The Pakistan Development Review
31 : 3 (Autumn 1992) pp. 287–293

Probability Distribution of Electricity Demand Forecast for the City of Karachi

SALIM CHISHTI

A prerequisite for developing the planning and demand management policies for energy is the appropriate projections of future demand. Capacity targets may be set particularly more efficiently in the light of a knowledge of the probability of shortages. However, such probabilities are usually not available. Such probability distributions are generated in this study. More specifically, a simple four-equation model has been developed to make projections of average and peak demand for the city of Karachi in the year 2000. Three of the four equations have been estimated using Full Generalised Least Squares with Prais-Winsten transformation in order to correct serially correlated errors. The estimated model, after making appropriate tests for hetroscedasticity, has been put to recursive bootstrapping to generate probability distributions of average and peak demand in order to assess the extent of uncertainty in point projections. Bootstrapping has been used because of the limitations of the conventional method with regard to imposing a pre-specified stochastic structure on the error term and assuming the knowledge of the values of the exogenous variables for the forecast period with certainty. Probability distributions are based on 1000 random samples. The results indicate that under quite plausible assumptions, the extent of uncertainty remains significant which should be taken into account for future policy planning.

1. INTRODUCTION

Over the last thirty years the annual growth in Pakistan's GDP has been 5.94 percent whereas electricity consumption has grown at about 12.56 percent. Karachi is the largest city of Pakistan. Its population, which was less than 0.2 million in 1947, has now reached 10 million approximately. Karachi Electric Service Corporation, being the main utility agency providing electricity to the city has more than one million metered consumers. The growth of electricity consumption in Karachi during the period 1974–1990 has been 6.97 percent per year. The pattern of consumption has also shown significant changes. For instance, the residential sector, which accounted for 18.7 percent of total consumption in 1974 now consumes 35 percent. The current installed capacity is 1318 MW as against the current peak demand of 1123 MW. The installed capacity has in fact remained above the peak demand upto now. But by the year 1997-98 the planned capacity of KESC would reach 2055 MW. This will be achieved by commissioning three additional plants of 210 MW each whereas some of the older plants producing a total of

Salim Chishti is Research Professor at the Applied Economics Research Centre, University of Karachi, Karachi.

298 MW will be retired. This planned capacity would certainly be insufficient to satisfy the expected increase in demand. Therefore, the private sector is also being involved to generate additional capacity. However, in order to do appropriate planning for generation of new capacity and develop appropriate demand management policies a prerequisite is the availability of reliable projections of average and peak demand as well as their probability distributions. This is essentially what has been attempted here. Non-parametric recursive bootstrapping technique has been employed to estimate the probability distribution.

The paper is organised as follows. Section 2 consists of a brief discussion of the model and data. Estimation and results are reported in Section 3. The last section presents the summary and conclusion.

2. MODEL AND DATA

The model consists of the following equations:

$$LPR = \alpha_0 + \alpha_1 T + \mu_1 \quad \ldots \quad \ldots \quad \ldots \quad (1)$$

$$LGDP = \beta_0 + \beta_1 T + \mu_2 \quad \ldots \quad \ldots \quad \ldots \quad (2)$$

$$LAMW = \tau_0 + \tau_1 LGDP + \tau_2 LPR + \mu_3 \quad \ldots \quad \ldots \quad (3)$$

$$LPEAK = \delta_0 + \delta_1 LAMW \quad \ldots \quad \ldots \quad \ldots \quad (4)$$

where *LPR* = Log of average sale price charged by KESC;
LGDP = Log of regional gross domestic product of Karachi;
LAMW = Log of the total units of electricity sold by KESC; and
LPEAK = Log of the peak demand.

Equations 1 and 2, although very simple, have been estimated with reasonable precision and are capable of providing reliable projections. The inclusion of these two equations has enabled us to treat price and income as stochastically determined instead of making a rather tenuous assumption that they are exogenously given with certainty for the forecasting period. The third equation is a conventional demand function with price and income as the dependent variables. Instead of taking peak demand as a constant proportion of average demand we have estimated it empirically using Equation 4.

The equations of the estimated model are put to a series of diagnostic tests before using them to make forecasts and develop corresponding probability distributions. Once appropriate point estimates of the required forecasts are obtained the model has been recursively bootstrapped to obtain the probability distributions of the average demand, peak demand and of forecast errors conditional on current information. Instead of using the bootstrapping technique one could get confidence intervals around the forecast by using the conventional method. However this approach has serious limitations because it requires to pre-specify a stochastic structure of the error. A more restrictive assumption for the validity of this approach is the knowledge of the values of the exogenous variables for the forecast period with certainty. Bootstrapping is a computer-intensive simulation technique. The main idea consists of re-sampling the residuals from a fitted relation a large number of times, preserving the own stochastic structure of the relationship and using these residuals to generate an empirical distribution which serves as an estimate of the theoretical distribution of the error term. Recursively bootstrapping a forecast based on a system of equations primarily consists of using the Monte-Carlo sample of residuals to construct 'pseudo-data' at each stage for a large number of times and using these data sets successively in the next higher equation [Jean and Vealp (1987)]. More details of the procedure are given in the next section.

The data covers a period of 17 years from 1974 to 1991. The main source is the Power System Statistics [KESC (1990)]. The estimates of the GDP of Karachi have been taken from Bengali (1990). It may also be mentioned here that this series is available only upto the period 1985, and therefore for the remaining four years it has been estimated by us by first estimating a linear relationship between the GDP of Karachi and of the province of Sindh. This estimated relationship is used to project Karachi GDP for the remaining years.

3. ESTIMATION AND RESULTS

Following are the OLS estimates of Equations 1 and 2

$$LPR = -0.78 + \underset{(6.3)}{0.138T} \qquad R^2 = 0.92 \qquad DW = 0.4 \qquad (1)$$

$$LGDP = 5.66 + \underset{(21.8)}{0.148T} \qquad R^2 = 0.99 \qquad DW = .97 \qquad (2)$$

Durbin-Watson values in both the equations show autocorrelation. Therefore both the equations have been re-estimated through Full Generalised Least Square using the Prais-Winsten transformation [Prais-Winsten (1954); Green (1990)] weighting the first observation by $(1-\delta^2)1/2$. The resulting equations were put to the Breusch-Pagan-Godfrey test of hetroscedasticity which led to the conclusion that the residuals are not hetroscedastic.

OLS estimates of Equation (3) are as follows:

$$LAMW = -1.57 + 0.741\,LGDP \quad -0.188LPR \quad \ldots \quad \ldots \qquad (3)$$

$$R^2 = 0.95 \qquad DW = 0.65$$

The Durbin-Watson value corresponding to this equation also indicates autocorrelation. Therefore, this equation has also been re-estimated with Generalised Least Squares using the Prais-Winsten transform. The resulting equation is as follows:

$$LAMW = -0.167 + \underset{(3.8)}{0.553\,LGDP} \quad \underset{(1.5)}{-0.023LPR} \quad \ldots \quad (3')$$
$$R^2 = .98 \quad DW = 1.7$$

Equation (3) is then put to some diagnostic tests. The goodness of fit test for normality of residuals yields an x^2 value of 6.8 with 1 d. f. The corresponding critical value at 5 percent level of significance is 7.8. The Jarque-Bera Asymtotic *LM* normality test yields an x^2 value of 0.74 with the critical value being 10.59 with 2 d. f. The Breusch-Pagan-Godfrey test of hetroscedasticity yields an x^2 value of 0.366 with 2 d. f. The corresponding critical value is 10.59. Therefore the hypothesis of homoscedasticity cannot be rejected, which is a pre-requisite for the validity of the bootstrapping estimation.

The demand for electricity according to the above equation seems to be inelastic with regard to income as well as price.

The OLS estimates of Equation (4) are as follows:

$$LPEAK = 0.95 + 0.95LAMW \quad \ldots \quad \ldots \quad \ldots \quad (4)$$
$$R^2 = .98 \quad DW = 1.8$$

This equation has also been put to similar diagnostic tests. The goodness of fit test for normality of residuals gives an x^2 value of 1.3 with 2 d. f. The Jarque-Bera asymtotic normality test gives an x^2 value of 0.99 with 2 d. f. The corresponding critical value at 5 percent is 10.59. The Breusch-Pagan-Godfrey test of hetroscedasticity gives an x^2 value of 0.374 with 7.8 being the critical value at the 5 percent significance level leading to accept the hypothesis of homoscedasticity.

4. BOOTSTRAPPING

To start with, 1000 random samples of 17 observations are obtained from the GLS residuals of Equation (1) and Equation (2). However, since the Prais-Winsten transformations were used in their estimation, the original stochastic structure had to be preserved by reversing the Prais-Winsten transformation. This is done by dividing the first element of the GLS residuals by $(1-\delta^2)$ and the successive residuals obtained using *AR*(1) process. These randomly re-sampled residuals are then added to the predicted *GLS* estimates of Equations (1) and (2) to obtain pseudo-data corresponding to *LPR* and *LGDP*. For future reference let us call them *Y*1 and *Y*2. These 1000 pseudo-data sets of 17 observation of *Y*1 and *Y*2 are then used as independent variables in Equation (3) to re-estimate it 1000 times. Finally, the *LPR* and *LGDP* forecasts obtained from Equations (1) and (2) for the year 2000 are used in these estimated equations to produce 1000 different forecasts of demand. (The residuals corresponding to these 1000 equations and the predicted values of *LAMW* will be used later. For future reference let us call them *LAME* and μ_3 respectively). the probability distribution of the forecasts of the average demand is presented in Diagram 1.

The mean of the probability distribution of the demand in electricity for Karachi in the year 2000 comes to 1347 MW. The 80 and 90 percent upper ceiling are 1481 MW and 1557 MW respectively. The standard error of the distribution is 169 MW.

In order to obtain the probability distribution of the peak demand, we first obtained 1000 random sets from μ_3 referred to above. However, since μ_3 were based on *AR*(1) with Prais-Winsten transformation, they were first re-transformed by reversing this Prais-Winsten transformation as described above. These sets of residuals are then added to the predicted values *LAME* to obtain pseudo-data on *LAMW*. Let us call it Y_3. These Y_3 values are then used as an independent variable to re-estimate Equation (4) again 1000 times. The pseudo-data corresponding to

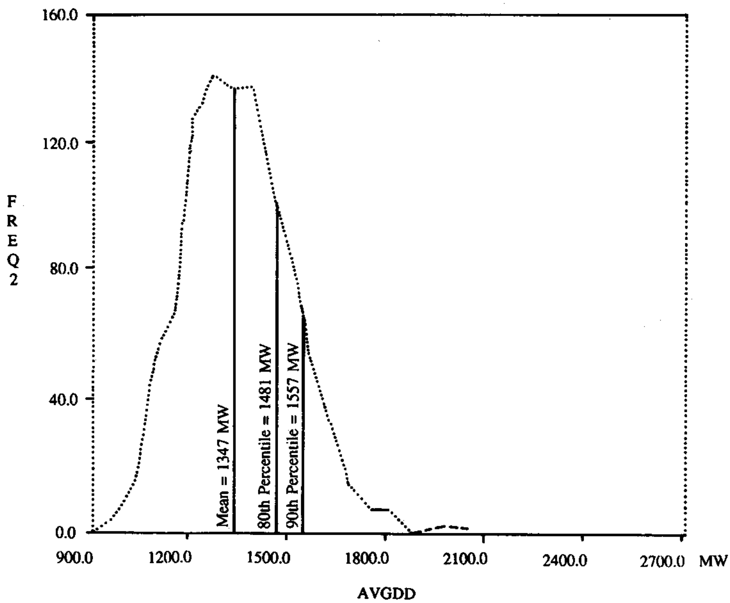

Diagram 1. Probability Distribution of Electricity Demand Forecast in Karachi for the Year 2000

LPEAK is then generated using the same procedure as for Y_3. Let us call it Y_4. Then Y_4 is regressed on Y_3, again 1000 times to generate 1000 sets of estimated coefficients. The bootstrapped forecasts of average demand are then used along with these coefficients to generate 1000 different estimates of *PEAK* demand. The resulting probability distribution of *PEAK* demand based on these estimates is presented in Diagram 2.

The mean of the peak demand distribution is 2458 MW. The 80th and 90th percentiles of the distribution are 2722 MW and 2919 MW respectively. The standard error of the distribution is 337 MW.

5. SUMMARY AND CONCLUSION

This article has provided the forecasts of electricity demand in Karachi and has illustrated the use of bootstrapping to estimate the extent of uncertainty in the regression forecasts. The point made here is that, even under somewhat conservative assumptions that there would be no structural shifts during the forecast period, the extent of uncertainty in the regression forecast are large when the uncertainty in the forecasts of the independent variables is also assumed. This type of approach may be of some help to utility agencies like the KESC to devise future demand management policies and plan for new capacity.

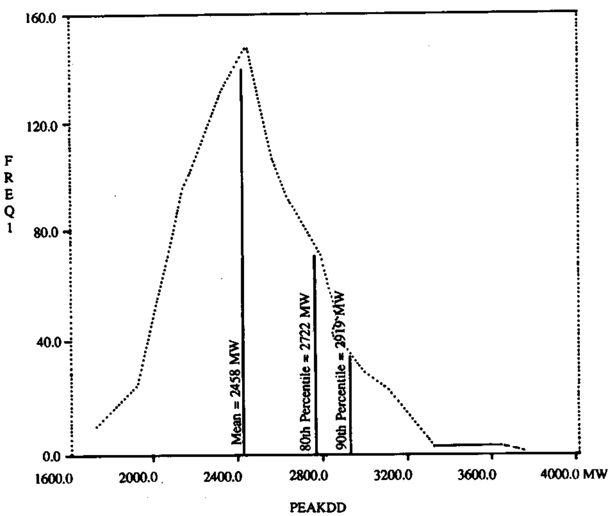

Diagram 2. Probability Density of the Forecast of Peak Demand of Electricity in Karachi for the Year 2000

REFERENCES

Bengali, K. (1988) The Economy of Karachi: Growth and Structural Change. Karachi: AERC, University of Karachi. (Research Report No. 62.)

Breusch, T. S., and A. R. Pagan (1979) A Simple Test for Hetroskedasticity and Random Coefficient Variation. *Econometrica* 47: 1287 – 1294.

Green, William H. (1990) *Econometrics Analysis*. London: MacMillan Publication Company.

Jean, T. B., and M. R. Vealp (1987) The Probability Distribution of Future Demand: The Case of Hydro Quebec. *Journal of Business and Economic Statistics* 5: 417 – 424.

Karachi Electric Supply Corporation Ltd. (1990) *Power System Statistics*. Karachi: Management Studies Department.

Prais, G. J., and C. B. Winsten (1954) Trend Estimators and Serial Correlation. Chicago: Cowles Commission. (Discussion Paper 383.)