CHAPTER 8
MODELING ADOPTION OF ENERGY EFFICIENT TECHNOLOGIES AND CONSERVATION HABITS IN INDIAN URBAN HOUSEHOLDS – A QUANTILE REGRESSION APPROACH

8.1 INTRODUCTION
It is said that a student scores at the $\tau^{th}$ quantile of a standardized exam if he performs better than the proportion ‘$\tau$’ of the reference group of students and worse than the proportion $(1-\tau)$. Thus, half of students perform better than the median student and half perform worse. Similarly, the quartiles divide the population into four segments with equal proportions of the reference population in each segment. The quintiles divide the population into five parts; the deciles into ten parts. The quantiles, or percentiles, or occasionally fractiles, refer to the general case. Quantile regression as introduced by Koenker and Bassett (Koenker and Hallock, 2001), seeks to extend these ideas to the estimation of conditional quantile functions—models in which quantiles of the conditional distribution of the response variable are expressed as functions of observed covariates. Quoting Mosteller and Tukey from their text book, Koenker and Hallock make this concluding remark in their paper: 'What the regression curve does is give a grand summary for the averages of the distributions corresponding to the set of $x$’s'. One could go further and compute several different regression curves corresponding to the various percentage points of the distributions and thus get a more complete picture of the set. Ordinarily this is not done, and so regression often gives a rather incomplete picture. Just as the mean gives an incomplete picture of a single distribution, so does the regression curve gives a corresponding incomplete picture for a set of distributions.

The main advantage of quantile regression over least-squares regression is its flexibility for modeling data with heterogeneous conditional distributions. Data of this type occur in many fields, including econometrics, survival analysis, and ecology (Koenker and Hallock, 2001). Quantile regression provides a complete picture of the covariate effect when a set of percentiles is modeled, and it makes no distributional assumption about the error term in the model.
Various aspects of quantile regression, including algorithms for estimating regression coefficients, confidence intervals, statistical tests, detection of leverage points and outliers, and quantile process plots have been discussed in detail by Colin (Lin) Chen and Cary in one of their papers on Quantile regression (Chen and Cary, 2005).

The purpose of regression analysis is to expose the relationship between a response variable and predictor variables. In real applications, the response variable cannot be predicted exactly from the predictor variables. Instead, the response for a fixed value of each predictor variable is a random variable. For this reason, we often summarize the behavior of the response for fixed values of the predictors using measures of central tendency. Typical measures of central tendency are the average value (mean), the middle value (median), or the most likely value (mode).

Traditional regression analysis is focused on the mean; that is, one summarizes the relationship between the response variable and predictor variables by describing the mean of the response for each fixed value of the predictors, using a function referred to as the conditional mean of the response. The idea of modeling and fitting the conditional-mean function is at the core of a broad family of regression-modeling approaches, including the familiar simple linear-regression model, multiple regression model, models with hetero-scedastic errors using weighted least squares, and nonlinear regression models (Hao and Naiman, 2007).

Conditional-mean models have certain attractive properties. Under ideal conditions, they are capable of providing a complete and parsimonious description of the relationship between the covariates and the response distribution. In addition, using conditional-mean models leads to estimators (least squares and maximum likelihood) that possess attractive statistical properties, are easy to calculate, and are straightforward to interpret. Such models have been generalized in various ways to allow for hetero-scedastic errors so that given the predictors, modeling of the conditional mean and conditional scale of the response can be carried out simultaneously.
Conditional-mean modeling has been applied widely in the social sciences, particularly in the past half century, and regression modeling of the relationship between a continuous response and covariates via least squares and its generalization is now seen as an essential tool. More recently, models for binary response data, such as logistic and probit models and Poisson regression models for count data, have become increasingly popular in social science research. These approaches fit naturally within the conditional mean modeling framework. While quantitative social-science researchers have applied advanced methods to relax some basic modeling assumptions under the conditional-mean framework, this framework itself is seldom questioned.

The conditional-mean framework has inherent limitations. First, when summarizing the response for fixed values of predictor variables, the conditional-mean model cannot be readily extended to non-central locations, which is precisely where the interests of social-science research often reside. For instance, studies of economic inequality and mobility have intrinsic interest in the poor (lower tail) and the rich (upper tail). Educational researchers seek to understand and reduce group gaps at pre-established achievement levels (e.g., the three criterion-referenced levels: basic, proficient, and advanced). Thus, the focus on the central location has long distracted researchers from using appropriate and relevant techniques to address research questions regarding non-central locations on the response distribution. Using conditional-mean models to address these questions may be inefficient or even miss the point of the research altogether.

Second, the model assumptions are not always met in the real world. In particular, the homoscedasticity assumption frequently fails, and focusing exclusively on central tendencies can fail to capture informative trends in the response distribution. Also, heavy-tailed distributions commonly occur in social phenomena, leading to a preponderance of outliers. The conditional mean can then become an inappropriate and misleading measure of central location, because it is heavily influenced by outliers.

Third, the focal point of central location has long steered researchers’ attention away from the properties of the whole distribution. It is quite natural to go beyond location and scale effects of predictor variables on the response and ask how changes in the predictor variables affect the
underlying shape of the distribution of the response. For example, much social-science research focuses on social stratification and inequality, areas that require close examination of the properties of a distribution. The central location, the scale, the skewness, and other higher-order properties—not central location alone—characterize a distribution. Thus, conditional-mean models are inherently ill equipped to characterize the relationship between a response distribution and predictor variables. Examples of inequality topics include economic inequality in wages, income, and wealth; educational inequality in academic achievement; health inequality in height, weight, incidence of disease, drug addiction, treatment, and life expectancy; and inequality in wellbeing induced by social policies. These topics have often been studied under the conditional-mean framework, while other, more relevant distributional properties have been ignored.

An alternative to conditional-mean modeling has roots that can be traced to the mid-18th century. This approach can be referred to as conditional median modeling, or simply median regression. It addresses some of the issues mentioned above regarding the choice of a measure of central tendency. The method replaces least-squares estimation with least-absolute distance estimation. While the least-squares method is simple to implement without high-powered computing capabilities, least-absolute-distance estimation demands significantly greater computing power. It was not until the late 1970s, when computing technology was combined with algorithmic developments such as linear programming, that median-regression modeling via least-absolute-distance estimation became practical. The median-regression model can be used to achieve the same goal as conditional-mean-regression modeling: to represent the relationship between the central location of the response and a set of covariates. However, when the distribution is highly skewed, the mean can be challenging to interpret while the median remains highly informative. As a consequence, conditional-median modeling has the potential to be more useful.

The median is a special quantile, one that describes the central location of a distribution. Conditional-median regression is a special case of quantile regression in which the conditional 0.5th quantile is modeled as a function of covariates. More generally, other quantiles can be used
to describe noncentral positions of a distribution. The quantile notion generalizes specific terms like quartile, quintile, decile, and percentile. The pth quantile denotes that value of the response below which the proportion of the population is p. Thus, quantiles can specify any position of a distribution. For example, 2.5% of the population lies below the 0.025th quantile.

The quantile-regression model is a natural extension of the linear-regression model. While the linear-regression model specifies the change in the conditional mean of the dependent variable associated with a change in the covariates, the quantile regression model specifies changes in the conditional quantile. Since any quantile can be used, it is possible to model any predetermined position of the distribution. Thus, researchers can choose positions that are tailored to their specific inquiries like poverty studies concerning the low-income population; tax-policy studies concerning the rich and so on. Since multiple quantiles can be modeled, it is possible to achieve a more complete understanding of how the response distribution is affected by predictors, including information about shape change. A set of equally spaced conditional quantiles (e.g., every 5% or 1% of the population) can characterize the shape of the conditional distribution in addition to its central location. The ability to model shape change provides a significant methodological leap forward in social research on inequality. Traditionally, inequality studies are non-model based; approaches include the Lorenz curve, the Gini coefficient, Theil’s measure of entropy, the coefficient of variation, and the standard deviation of the log-transformed distribution (Hao and Naiman, 2007).

Quantile-regression models can be easily fit by minimizing a generalized measure of distance using algorithms based on linear programming. As a result, quantile regression is now a practical tool for researchers. Software packages familiar to social scientists offer readily accessed commands for fitting quantile-regression models. A decade and a half after Koenker and Bassett first introduced quantile regression, empirical applications of quantile regression started to grow rapidly. Empirical researchers took advantage of quantile regression’s ability to examine the impact of predictor variables on the response distribution. But, looking at the available literature it is felt that the application of Quantile regression technique in India is yet to pick up. Balasubrahmanya and Sudhir Kumar have applied this technique for their study of Technological
innovations and energy intensity of machine tool SMEs in Bangalore in the southern state of India (Balasubrahmany and Sudhir Kumar 2011). Nikhil Kaza has applied Quantile regression technique to study the residential energy consumption in United States, but it is limited to only the dwelling characteristics (Kaza, 2010). Possibly, this study is the first instance where Quantile Regression has been used to model household energy consumption in India.

8.2 MODELING ADOPTION OF EETs USING QUANTILE REGRESSION

The OLS (Ordinary Least Square) regression discussed under section 7.2 has revealed that the 'Financial status' is the most important factor and the 'Willingness to adopt EETs' is the least important factor influencing the adoption of energy efficient devices in Indian urban households. But as discussed above, for evolving policy analyses targeting different tiers of energy consumers, it was decided to use Quantile regression approach to generate additional input. For adoption of EETs, it was decided to observe whether there is any difference in the perception of households towards EETs, at the lower and higher quantiles using only nine factors as the fifth factor was found to be statistically not significant under OLS. For ease of analysis and interpretation of results, it is decided to observe the same at 30%, 60% and 85% quantiles. The entire Quantile regression analysis has been carried out using the EasyReg International software developed by Professor Herman J. Bierens of Pennsylvania State University. For the outputs obtained from this analysis, refer Appendix 8. The results of the Quantile Regression Analysis at 30%, 60% and 85% quantiles are shown in Table 8.1. A discussion of the obtained results is presented below:

The OLS coefficient of the first factor namely the 'Financial status' is 0.336. At 30% quantile it is 0.316, at 60% it is 0.350 and at 85% it is 0.378. So, it is seen that the financial status becomes more significant in explaining the adoption of energy efficient devices at higher quantiles. The OLS coefficient of the second factor, 'Attitude towards EET' is 0.088. At 30% quantile it is 0.095, at 60% it is 0.072 and at 85% it is 0.085. It may be recalled from the discussion under section 6.2 that, the variables considered under the second factor are Ego factor and Technology savvy nature of the households. Hence, for the households in the lower quantiles, being proud
about their EETs and showing interest in technical devices, is more dominant in explaining the adoption of energy efficient devices.

### Table 8.1 Results of Quantile Regression on Adoption of EETs as Dependent Variable

<table>
<thead>
<tr>
<th>Quantiles</th>
<th>30%</th>
<th>60%</th>
<th>85%</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.183090 (77.350) [0.00000]</td>
<td>2.477843 (91.360) [0.00000]</td>
<td>2.796134 (77.379) [0.00000]</td>
<td>2.389</td>
</tr>
<tr>
<td>b1</td>
<td>0.316054 (11.177) [0.00000]</td>
<td>0.350739 (12.908) [0.00000]</td>
<td>0.378280 (10.449) [0.00000]</td>
<td>0.336</td>
</tr>
<tr>
<td>b2</td>
<td>0.095638 (3.382) [0.00072]</td>
<td>0.072408 (2.665) [0.00770]</td>
<td>0.085020 (2.348) [0.01885]</td>
<td>0.088</td>
</tr>
<tr>
<td>b3</td>
<td>0.209598 (7.413) [0.00000]</td>
<td>0.252353 (9.287) [0.00000]</td>
<td>0.239361 (6.612) [0.00000]</td>
<td>0.230</td>
</tr>
<tr>
<td>b4</td>
<td>0.122428 (4.330) [0.00001]</td>
<td>0.110068 (4.051) [0.00005]</td>
<td>0.117137 (3.236) [0.00121]</td>
<td>0.118</td>
</tr>
<tr>
<td>b6</td>
<td>0.190116 (6.724) [0.00000]</td>
<td>0.166769 (6.137) [0.00000]</td>
<td>0.225212 (6.221) [0.00000]</td>
<td>0.201</td>
</tr>
<tr>
<td>b7</td>
<td>0.054006 (1.910) [0.05614]</td>
<td>0.058291 (2.145) [0.03193]</td>
<td>0.046475 (1.284) [0.19923]</td>
<td>0.068</td>
</tr>
<tr>
<td>b8</td>
<td>-0.011688 (-0.413) [0.67935]</td>
<td>-0.046647 (-1.717) [0.08603]</td>
<td>-0.129044 (-3.564) [0.00036]</td>
<td>-0.054</td>
</tr>
<tr>
<td>b9</td>
<td>0.126063 (4.458) [0.00001]</td>
<td>0.078038 (2.872) [0.00408]</td>
<td>0.111739 (3.086) [0.00203]</td>
<td>0.111</td>
</tr>
<tr>
<td>b10</td>
<td>-0.048775 (-1.725) [0.08453]</td>
<td>-0.033681 (-1.240) [0.21515]</td>
<td>-0.015489 (-0.428) [0.66878]</td>
<td>-0.057</td>
</tr>
</tbody>
</table>

(t values are in parentheses) [p values are in parentheses]

The third factor, namely ‘Awareness about energy and environment' shows a consistent increasing trend in explaining the adoption of EETs, from lower to higher quantiles (From 60% to 85% there is a very small reduction in its strength). The fourth factor, the ‘Cost driven concern
for environment', shows a slight dip at 60% quantile and then increases again at higher quantiles in its importance in explaining the adoption of EETs.

The sixth factor, the 'Dwelling characteristics' (includes the variables: ownership of residence and degree of satisfaction) also shows a jump at higher quantiles. Dwelling characteristics is the third most important factor in all the quantiles. The seventh factor, the 'Attitude towards risk' (the important variables here being Risk coverage, Government subsidies and Government regulations) shows a greater dominance at lower quantiles.

The eighth factor, the 'Government efforts' (includes variables like Government incentives, Government efforts in facilitating EETs and Adequacy of information about EETs) shows a steep increase in its importance at higher quantiles. Also it is statistically not significant at lower quantiles (p value 0.68). The reasons for the negative beta values of this factor have already been discussed under section 7.2. The ninth factor, the 'family size' is found to be more important only at lower and higher quantiles. Hence, for a household which is average in the adoption of EETs, family size is not such an influencing factor. The tenth factor, the 'Qualification of the head of the household', takes a sudden dip in its importance at 85% quantile. It is also not statistically significant at all quantiles (more so at 85% quantile with p value of 0.67).

The graphical plots for the two most dominant factors namely, the 'Financial status' and 'Awareness about energy and environment' have been obtained by performing simple quantile regression. The intercepts and the coefficients for the three selected quantiles are shown in Table 8.2 and the plots are shown in Figures 8.1 and 8.2. (The objective of the graphical plot is only to illustrate as to how the slopes vary for different quantiles and as an example the plot is drawn for the two most dominant factors with OLS coefficients 0.336 and 0.230 respectively). The main thing to be observed here is the way in which the slopes of the lines change with different quantiles, indicating that the strength of the influence of these factors significantly change for different quantiles.
### Table 8.2 Results of Simple Quantile Regression on Adoption of EETs as Dependent Variable

<table>
<thead>
<tr>
<th>Factor</th>
<th>Quantile</th>
<th>Intercept</th>
<th>B value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial status</td>
<td>30%</td>
<td>2.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>60%</td>
<td>2.4995</td>
<td>0.3946</td>
</tr>
<tr>
<td></td>
<td>85%</td>
<td>2.9378</td>
<td>0.4648</td>
</tr>
<tr>
<td>Awareness about energy and environment</td>
<td>30%</td>
<td>2.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>60%</td>
<td>2.4435</td>
<td>0.375</td>
</tr>
<tr>
<td></td>
<td>85%</td>
<td>3.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Figure 8.1 Quantile Regression on 'Financial status'**
8.3 MODELING ADOPTION OF CHs USING QUANTILE REGRESSION

The OLS stepwise regression of the level of energy conservation habits resulted in four factors explaining the level of conservation habits in Indian urban households. These four factors in order of their strength of relationship were: 'Cost driven concern for environment', 'Awareness about energy and environment', 'Attitude towards EETs' and 'Dwelling characteristics'. Behavioral Economic Theory (BET) suggests four categories of consumers based on environmental and cognitive parameters (Faiers et al., 2007). It says that consumers can be influenced in their moods or desires through communication techniques and will respond, in part, according to their level of ‘education’ in respect of the product in question. The four categories are ‘initiators’, ‘early’ and ‘later’ imitators, and finally, ‘last’ adopters. Within the categories, the consumers would demonstrate their type of behaviour along a high-low continuum according to the level of antecedent influences, such as their ‘moods’, their ‘ability to pay’, levels of ‘deprivation’, and the level of their ‘learning’ in respect of the product or service. With this backdrop, it was felt that there is a need to look into the behavioural attitudes of households at
different quantiles with regard to the adoption of CHs. The quantile regression results for this at 30%, 60% and 90% quantiles are shown in Table 8.3.

**Table 8.3 Results of Quantile Regression on 'Level of Conservation habits' as Dependent Variable**

<table>
<thead>
<tr>
<th>Quantiles</th>
<th>30%</th>
<th>60%</th>
<th>90%</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.429636</td>
<td>3.070667</td>
<td>3.892149</td>
<td>2.867</td>
</tr>
<tr>
<td></td>
<td>(44.901) [0.00000]</td>
<td>(53.236) [0.00000]</td>
<td>(50.962) [0.00000]</td>
<td></td>
</tr>
<tr>
<td>b2</td>
<td>0.181659</td>
<td>0.219228</td>
<td>0.321584</td>
<td>0.223</td>
</tr>
<tr>
<td></td>
<td>(3.351) [0.00081]</td>
<td>(3.794) [0.00015]</td>
<td>(4.203) [0.00003]</td>
<td></td>
</tr>
<tr>
<td>b3</td>
<td>0.269822</td>
<td>0.291020</td>
<td>0.309957</td>
<td>0.251</td>
</tr>
<tr>
<td></td>
<td>(4.977) [0.00000]</td>
<td>(5.036) [0.00000]</td>
<td>(4.051) [0.00005]</td>
<td></td>
</tr>
<tr>
<td>b4</td>
<td>0.600192</td>
<td>0.724901</td>
<td>0.845145</td>
<td>0.716</td>
</tr>
<tr>
<td></td>
<td>(11.071) [0.00000]</td>
<td>(12.544) [0.00000]</td>
<td>(11.045) [0.00000]</td>
<td></td>
</tr>
<tr>
<td>b6</td>
<td>-0.030392</td>
<td>-0.171596</td>
<td>-0.184652</td>
<td>-0.092</td>
</tr>
<tr>
<td></td>
<td>(-0.561) [0.57505]</td>
<td>(-2.969) [0.00298]</td>
<td>(-2.413) [0.01581]</td>
<td></td>
</tr>
</tbody>
</table>

b2 – Coefficient of factor 'Attitude towards EET'
b3 – Coefficient of factor 'Awareness about energy and environment'
b4 – Coefficient of factor 'Cost driven concern for environment'
b6 – Coefficient of factor ' Dwelling characteristics'

The factor 'Attitude towards EETs', shows a strong increasing trend in its influence over the level of conservation habits at higher quantiles. The factor 'Awareness about energy and environment' also shows an increasing trend though not as strong as other factors. The factor 'Cost driven concern for environment' which is the strongest factor influencing the energy conservation habits of Indian urban households according to this survey, also shows an increasing trend from lower to higher quantiles peaking to 0.845 at 90% quantile. The factor 'Dwelling characteristics' shows a substantial slump from -0.03 to -0.172 corresponding to 30% to 60% quantile and a marginal slump from -0.172 to -0.185 corresponding to 60% to 90% quantile. It is to be noted that at 30%
quantile, its coefficient is not statistically significant (p value 0.575). The reasons for negative sign have been explained under section 7.3.

The graphical plots for the two most dominant factors in this case namely, the 'Cost driven concern for environment' and 'Awareness about energy and environment' were again obtained by performing simple quantile regression. The intercepts and the coefficients for the three selected quantiles are shown in Table 8.4 and the plots are shown in Figures 8.3 and 8.4.

**Table 8.4 Results of Simple Quantile Regression on Adoption of CHIs as Dependent Variable**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Quantile</th>
<th>Intercept</th>
<th>B value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness about energy and environment</td>
<td>30%</td>
<td>2.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>60%</td>
<td>3.00</td>
<td>0.2923</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>4.4532</td>
<td>0.3787</td>
</tr>
<tr>
<td>Cost driven concern for environment</td>
<td>30%</td>
<td>2.3837</td>
<td>0.5527</td>
</tr>
<tr>
<td></td>
<td>60%</td>
<td>3.0138</td>
<td>0.7793</td>
</tr>
<tr>
<td></td>
<td>90%</td>
<td>4.0047</td>
<td>0.7612</td>
</tr>
</tbody>
</table>
Figure 8.3 Quantile Regression on Cost driven concern for environment

Figure 8.4 Quantile Regression on Awareness about energy and environment
8.4 SUMMARY

Careful targeting of public policy towards energy consumption and conservation is not possible if only effects on the averages are considered. Policy analyses using standard OLS techniques are not particularly suitable to target conservation policies towards high energy consumers. Quantile regression, on the other hand, estimates the effects of individual independent variables on specific quantiles of the dependent variable and therefore offers a simple method to estimate how different tiers of energy consumers respond to changes in the dependent variables. It may be tempting to pick a subsample of high energy users and analyze the effects of various variables to target policies that promote energy efficiency and conservation in that group. However, such analyses are clouded by sampling bias. Quantile regression approach uses the entire sample to estimate the effect on the distribution, not merely the data in the neighborhood of the quantile of interest of the dependent variable. One of the advantages of quantile regression approach is to understand the differential effect of variables on the entire distribution of consumption. Looking only at the average effect of the independent variable masks the impact and the explanatory power of the dependent variable. In this research study, to achieve the objective of examining whether there is any significant differences in the importance of the influencing factors at higher and lower quantiles, Quantile regression technique was used. It showed that all the factors influencing the adoption of EETs and CHs showed significant differences at lower and higher quantiles. These findings have important bearing on policy implications, manufacturers of EETs and NGOs which shall be discussed in the next chapter.