531.82	0.983	2.98	11253.9	33006.13	0.298
GC PETC POP	62795.71	0.896	18.7	46822.7	73877.29	0.159
GC PETC POP GDP	31679.33	0.979	6.00	17361.3	39135.50	0.478

Table 2.4.1. Output of factor analysis for gas consumption (extraction of factors by maximum likelihood method)

Initial Statistics Variable	Communality	*	Factor	Eigen value	% of Var.	Cum. %
GP	1.000	*	1	2.9873	59.83	59.83
EC	1.000	*	2	1.7880	35.81	95.64
PETC	1.000	*	3	0.1010	2.03	97.67
GDP	1.000	*	4	0.0920	1.84	99.51
POP	1.000	*	5	0.0243	0.49	100.00

ML extracted two factors
**Chi-square = 5.624, DF= 3

Final statistics

Variable	Communality	*	Factor	Eigen value	% of Var.	Cum. %
GP	0.9586	*	2	2.9873	59.83	59.83
EC	0.9634	*		1.7880	35.81	95.64
PETC	0.8199	*				
GDP	0.6769	*				
POP	0.9384	*				

** upper tailed tests

Table 2.4.2. Output of factor analysis for electricity consumption (extraction of factors by maximum likelihood method)

Initial Statistics Variable	Communality	*	Factor	Eigen value	% of Var.	Cum. %
GC	1.000	*	1	2.8961	59.22	59.22
PETC	1.000	*	2	1.6654	34.05	93.27
GDP	1.000	*	3	0.2037	4.16	97.43
POP	1.000	*	4	0.1255	2.57	100.00

ML extracted two factors
**Chi-square = 6.223, DF= 1

Final statistics Variable	Communality	*	Factor	Eigen value	% of Var.	Cum. %
EC	0.9511	*	2	2.8961	59.22	59.22
PETC	0.9032	*		1.6654	34.05	93.27
GDP	0.7156	*				
POP	0.8677	*				

** upper tailed tests