Density Based Regression for Inhomogeneous Data: Application to Lottery Experiments

Kontek, Krzysztof

Artal Investments

21 April 2010

Density Based Regression for Inhomogeneous Data: Application to Lottery Experiments

Krzysztof Kontek[1]
Artal Investments, Warsaw[2]

Abstract

This paper presents a regression procedure for inhomogeneous data characterized by varying variance, skewness and kurtosis or by an unequal amount of data over the estimation domain. The concept is based first on the estimation of the densities of an observed variable for given values of explanatory variable(s). These density functions are then used to estimate the relation between all the variables. The mean, quantile (including median) and mode regression estimators are proposed, with the last one appearing to be the maximum likelihood estimator in the density based approach. The paper demonstrates the advantages of the proposed methodology, which eliminates most of the estimation problems arising from data inhomogeneity. These include the computational inconveniences of the standard quantile and mode regression techniques. The proposed methodology, when applied to lottery experiments, makes it possible to confirm and to extend the previously presented conclusion (Kontek, 2010) that lottery valuations are only nonlinear with respect to probability when medians and means are considered. Such nonlinearity disappears once modes are considered. This means that the most likely behavior of a group is fully rational. The regression procedure presented in this paper is, however, very general and may be applied in many other cases of data inhomogeneity and not just lottery experiments.

JEL classification: C01, C13, C16, C21, C46, C51, C81, C91, D03, D81, D87

Keywords: Density Distribution; Least Squares, Quantile, Median, Mode, Maximum Likelihood Estimators; Lottery experiments; Relative Utility Function; Prospect Theory.

[1] The author is grateful to Stefan Traub from the Department of Economics, University of Bremen, Ulrich Schmidt from the Department of Economics, University of Kiel and Katarzyna Idzikowska from the Center of Economic Psychology and Decision Sciences, Kozminski University, Warsaw, for making the results of their experiments available. This paper would never have been written without their data.
[2] Contact: ul. Chrościckiego 93/105, 02-414 Warsaw, Poland,
e-mail: kontek@artal.com.pl, kkontek2000@yahoo.com.

1. Introduction

1.1. There are at least three ways in which data can be inhomogeneous. The first, frequently referred to as heterogeneity in the micro-economics literature, is encountered when the object or system described consists of multiple items having a large number of structural variations.

The second type is when the observed variable changes its properties over the estimation domain. An example of this is heteroskedasticity, meaning that variance varies over time. This type of inhomogeneity is, however, broader in meaning as higher moments of the distribution (in particular skewness and kurtosis) may vary as well. An example of this is changes in stock market indexes, which are usually negatively skewed during declines and positively skewed during gains. In this paper, the term inhomogeneity refers to this second type of inhomogeneity unless otherwise stated.

The third type of inhomogeneity is encountered when the data are not uniformly distributed in the estimation domain. This can occur when there are different amounts of data for given values of explanatory variables and/or when the distribution of explanatory variable values is non-uniform. For example, the amount of data on hourly wages can differ for different values of the explanatory variable "Years of Schooling" (see Cameron, Trivedi, 2005, Figure 9.3). Obviously this type of inhomogeneity can result from the nature of the phenomenon under observation and can be deliberate in experiments as in the case of stratified sampling.

All these types of inhomogeneity have to be carefully checked and addressed using proper econometric methods in order to derive the correct relationships between the observed and explanatory variables.

1.2. Each type of data inhomogeneity presented above is encountered in experiments determining the certainty equivalents of lotteries. This was considered in detail in another paper by Kontek (2010), although the term "inhomogeneity" did not appear there.

To recapitulate, certainty equivalents are typically determined in lottery experiments for several combinations of outcomes and for several probabilities of winning. A wide range of certainty equivalent values was observed for specific probabilities in two analyzed sets of data. This points to very diverse risk attitudes among the examined subjects. Different risk attitudes were stated earlier by Tversky and Kahneman (1992) and by other researchers. These observations point to the first type of inhomogeneity (heterogeneity) described in 1.1.

As stated, the variance, skewness and kurtosis of relative certainty equivalents vary

substantially with probability. The most striking results were observed for skewness. The data are positively skewed for low probabilities, negatively skewed for high probabilities and not skewed for medium probabilities. These results point to the second type of inhomogeneity described in point 1.1.

Finally, different numbers of lotteries are examined for specific probabilities in every set of data under consideration. Moreover, those probabilities are non-uniformly distributed in the range [0,1]. This is an example of the third type of inhomogeneity described in point 1.1.

1.3. Kontek (2010) presented a wide range of regression methods in use for lottery experimental data. These include standard least squares (mean), quantile (including median), and mode estimators, all performed parametrically and nonparametrically.

Despite presenting several important conclusions, standard regression methods were not able to dispel all of the doubts regarding the interpretation of the obtained results. The standard procedures assume variance to be constant over the estimation domain. Varying variance, or heteroskedasticity, requires more advanced procedures like weighted, generalized or feasible generalized least squares estimators (Cameron, Trivedi, 2005). However, the problem of varying skewness or kurtosis is not considered even in the advanced textbooks on regression methods as in the one mentioned above. This makes it difficult to predict how the stated inhomogeneity of the data will impact the estimation results. For example, in one of the examined sets, the median and mode regression estimations are practically the same but the mean estimation is different. As all these estimations should be different when the data are skewed, this can raise some objections as to the result.

Additionally the standard median and (especially) mode estimators are characterized by computational inconveniences, which may lead to difficulties in finding the global optimum. This raises the question of whether, and if so how, these inconveniences might be overcome. The important question of how to define the maximum likelihood estimator for the data considered has also been left open.

1.4. In view of the above, a novel approach is proposed in this paper, an approach based on parametrical estimation of densities of certainty equivalents (observed variable) for given probability values (explanatory variable). These density distributions are later used for mean, quantile and mode estimations of the relationship between the two variables. As shown, the proposed methodology results in the maximum likelihood estimator being the mode estimator. The paper compares the regression results obtained using the proposed approach with those using standard regression procedures. In these cases the paper refers to Kontek (2010).

The paper demonstrates the advantages of the proposed methodology, which eliminates most of the estimation problems resulting from the inhomogeneity of the data. These include the computational problems of the standard quantile and mode regression estimators. As the concept presented here is very general, it can easily be applied to other cases of data inhomogeneity and is not limited to lottery experiments.

Clearly, this method cannot be used when there are few data points for each value of the explanatory variable. One disadvantage of the proposed method is that determining the data densities requires an intermediate step. The accuracy of the density estimation therefore has a vital impact on the final result. Another disadvantage is the additional computation time required. On the other hand, determining the distributions will speed up the regression procedure considerably. This method additionally comes with the nice feature of being able to predict the error distribution *before* the regression procedure, as the data distributions are already known.

1.5. The presented methodology, when applied to lottery experiments, makes it possible to confirm and extend the previously presented conclusion that lottery valuations are only nonlinear with respect to probability when medians and means are considered. Such nonlinearity is not confirmed by the mode (maximum likelihood) estimator. This means that the most likely behavior of a group is fully rational.

To the best of the author's knowledge, this is the first paper to present a true maximum likelihood estimation for lottery experiments where "true" is to be understood as meaning based on the properly defined distribution of the observed variable (the typically assumed normal error distribution reduces the maximum likelihood to a standard least squares procedure).

1.6. An extensive number of estimation methods presented in this and the former paper were made possible by using the relative utility model which, in contrast to the Prospect Theory model, adopts a classical econometric approach to data description. This model is briefly outlined in Point 2. Point 3 presents the data sets examined in the present research. Point 4 demonstrates calculated densities of the relative certainty equivalents for given probabilities, which are then used for mean (Point 5), quantile (Point 6), and mode (Point 7) regression estimation. Point 8 summarizes results of the paper.

2. Relative Utility Model

2.1. The relative utility model assumes a direct relationship between probability and the relative certainty equivalents for a two outcome lottery:

$$p = Q(r), \tag{2.1}$$

where p denotes probability, Q denotes a relative utility function, which should have the form of a cumulative density function defined over the range [0,1], and r denotes the relative certainty equivalent defined as:

$$r = \frac{ce - A}{P - A}, \tag{2.2}$$

where ce denotes the certainty equivalent, $P = \text{Max}(x)$ is the maximum lottery outcome, and $A = Min(x)$ is the minimum lottery outcome. The relationship described by (2.2) ensures that r assumes values from the range [0, 1], even in the case of lotteries with a risk free component.

2.2. As probability p is a single-variable function of the relative certainty equivalent r, r can be easily represented as a function of p:

$$r = Q^{-1}(p), \tag{2.3}$$

where Q^{-1} is the inverse form of the relative utility function. Because there are certainty equivalents which are typically determined in experiments rather than probabilities, the inverse form (2.3) of the relative utility function will be mostly used throughout the paper.

2.3. It is possible to propose several functional forms for the relative utility function Q. Beta distribution is the only one used in this paper, as it is the best known and most widely used distribution defined over the interval [0,1]. Hence, the function Q is described using Cumulative Beta Distribution as follows:

$$p = Q(r) = I(r; \alpha, \beta), \tag{2.4}$$

where I denotes the regularized incomplete beta function. The curve is S-shaped for $\alpha > 1$ and $\beta > 1$, inversed S-shaped for $0 < \alpha < 1$ and $0 < \beta < 1$, J-shaped for $\alpha > 1$ and $0 < \beta < 1$, and inverse J-shaped for $0 < \alpha < 1$ and $\beta > 1$. For $\alpha = 1$ and $\beta = 1$ the curve is linear. The inversed form of (2.4) is:

$$r = Q^{-1}(p) = I^{-1}(p; \alpha, \beta), \tag{2.5}$$

where I^{-1} denotes the inverse of the regularized incomplete beta function. More on the relative utility function, especially in the case of multi-prize lotteries can be found in Kontek (2009).

3. Data Sets

3.1. Two data sets are considered in the present study.

Set 1 - the experimental data presented by Traub and Schmidt (2009), whose research concerned the relationship between WTP (Willingness to Pay) and WTA (Willingness to Accept). Twenty four subjects participated in the experiment. Only that subset of the data covering the certainty equivalents of two outcome lotteries was used in further analyses.

Set 2 - the experimental data of Idzikowska (2009), whose research concerns the question of whether the form in which probability is presented has any impact on the shape of the probability weighting function. Twenty five subjects participated in the experiment but some of the responses were disregarded by Idzikowska on account of their inconsistency. The present research uses that subset of the data related to experimentally learned probabilities.

3.2. It is noted that the amount of collected data differs for specific probabilities. This is shown in Tables 3.1 and 3.2.

Probability	.15	.17	.20	.25	.30	.40	.50	.60	.65	.80	.85	.90	Total
No of data	24	24	48	48	72	24	144	48	24	96	24	24	600

Table 3.1. Number of data for Set 1.

Probability	.01	.05	.10	.25	.50	.75	.90	.95	.99	Total
No of data	54	53	57	55	116	61	54	58	60	568

Table 3.2. Number of data for Set 2.

The greatest number of data in both sets exists for the probability of 0.5. Additionally, the data count is pretty high in Set 1 for probabilities 0.8 and 0.3. This means that these data can dominate the data concerning the remaining probabilities during the estimation procedure, which can have undesirable effects. A similar example of inhomogeneity may be found in Tversky and Kahneman's data (1992), which is presented in Table 3.3.

Probability	.01	.05	.10	.25	.50	.75	.90	.95	.99
No of lotteries	2	3	3	3	6	3	3	3	2

Table 3.3. Number of lotteries having the given probability of winning the main prize.

It should be obvious that determining the densities of the relative certainty equivalents for given probabilities solves the problem of unequal amounts of data. There is one density function for each probability, no matter how many data are used to determine it.

4. Densities of Relative Certainty Equivalents *r*

4.1. In order to estimate a density function, its functional form must first be specified. There are a few possibilities, although not as many as in the case of unbounded distributions. As in the case of the relative utility function, we will concentrate further on beta distribution and its generalization[3].

4.1.1. Beta distribution (*bt*) is defined as:

$$bt(r;\gamma,\delta)=\frac{r^{\gamma-1}\left(1-r\right)^{\delta-1}}{B\left(\gamma,\delta\right)}. \tag{4.1}$$

4.1.2. There are several generalizations of beta distribution (Gupta, Nadarajah, 2004) with 3, 4 or even 5 parameters. It should be noted that the number of parameters here is not a problem and need not be limited. This is because determining the density function is only an intermediate result where high accuracy is of great importance. The most adequate seems to be the generalization *gbt* of Libby and Novick (1982):

$$gbt(r;\gamma,\delta,\lambda)=\frac{r^{\gamma-1}\left(1-r\right)^{\delta-1}}{B\left(\gamma,\delta\right)}\frac{\lambda^{\gamma}}{\left[\left(\lambda-1\right)r+1\right]^{\gamma+\delta}}. \tag{4.2}$$

4.2. Knowing the density function allows the characteristic points of the distribution to be determined analytically. The mean value of the *gbt* is[4]:

$$Mean\left[gbt\left(r;\gamma,\delta,\lambda\right)\right]=\frac{\gamma}{\gamma+\delta}\varphi=m, \tag{4.3}$$

where

$$\varphi={}_2F_1\left(1,\delta;1+\gamma+\delta;1-\lambda\right), \tag{4.4}$$

denotes the hypergeometric function. For $\lambda = 1$, (4.3) reduces to the known formula for the mean value of the beta distribution:

$$Mean\left[bt\left(r;\gamma,\delta\right)\right]=\frac{\gamma}{\gamma+\delta}. \tag{4.5}$$

The *gbt* mode (and anti-mode) is given by:

---

[3] γ, δ, and λ are used throughout the paper to denote the parameters of the density function, whereas α and β are used to describe the relative utility function. Distribution written with small letters like *bt*, *gbt* or *d* denote density functions, whereas written with capital letters like *GBT* or *D* denote cumulative density functions.

[4] This and some other *gbt* properties given in this subsection are not to be found in the handbook by Gupta and Nadarajah (2004). There is also a mistake in the formula for the mode value (p.121).

$$Mode\left[gbt\left(x;\gamma,\delta,\lambda\right)\right]=\frac{\gamma+\lambda+\delta\lambda-3\mp\sqrt{\left(\gamma+\lambda+\delta\lambda-3\right)^2-8\left(\gamma-1\right)\left(\lambda-1\right)}}{4\left(\lambda-1\right)}. \quad (4.6)$$

For $\lambda = 1$, (4.6) reduces to:

$$Mode\left[bt\left(r;\gamma,\delta\right)\right]=\frac{\gamma-1}{\gamma+\delta-2}, \quad (4.7)$$

which is the beta distribution mode. The cumulative generalized beta distribution is defined as:

$$GBT\left(r;\gamma,\delta,\lambda\right)=I\left(\frac{\lambda r}{\left(\lambda-1\right)r+1};\gamma,\delta\right), \quad (4.8)$$

where I denotes the regularized incomplete beta function. For $\lambda = 1$, (4.8) simplifies to the cumulative beta distribution:

$$BT\left(r;\gamma,\delta\right)=I\left(r;\gamma,\delta\right). \quad (4.9)$$

After inverting (4.8), the *gbt* quantile can be determined from:

$$Quantile\left[q,gbt\left(r;\gamma,\delta,\lambda\right)\right]=\frac{1}{1-\lambda+\frac{\lambda}{I^{-1}\left(q;\gamma,\delta\right)}}, \quad (4.10)$$

which, for $\lambda = 1$, expresses the beta distribution quantile:

$$Quantile\left[q,bt\left(r;\gamma,\delta\right)\right]=I^{-1}\left(q;\gamma,\delta\right). \quad (4.11)$$

In the special case where $q = 0.5$, the *gbt* median is given by:

$$Median\left[gbt\left(r;\gamma,\delta,\lambda\right)\right]=\frac{1}{1-\lambda+\frac{\lambda}{I^{-1}\left(0.5;\gamma,\delta\right)}}. \quad (4.12)$$

The *gbt* variance is given by:

$$\begin{aligned} Variance\left[gbt\left(r;\gamma,\delta,\lambda\right)\right] &= \frac{\gamma\left\{\left(\gamma+\delta\right)^2-\left(\gamma+\delta\right)\left[1+\gamma+\left(\delta-1\right)\lambda\right]\phi+\gamma\left(1-\lambda\right)\varphi^2\right\}}{\left(\gamma+\delta\right)^2\left(\lambda-1\right)} \\ &= \frac{\left(1-m\right)\left(\gamma-m\right)+m\left(1-m-\delta\right)\lambda}{\lambda-1}=var, \end{aligned} \quad (4.13)$$

where m is given by (4.3). For $\lambda = 1$ this reduces to the known formula for beta distribution:

$$Variance\left[bt\left(r;\gamma,\delta\right)\right]=\frac{\gamma\,\delta}{\left(\gamma+\delta\right)^{2}\left(\gamma+\delta+1\right)}. \tag{4.14}$$

4.3. Estimating densities is commonly achieved using the maximum likelihood estimator. The likelihood principle involves choosing the parameter vector θ that maximizes the likelihood of observing the actual sample. The likelihood function is expressed by $L_N(\theta|y,x)$ and is a function of θ given the data (y, x). Maximizing $L_N(\theta)$ is equivalent to maximizing the (conditional) log-likelihood function:

$$\ln L_N\left(\theta\right)=\sum_{i=1}^{N}\ln d\left(y_i\mid x_i,\theta\right), \tag{4.15}$$

where d denotes a (conditional) density function. The log-likelihood function for *gbt* is:

$$\begin{aligned}\ln L_N\left(\gamma,\delta,\lambda\right)=&\left(\gamma-1\right)\sum_{i=1}^{N}\ln r_i+\left(\delta-1\right)\sum_{i=1}^{N}\ln\left(1-r_i\right)-N\ln B\left(\gamma,\delta\right)\\&+\gamma\,N\ln\lambda-\left(\gamma+\delta\right)\sum_{i=1}^{N}\ln\left[\left(\lambda-1\right)r_i+1\right],\end{aligned} \tag{4.16}$$

and for *bt* reduces to:

$$\ln L_N\left(\gamma,\delta\right)=\left(\gamma-1\right)\sum_{i=1}^{N}\ln r_i+\left(\delta-1\right)\sum_{i=1}^{N}\ln\left(1-r_i\right)-N\ln B\left(\gamma,\delta\right). \tag{4.17}$$

4.4. A problem was encountered during the estimation procedure in that r sometimes assumed a value of 0 or 1 in Set 1. This resulted from providing certainty equivalents equal to the minimum or maximum outcomes of the lottery (for instance certainty equivalents of \$0 or \$40 for a \$0-40 lottery with a 0.5 probability of winning). These seemingly illogical answers were possible because of the way in which the experiment was set up, i.e. responses were not checked for inconsistencies. The values on the bounds of the density function make maximum likelihood estimation impossible as one of the first two elements in (4.16) or (4.17) is infinite. A similar case was encountered in Set 2. Although values 0 and 1 were prohibited there, a number of minimum (1/200) and maximum (199/200) values, which were close to the bounds distorted the shape of the estimated density curve. It was therefore decided to exclude all values in the vicinity of the bounds from the estimation procedure, but to nevertheless consider them in subsequent analyses. The technical details are given in Appendix 1, as this problem does not appear to greatly improve our knowledge on the main subject of this paper.

4.5. The results of the maximum likelihood estimator using g*bt* and *bt* are presented in Figures 4.1 and 4.2. Figure 4.1 presents the densities of the relative certainty equivalents for

the specific probabilities in Set 1.

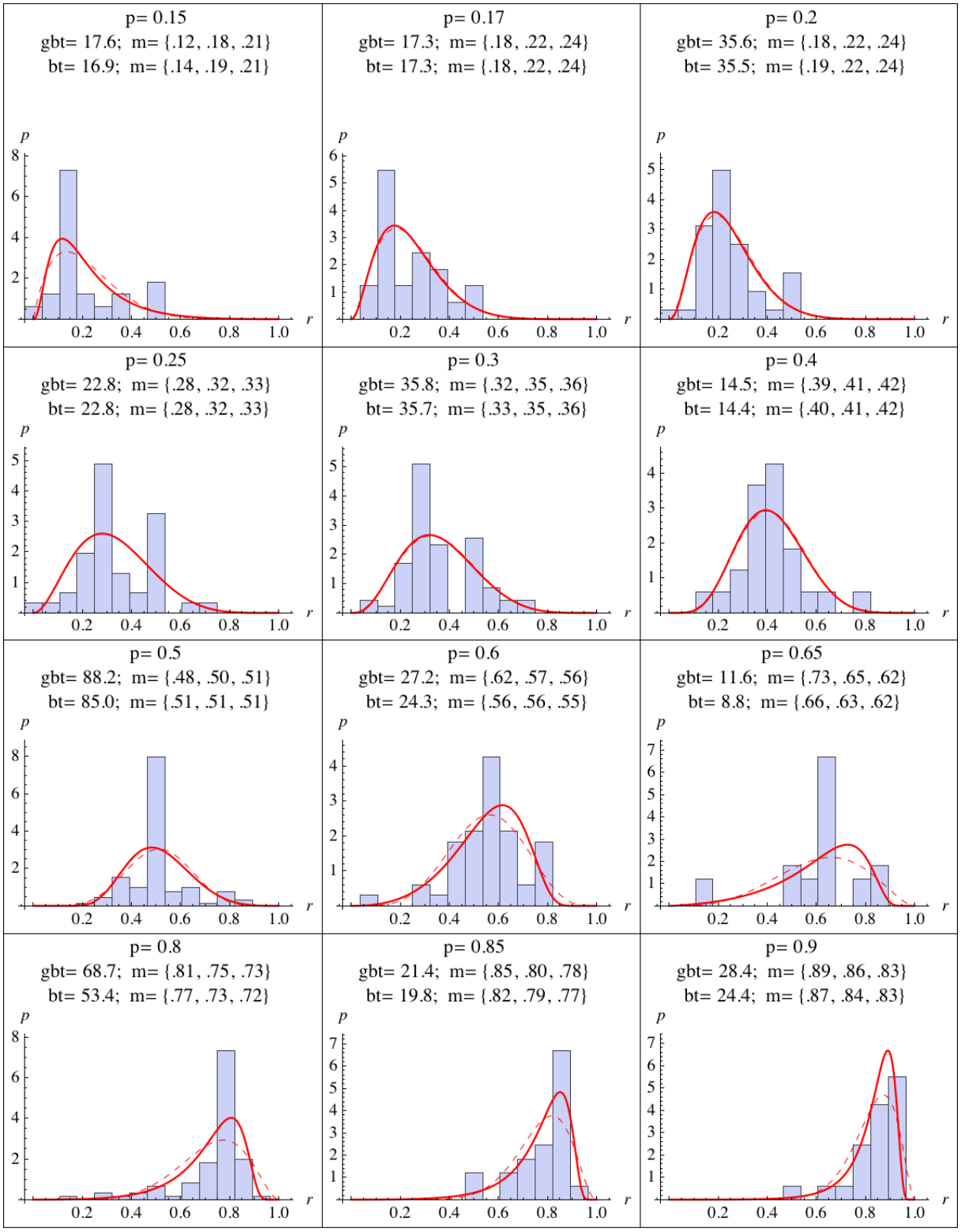

Figure 4.1. Densities of *r* for respective probabilities in Set 1 for *gbt* (red) and *bt* (red, dashed). The values of *gbt* and *bt* indicate the maximized log-likelihood value for the respective density function. The parentheses contain the characteristic points *m* of the distributions: the mode, median and mean calculated using the formulas presented in 4.2.

The *gbt* performs slightly better than *bt* - especially for probabilities approaching 0

and 1. This is indicated by the maximized log-likelihood values. The better estimation results for *gbt* are not surprising because it has one more parameter than *bt*. It should be noted, however, that the estimated density functions do not fit the histograms exactly, even when *gbt* is used. The histograms show that the data are characterized by high peakedness, which confirms the high kurtosis values calculated using standard distribution moment measures. This may mean that *gbt* and *bt* are not the best functions to estimate such peaked distributions.

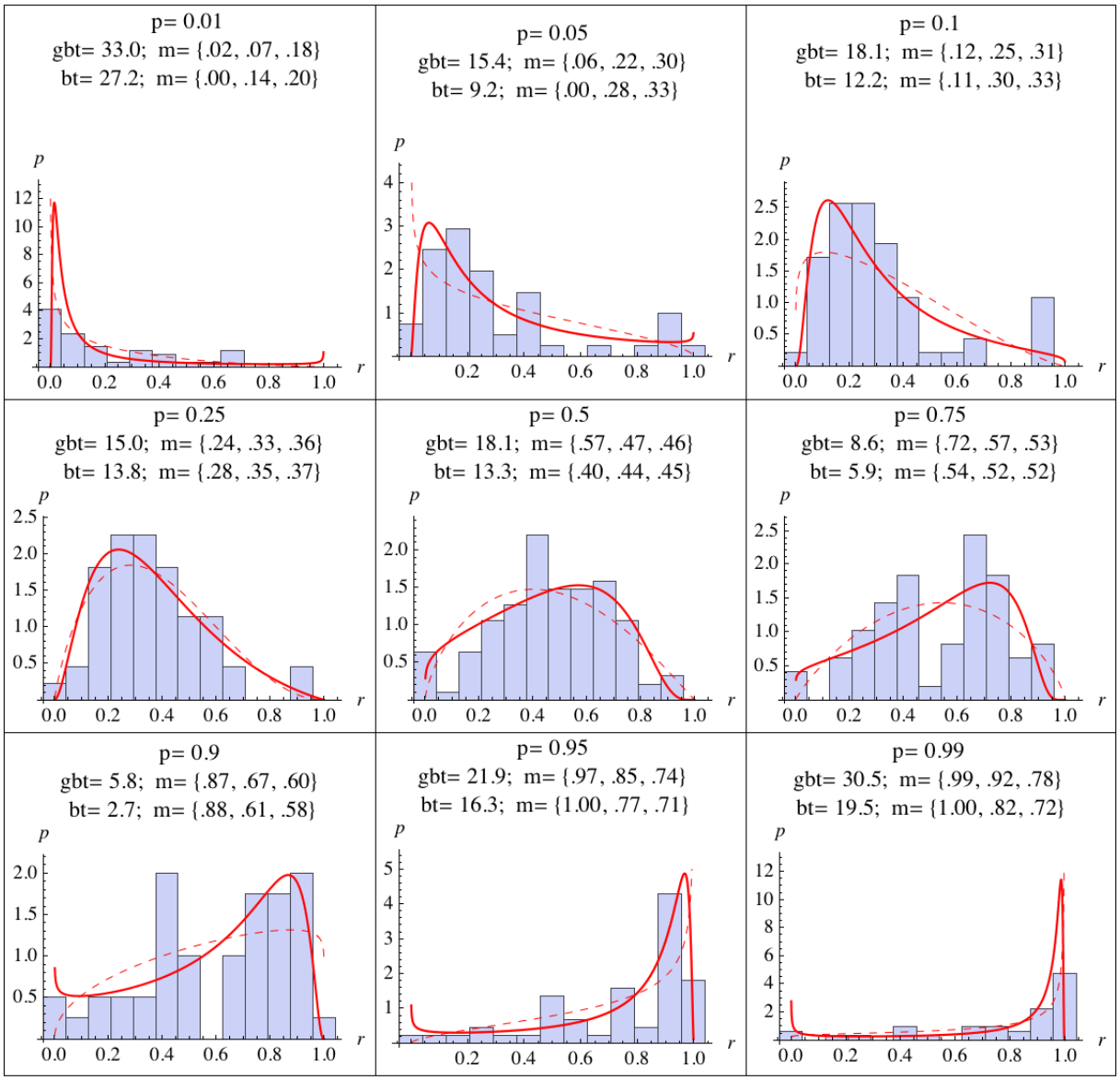

Figure 4.2. Densities of *r* for Set 2. The description as in Figure 4.1.

In any case, it can be observed that the distributions are positively skewed for low probabilities, negatively skewed for high probabilities, and roughly symmetric for medium probabilities. This is confirmed by the characteristic points of the distributions. For lower probabilities, the mode is less than the median, which is less than the mean. For higher probabilities, the reverse holds true. For medium probabilities the mode, median and mean assume

similar values. It is important to note that the mode values for both *gbt* an *bt* estimations are very similar to the values of probabilities. This means that the most likely certainty equivalents correspond with the expected values of lotteries.

Figure 4.2 presents the densities of the relative certainty equivalents *r* in Set 2. Here, the data are more scattered and the histograms are flatter. This confirms the results obtained using standard distribution moments. In this case, *gbt* performs much better than *bt*. For probabilities 0.01, 0.05, 0.95, and 0.99, *bt* was unable to detect the peak, and has an inverse J-shape for the first two probabilities and a J-shape for the remaining two. Peaks are, however, to be expected as the density at the bounds should have a value of 0. The maximized log-likelihood values for *bt* confirm that these estimations are much worse than those with peaks obtained using *gbt.*

The presented estimations confirm the conclusions drawn for Set 1. The distributions are positively skewed for low probabilities and negatively skewed for high probabilities. The values of modes for *gbt* are almost identical with the values of probabilities.

4.6. It is interesting to compare all the density functions obtained using *gbt* on a single graph (Figure 4.3). Quite clearly, the densities gradually change its character from being positively to negatively skewed. The correspondence of the probability values with the values of corresponding modes is readily observed.

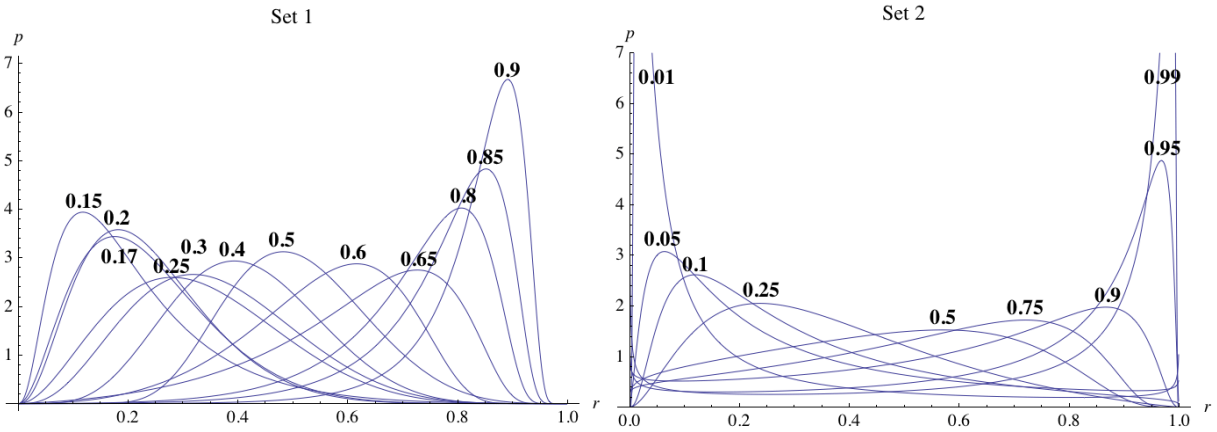

Figure 4.3. Densities obtained using *gbt* for Sets 1 and 2 plotted together. The numbers indicate the respective probabilities for which the estimations were made.

4.7. Despite providing worse estimation results, *bt* has some computational advantages over *gbt*. First, it requires much less time to find the optimum. Second, it is more robust - especially for probabilities approaching 0 and 1. Finally, it does not require performing high precision calculations, as is the case with *gbt*. It may therefore be advantageous to use *bt* first so as to have a view of the sought densities. In some cases, as for Set 1, the *bt* densities may be sufficient to get a good approximation of the final results.

5. Mean Regression Based on Densities of *r*

5.1. The means of density functions can easily be calculated using (4.3) or (4.5). These values have already been presented in Figures 4.1 and 4.2. According to the methodology proposed in 1.4, the mean regression estimator should minimize the distances of the estimated relative utility curve from those mean values. Using a square error loss function results in:

$$S^d_{mean}(\theta) = \sum_{j=1}^{M} \left(m_j - r_j \right)^2, \tag{5.1}$$

where superscript *d* denotes an approach based on densities, $j = 1 \ldots M$ denotes consecutive probabilities p_j, m_j denotes the mean of the density distribution for probability p_j, and r_j is the value of the relative utility function for probability p_j. S^d_{mean} is a function of the relative utility parameters θ because $r_j = Q^{-1}\left(p_j ; \theta \right)$ (cf. (2.3)). It is natural to express the squares of the errors in relative terms to minimize the impact of the differences in variances for consecutive probabilities:

$$S^d_{mean}(\theta) = \sum_{j=1}^{M} \frac{\left(m_j - r_j \right)^2}{var_j}, \tag{5.2}$$

where var_j is the variance of the density distribution for probability p_j. This is an extremely easy method of eliminating the problem of heteroskedasticity, as the variance can be determined analytically from the density function. Substituting *var* with (4.13) and *r* with (2.5) results, in the case of *gbt*, in:

$$S^d_{mean}(\alpha, \beta) = \sum_{j=1}^{M} \frac{\left[m_j - I^{-1}\left(p_j ; \alpha, \beta \right) \right]^2 \left(\lambda_j - 1 \right)}{\left(1 - m_j \right)\left(\gamma_j - m_j \right) + m_j \left(1 - m_j - \delta_j \right) \lambda_j}, \tag{5.3}$$

where *m* is given by (4.3) and γ_j , δ_j , λ_j denote the density function parameters for probability p_j . Similarly the result for *bt* is:

$$S^d_{mean}(\alpha, \beta) = \sum_{j=1}^{M} \frac{\left(\gamma_j + \delta_j + 1 \right)\left[\left(\gamma_j + \delta_j \right) I^{-1}\left(p_j ; \alpha, \beta \right) - \gamma_j \right]^2}{\gamma_j \, \delta_j}. \tag{5.4}$$

5.2. The value of S^d_{mean} is expressed in terms of normalized variances. The fit error is therefore given by:

$$err^d_{mean} = \sqrt{\frac{Min\left(S^d_{mean} \right)}{M - k}}, \tag{5.5}$$

where M - k specifies the number of degrees of freedom resulting from M density functions and k parameters of the relative utility function Q. The error (5.5) is expressed in terms of the normalized distance which is the square root of the variance.

5.3. Figure 5.1 presents the mean regression estimations obtained using densities of r. The shapes of the relative utility curve for Set 1 and Set 2 are almost the same as those obtained using the standard least squares regression procedure. The results presented in boxes show that the average error is equal to 12 percent of the square root of the variance for Set 1 and 15 percent for Set 2.

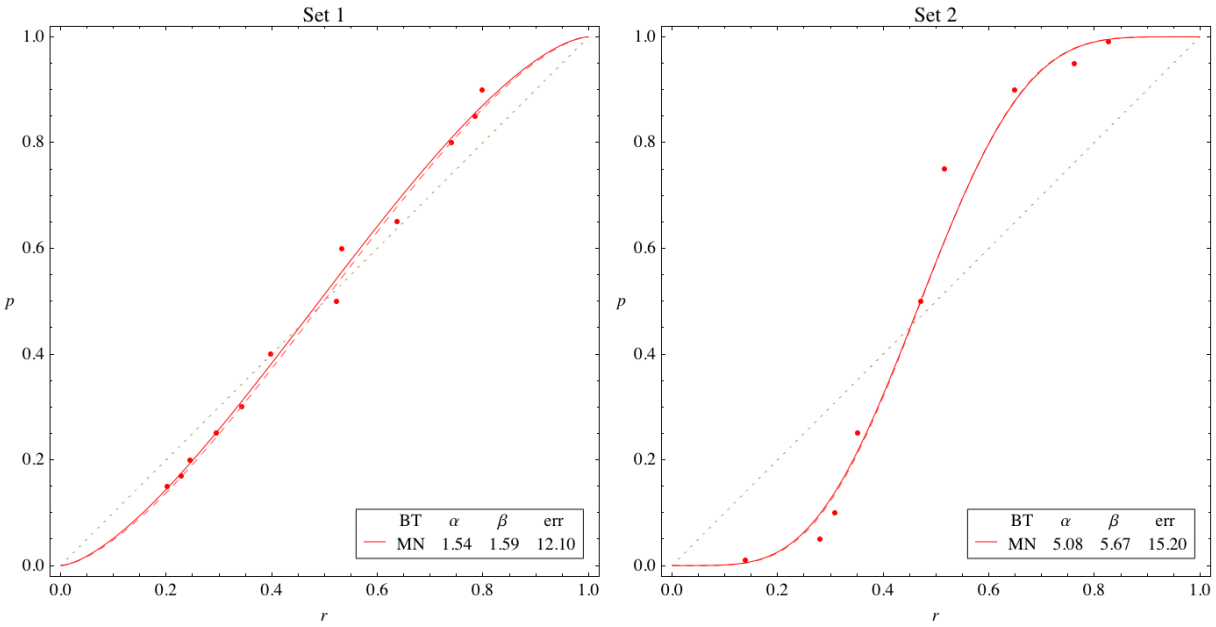

Figure 5.1. Mean regression estimations based on densities of r. The *gbt* densities were used for optimization. The dashed curves are those obtained using the standard least squares regression method.

6. Quantile Regression Based on Densities of *r*

6.1. The density quantile for a given value of r can easily be calculated using the cumulative density distributions (4.8) or (4.9). According to the methodology presented, the qth quantile regression estimator should therefore result in a curve which is as close as possible to the assumed qth quantile of the density functions. One possible way of using the absolute error loss function results in:

$$S_q^d(\theta) = \sum_{j=1}^{M} \left| q - D\left(r_j; \theta_j^d\right) \right|, \tag{6.1}$$

where q denotes the quantile under consideration, D denotes the cumulative density distribution having parameters θ_j^d for probability p_j, and which value expresses the quantile at r_j. S_q^d is a function of the relative utility parameters θ because $r_j = Q^{-1}(p_j; \theta)$. Equation (6.1) is

clearer than the standard formula for quantile regression (Cameron, Trivedi, 2005, Kontek, 2010), because it directly shows the interest in the minimization of the distance expressed in quantiles rather than in absolute values of *r*. However, the loss function is not smooth in this case and this estimator shares the adverse feature of the standard regression estimator in trying to force the estimated curve to pass through any two quantile points (for a two-parameter *Q*) exactly. This runs the risk of failing to find a global minimum. For this reason, it is more practical to use the squares of quantile distances:

$$S_q^d(\theta) = \sum_{j=1}^{M} \left[q - D\left(r_j; \theta_j^d\right) \right]^2. \tag{6.2}$$

It is important to note that, despite using the least squares procedure, this remains the quantile regression estimator as it minimizes the quantile distances. Considering the functional form of the density functions gives in case of *gbt* :

$$S_q^d(\alpha,\beta) = \sum_{j=1}^{M} \left[q - GBT\left(I^{-1}\left(p_j;\alpha,\beta\right);\gamma_j,\delta_j,\lambda_j\right) \right]^2, \tag{6.3}$$

and in case of *bt*:

$$S_q^d(\alpha,\beta) = \sum_{j=1}^{M} \left[q - BT\left(I^{-1}\left(p_j;\alpha,\beta\right);\gamma_j,\delta_j\right) \right]^2. \tag{6.4}$$

6.2. The minimum value of S_q^d obtained by the minimization procedure is the square of the quantile distances. The fit error is therefore given by:

$$err_q^d = \sqrt{\frac{Min\left(S_q^d\right)}{M-k}}, \tag{6.5}$$

and expresses the average quantile distance of the relative utility function from the quantile considered.

6.3. The quantile regression estimations using densities of *r* are shown in Figure 6.1. A few observations are in order. The most important is that the median regression for Set 1 is no longer a straight line as for standard quantile regression, but a slightly S-shaped curve. This is confirmed by the beta distribution parameters, which are greater than 1 (α = 1.3, β = 1.32). The range between the lower and upper quantiles is wider than for the standard quantile regression. This may partially be the result of *gbt* having imperfectly estimated the peaked densities.

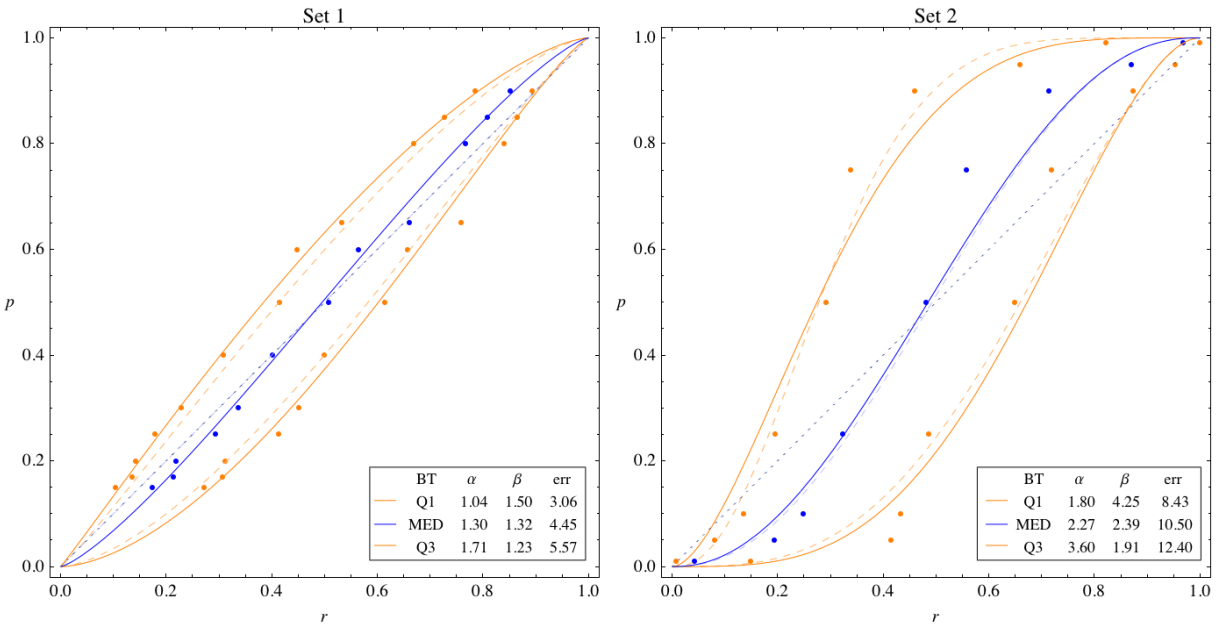

Figure 6.1. Quantile regressions based on the densities of *r*. The *gbt* densities were used for optimization. The dashed curves are those obtained using the standard quantile regression estimators.

The results for Set 2 also differ slightly from those obtained using standard regression estimator. Examining the plots for Set 2 gives the impression that the fit is not the best. The possibility that the inexactness in the estimation is caused by the functional form of the relative utility function Q cannot be excluded. This impression, however, also arises from being accustomed to minimization of the distance expressed in absolute terms of r whereas the quantile distance does not correspond with this value. The results presented in the boxes show that the average error is between 3.06 and 5.57 quantiles for Set 1 and between 8.43 and 12.40 quantiles for Set 2, depending on whether lower, median or upper quartile is considered.

7. Mode (Maximum Likelihood) Regression Based on Densities of *r*

7.1. The modes of the density functions can be easily calculated using (4.6) or (4.7), and have already been presented in Figures 4.1 and 4.2. The mode regression estimator should therefore result in r values of the relative utility function which are as close as possible to the modes of the distributions. One possibility is to proceed similarly as with mean regression and to use (5.2) with the mean values replaced with the modes. This would result in minimizing the distances of the estimated curve from the modes. However, a more interesting possibility is the ratio of the density at the point assumed by the relative utility function to the density at the mode (where it achieves its maximum). This would produce a function which should be maximized:

$$S^d_{mode}(\theta) = \prod_{j=1}^{M} \frac{d\left(r_j; \theta^d_j\right)}{d\left(r_{\text{mode},j}; \theta^d_j\right)}, \tag{7.1}$$

where $j = 1 \ldots M$ denotes consecutive probabilities p_j, d denotes the density function having parameters θ^d_j for probability p_j, r_j is the value of the relative utility function for this probability, and $r_{mode,j}$ denotes the value at the mode for density j. S^d_{mode} is a function of the relative utility parameters θ because $r_j = Q^{-1}(p_j; \theta)$. As the denominator of (7.1) has no impact on the maximization procedure, (7.1) may be presented as:

$$S^d_{mode}(\theta) = \prod_{j=1}^{M} d\left(r_j; \theta^d_j\right), \tag{7.2}$$

where the right-hand side may be recognized as the joint probability of observing r_j values given parameters θ. On the other hand, it can also be seen as a function of θ given the densities d for the respective probabilities. In this way, (7.2) can be recognized as the likelihood of observing the actual densities of r. Maximizing (7.2) with respect to θ therefore leads to an estimator of the relative utility function which maximizes the likelihood of observing the stated densities of r. The mode regression estimator thus appeared to be the maximum likelihood estimator, which is one of the most interesting results of this paper.

7.2. The logarithm of the likelihood function (7.2) yields:

$$\ln L^d(\theta) = \sum_{j=1}^{M} \ln d\left(r_j; \theta^d_j\right), \tag{7.3}$$

where S^d_{mode} was replaced by L^d. As the density function is described by *gbt* and the relative utility function Q by cumulative beta distribution, we obtain:

$$\begin{aligned} \ln L^d(\alpha, \beta) &= \sum_{j=1}^{M} \ln gbt\left(I^{-1}\left(p_j; \alpha, \beta\right); \gamma_j, \delta_j, \lambda_j\right) \\ &= \sum_{j=1}^{M} \left(\gamma_j - 1\right) \ln I^{-1}\left(p_j; \alpha, \beta\right) + \sum_{j=1}^{M} \left(\delta_j - 1\right) \ln\left(1 - I^{-1}\left(p_j; \alpha, \beta\right)\right) - M \ln B\left(\gamma_j, \delta_j\right) \\ &\quad - \sum_{i=1}^{N} \left(\gamma_j + \delta_j\right) \ln\left[\left(\lambda_j - 1\right) I^{-1}\left(p_j; \alpha, \beta\right) + 1\right] + \gamma_j M \ln \lambda_j . \end{aligned} \tag{7.4}$$

7.3. Let L_o denote the maximum value of the likelihood function which can be achieved as a result of the estimation. The mode can be calculated from (4.6) or from (4.7), so the density at the mode can also be determined analytically. Hence:

$$\log L_0 = \sum_{j=1}^{M} \ln d(r = r_{\mathrm{mode},j}; \theta_j^d). \tag{7.5}$$

The Goodness of Fit measure χ^2, which resembles other pseudo-R^2 measures, can be given:

$$\chi^2_{ml1} = \frac{\log L_e}{\log L_0}, \tag{7.6}$$

where L_e is the maximized likelihood function. The meaning of the measure when presented in this way, however, is not clear. Bearing in mind that the likelihood function determines probability, the (geometric) mean ratio of obtained to possible probability (cf. (7.1)) can be written as:

$$\chi^2_{ml2} = \left[\prod_{j=1}^{M} \frac{d\left(r_j; \theta_j^d\right)}{d\left(r_{\mathrm{mode},j}; \theta_j^d\right)} \right]^{\frac{1}{M}} = \left(\frac{L_e}{L_0} \right)^{\frac{1}{M}} = \frac{e^{\frac{\log L_e}{M}}}{e^{\frac{\log L_0}{M}}} = e^{\frac{\log L_e - \log L_0}{M}}. \tag{7.7}$$

A penalty for the number of parameters of the relative utility function can be introduced:

$$\chi^2_{ml3} = e^{\frac{\log L_e - \log L_0}{M-k}}. \tag{7.8}$$

This measure can be expressed as an error:

$$err^d_{ml} = 1 - \chi^2_{ml3}. \tag{7.9}$$

7.4. The results of the estimation are presented in Figure 7.1 along with the values of the distribution modes.

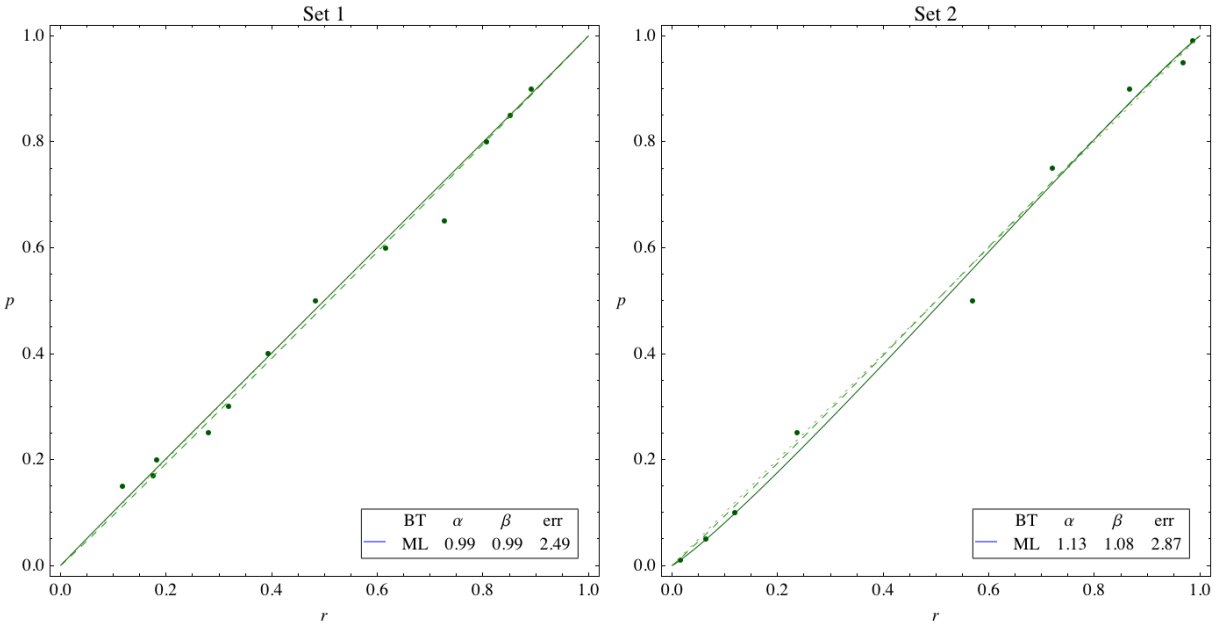

Figure 7.1. Mode regressions based on densities of *r*. The *gbt* densities were used for optimization. The dashed curves are those determined using the standard mode regression estimator.

As can be seen, the obtained function is almost linear for both sets of data. These results confirm the earlier observation achieved using the nonparametric and parametric approaches and are of great importance. It transpires that the most likely value of the relative certainty equivalent equals the probability of winning the lottery.

8. Summary

8.1. Figure 8.1 shows the joint estimation results obtained in Points 5, 6, and 7. This figure can be compared to the results of the nonparametric and parametric regression estimations (Kontek, 2010). In Set 1, the mode regression is almost linear while the median and mean exhibit some curvature. The graphs for Set 2 are similar. This means that the most likely lottery valuation is close to its expected value in both cases. Another way of saying this is that the most likely behavior of the examined groups was fully rational.

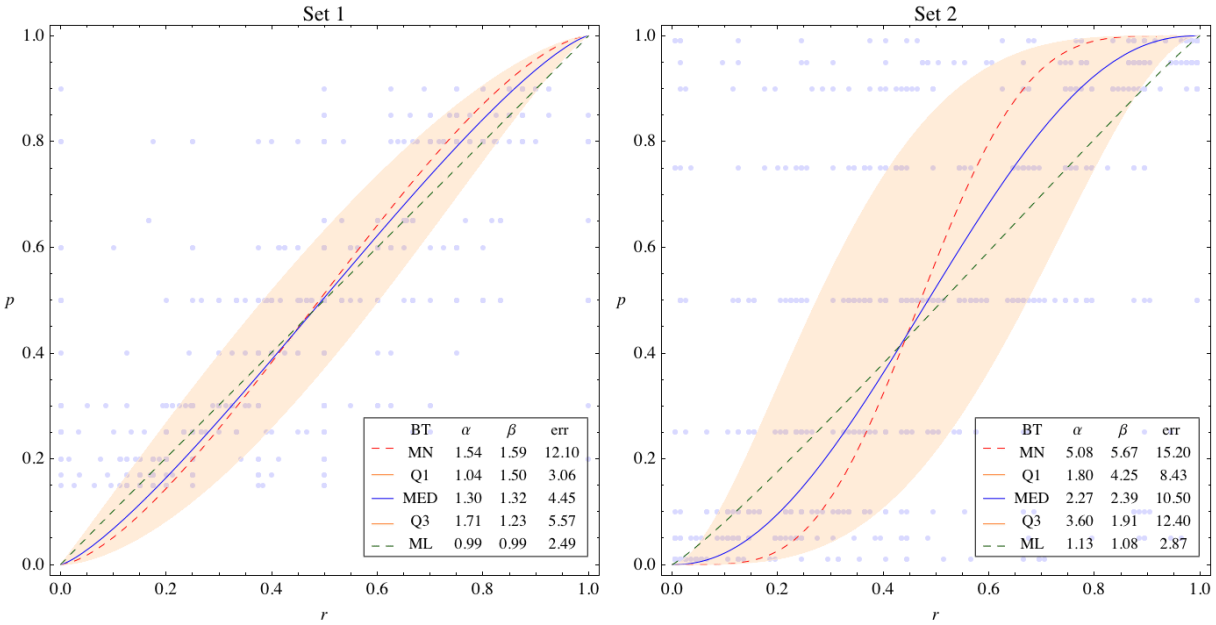

Figure 8.1. Mean, quantile, and mode regression estimations plotted together.

8.2. This paper concentrated on presenting a regression methodology for inhomogeneous data encountered in lottery experiments. The paper covered mean, quantile (including median), and mode (maximum likelihood) regression estimators based on densities of observed variable. The proposed methodology would appear to have several advantages over the standard procedures. First and foremost, it enables the estimation problems caused by the inhomogeneity of the data (like heteroskedascicity) to be easily eliminated. Second, the computational inconveniences of the median and (especially mode) estimators can be eliminated. Third, it represents errors in a meaningful way (e.g. in terms of quantiles in the quantile regression). Finally, the computational time may be very short once the densities are deter-

mined.

It is not possible to cover all the details and the other subjects related to those already mentioned in one paper. Some of the subjects touched on require a deeper analysis. These include other functional forms of density functions especially suited to peaked densities, other functional forms of the relative utility function, assessing data at the individual level, and estimating the relative utility function for multi-prize lotteries. The results presented in this paper should obviously be confirmed for other data sets as well. All this is left for future papers.

8.3. This paper nevertheless proves that inhomogeneity of data has to be taken into account during analysis and evidences the usefulness of the proposed estimation methods. The proposed methodology seems to have much wider application in econometric research than just lottery experiments.

Appendix 1:

As stated in 4.6, there appeared to be cases in Set 1 where relative certainty equivalents assumed values of either 0 or 1. This made it impossible to perform the maximum likelihood procedure. In Set 2, relative certainty equivalents which were close to the bounds of the density function domain distorted the shape of the estimated density curve. These values were disregarded during the estimation procedure so as to eliminate this problem, but were then used in further analyses.

If, in a subset of data related to a specific probability, there are k_1 data items with a value of 0, k_2 data items with values in the range (0,1), and k_3 data items with a value of 1, then the densities may be defined as follows:

$$\rho_1 = \frac{k_1}{K}, \rho_2 = \frac{k_2}{K}, \text{and } \rho_3 = \frac{k_3}{K}, \text{where } k_1 + k_2 + k_3 = K. \tag{A1.1}$$

The new properties of *bt* may be given:

$$m = Mean\left[\rho_2, \rho_3, bt\left(r;\gamma,\delta\right)\right] = \frac{\gamma\,\rho_2}{\gamma+\delta} + \rho_3, \tag{A1.2}$$

$$var = Variance\left[\rho_2, \rho_3, bt\left(r;\gamma,\delta\right)\right] = \frac{m\left(1+\gamma\right) - m^2\left(1+\gamma+\delta\right) + \delta\,\rho_3}{1+\gamma+\delta}, \tag{A1.3}$$

$$loss = \frac{\left(m-r\right)^2\left(1+\gamma+\delta\right)}{m\left(1+\gamma\right) - m^2\left(1+\gamma+\delta\right) + \delta\,\rho_3}, \tag{A1.4}$$

which respectively replace (4.5), (4.14) and the loss function inside (5.4). Similarly, the new

gbt properties are as follows:

$$m = Mean\left[\rho_2, \rho_3, gbt\left(r;\gamma,\delta,\lambda\right)\right] = \frac{\gamma\,\rho_2\,\varphi}{\gamma+\delta} + \rho_3, \tag{A1.5}$$

$$\begin{aligned} var &= Variance\left[\rho_2, \rho_3, bt\left(r;\gamma,\delta\right)\right] \\ &= \frac{m^2\left(1-\lambda\right) - m\left[1+\gamma+\left(\delta-1\right)\lambda\right] + \delta\,\lambda\,\rho_3 + \gamma\left(\rho_2+\rho_3\right)}{\lambda-1}, \end{aligned} \tag{A1.6}$$

$$loss = \frac{\left(m-r\right)^2\left(\lambda-1\right)}{m^2\left(1-\lambda\right) - m\left[1+\gamma+\left(\delta-1\right)\lambda\right] + \delta\,\lambda\,\rho_3 + \gamma\left(\rho_2+\rho_3\right)}, \tag{A1.7}$$

which respectively replace (4.3), (4.13) and the loss function inside (5.3).

In the case of the densities based quantile estimator, the quantile q_r is calculated as:

$$q_r = \begin{cases} 0 & if\ q \le \rho_1, \\ \rho_1 + \rho_2\, D\left(r_j;\theta_j^d\right) & if\ \rho_1 < q < \rho_1+\rho_2, \\ 1 & if\ q \ge \rho_1+\rho_2. \end{cases} \tag{A1.8}$$

This value is then used in (6.3).

References:

1. Cameron, A. C., Trivedi, P. K., (2005). *Microeconometrics. Methods and Applications*, Cambridge University Press.

2. Gupta, A. K., Nadarajah, S., (2004), *Handbook of Beta Distribution and Its Applications*, Marcel Dekker, Inc, New York.

3. Idzikowska, K., (2009). *Determinants of the probability weighting function*, presented under name Katarzyna Domurat at SPUSM22 Conference, Rovereto, Italy, August 2009, paper Id = 182, http://discof.unitn.it/spudm22/program_24.jsp .

4. Kontek, K., (2010). *Mean, median or Mode? A Striking Conclusion from Lottery Experiments*, MPRA Paper http://mpra.ub.uni-muenchen.de/21758/, Available at SSRN: http://ssrn.com/abstract=1581436

5. Kontek, K., (2009). *Lottery Valuation Using the Aspiration / Relative Utility Function,* Warsaw School of Economics, Department of Applied Econometrics Working Paper no. 39. Available at SSRN: http://ssrn.com/abstract=1437420 and RePec:wse:wpaper:39.

6. Libby, D. L., Novick, M. R., (1982). *Multivariate generalized beta distributions with applications to utility assessment*, Journal of Educational Statistics, 7, pp 271-294.

7. Traub, S., Schmidt, U., (2009). *An Experimental Investigation of the Disparity between WTA and WTP for Lotteries,* Theory & Decision 66 (2009), pp 229-262.

8. Tversky A., Kahneman D., (1992). *Advances in Prospect Theory: Cumulative Representation of Uncertainty,* Journal of Risk and Uncertainty, vol. 5(4), October, pp 297-323.