

Expected Lifetime Analysis in OECD Member Countries in Europe with the Method of Quantile Regression and Introduction to the R Commands

Özge AKKUŞ*, Barış KESKİN

Department of Statistics, Muğla Sıtkı Koçman University, Muğla, Turkey

*Corresponding Author: E-mail: ozge.akkus@mu.edu.tr; ozgeakku@gmail.com

Abstract

This paper focuses on some crucial points in statistical model selection and interpretation over the data on the expected lifetime in OECD member countries in Europe. When we study with a quantitative variable, the most common mistake we encounter is the usage of the conventional linear regression method without testing the validity of the model assumptions. The distribution of the dependent variable may be skewed and researcher may not be interested in modelling the mean value of the distribution. This is the case frequently observed in data type in Economics and Social researches. In order to reveal the degree of misspecification of the model, we have modelled the data considering some economic and health indicators first with the conventional Least Square Regression (LSR) and then Quantile Regression (QR), respectively. Since the distribution of the lifetime variable is not strictly symmetrical, it is proposed to construct QR models based on the quantiles instead of the classical LSR model with a view to decrease the loss of information obtained from the data. In the final part of the study, comprehensive interpretations are given at the 5% and 95% quantiles and application commands are presented with their explanations especially for applied researchers in R.

Key words: Economic indicators, Lifetime data, Quantiles, R programming

{**Citation:** Ozge Akkus, Barış Keskin. Expected lifetime analysis in OECD member countries in europe with the method of quantile regression and introduction to the R commands. American Journal of Research Communication, 2016,4(1): 30-53} www.usa-journals.com, ISSN: 2325-4076.

Introduction

Regression models generally try to find out the degree of correlation between dependent and independent variables using the measures of central tendency. Especially, mean of the dependent

variable conditional on the independent variable values is focused in most applications. These models are called conditional mean models and have attractive features. That is, under optimum conditions, the Least Squares Regression (LSR) and the Maximum Likelihood Estimation (MLE) methods have effective statistical properties and they are easy to apply and interpret. In spite of the attractive properties of them, they are inadequate in some cases. For instance, a researcher may not be interested in the mean of the variable of income level but the lower or upper tail of the distribution. In this case, conditional mean models are not suitable for the purpose of the study. Similarly, when the distribution of the dependent variable is thick-tailed and outliers exist, again the use of the conditional mean models is inconvenient. Furthermore, these models may not yield good results when the dependent variable has a skewed distribution (Hao and Naiman 2007). In such cases, the method of Quantile Regression (QR), which considers the quantiles of the distribution, is the primary method among the non-central methods for modelling a continuous dependent variable.

In our study, we tried to determine the important factors affecting the length of the survival time. For this purpose, we first applied the model of LSR under the assumption that the distribution of the expected lifetime variable is symmetrical. The results of the model give information about the expected mean value of the expected lifetime data. However, when the necessary statistical tests and graphical representations are obtained, we see that the data is pooled on the right side of the distribution. This means that we should focus on this side of the data. At this point, QR models are constructed and interpretations related to the both tails of the distribution are given. All statistical analyses are made with the package of R. We presented the R commands consisting of the estimation steps of QR methods in Appendix.

Literature Review

The first studies are made by Koenker and Bassett (1978) related to the conditional Quantile Regression (QR) models and they considered the models as an extension of the linear regression models. Powell (1983) generated the censored QR analysis and proposed an estimator in his work. Koenker and d'Orey (1987) showed that the algorithm developed by Barrodale and Roberts (1974), which was extended for the median regression could also be improved for QR. Buchinsky (1994) made an application on changes in the U.S wage structure between the years 1963-1987. Manning, Blumberg and Moulton (1995) used QR model to estimate the alcohol demand based on the National Health Survey Data including 18844 observation. Hyndman, Bashtannyk and Grunwald (1996) constituted semi-parametric models for daily temperature values for Melbourne city of Australia and applied QR method. Buchinsky and Hahn (1998) and Buchinsky, Hahn and Powell (2001), proposed two-step estimator in order to get out of the problems encountered in small sample situation in conjunction with the development in computer technology. Shapiro and Friedman (2001) used QR method in their study on taxation policies for focusing on the richness level of sorting for the uppermost population of 4%. Budd and McCall (2001) studied on the wage distribution for some sectors and they investigated the situation in right and left tails of the wage distribution curved by QR. Bedi and Edwards (2002), Eide, Showalter and Sims (2002), discussed

the relationship between the income and the quality of school by QR method in various points of conditional income distribution. Taddy and Kottas (2007), proposed some Bayesian techniques for all non-parametric models based on QR. Scaccia and Green (2003), modelled the conditional distribution of the dependent variable based on only one continuous independent variable for inference using nonlinear QR functions. Bassole (2007) in his study on inadequate nutrition of children in Senegal, investigated the effect of transition public basic facilities to the child health using the data from National Household Survey in 2001 recorded over 2868 household. Schaeck (2008), in his study on liberal structure of Banks emphasized the inadequacy of making inference by standard econometric techniques because the same factors did not have homogenous effect on the dependent variable throughout all the quantile values. Chen and Yu (2009), in their study on the estimation of generalized QR model made transformation to the censored regression model and proposed maximum score estimator by combining binary choice models on condition that the data is censored randomly or specifically. Kuan and Lin (2010) made comprehensive tests for non-nested nonlinear QR models and showed that their distribution is asymptotically chi-square. Koenker (2011) made a study introducing the application steps of QR model in statistical programme R. The paper by Belmaker and Jetz (2012) is on the species richness. They used quantile regression method for determining the factors affecting the species diversity.

Method

When the observation values of a series are ranked in ascending order, the values dividing the total frequency of the series into two, four, ten or a hundred equal parts are defined as “quantiles”. When there is a non-linear correlation between the explanatory variable X and the dependent variable Y , QR method is regarded as an extension of the classical linear regression. The simple linear QR model is provided in Eq.(1):

$$F^{-1}(\tau / X) = \alpha + \beta X + F_{\varepsilon}^{-1}(\tau) \quad (1)$$

where the value $F^{-1}(t/X)$ denotes the conditional τ^{th} quantile of the dependent variable obtained from the reverse of the distribution $F(Y/X)$, $F^{-1}(\tau)$ denotes the error value of the τ^{th} quantile and α and β are the constant term and the parameter vector of the model, respectively.

The QR model estimates the conditional quantile functions. In other words, the QR model estimates the models of the dependent variable in specific quantile values as a function of the set of independent variables (Koenker and Hallock 2000). When the data are nonhomogeneous, the QR method allows making inferences about the effects of the independent variable conditions for the distribution of the dependent variable and provides information on the slopes of the distribution of the dependent variable at its different points.

It is also known that the method is more robust when the distribution of the dependent variable deviates from normality. Quantiles divide the cumulative distribution function of the random

variable in equal sizes. The aim of the QR model is to obtain a family of the regression models. In this way, one has more information on the conditional distribution of the dependent variable.

Let the τ^{th} quantile ($0 < \tau < 1$) of a distribution of the variable Y be indicated with m_τ and its cumulative distribution function be shown with $F_Y(y) = P(Y \leq y)$. Following the Eq.(2), sample quantile function m_τ is defined as in Eq.(3).

$$\tau = P(y \leq m_\tau) = F_Y(m_\tau) \tag{2}$$

$$m_\tau = F_Y^{-1}(\tau) = \inf \{y / F_Y(y) \geq \tau\} \tag{3}$$

In Eq.(3), “inf” represents the smallest y value of the finite data set which provides the condition $F_Y(y) \geq \tau$. For instance, for $\tau = 0.50$, it is ensured that $Y_{\text{median}} = F_Y^{-1}(0.50)$.

Functions m_τ and F_Y are continuously increasing functions which are equal to the reverse of each other.

Quantiles are obtained by solving a simple optimization problem. This problem might be thought as the point estimation problem by making use of the prior distribution functions of the random variables. For the QR model, the control function is defined as in Eq.(4):

$$\rho_\tau(u) = u(\tau - I(u < 0)) \tag{4}$$

where $I(u < 0)$ is the indicator function and shown in Eq.(5).

$$\rho_\tau(u) = u(\tau - I(u < 0))$$

$$I(u < 0) = \begin{cases} 1, & u < 0 \\ 0, & u > 0 \end{cases} \tag{5}$$

The image of the bipartite linear function of this function is provided in Figure 1 (Koenker, 2005).

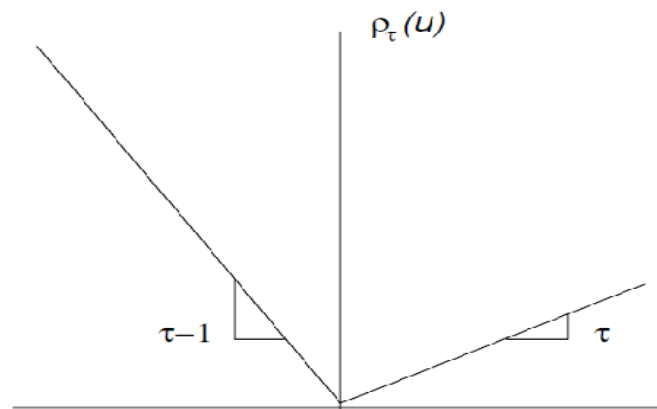


Figure 1: The Bipartite Linear Function for the Quantile Regression.

Here, the aim is to find \hat{y} estimations to minimize the expected loss. These estimations are obtained as follows:

$$E_{\rho_{\tau}}(Y - \hat{y}) = (\tau - 1) \int_{-\infty}^{\hat{y}} (y - \hat{y}) dF_Y(y) + (\tau) \int_{\hat{y}}^{\infty} (y - \hat{y}) dF_Y(y) \quad (6)$$

When the first order partial derivative of Eq.(6) is taken on \hat{y} , the following result is found:

$$0 = (\tau - 1) \int_{-\infty}^{\hat{y}} dF_Y(y) - (\tau) \int_{\hat{y}}^{\infty} dF_Y(y) = F_Y(\hat{y}) - \tau \quad (7)$$

$$F_Y(\hat{y}) = \tau \quad (8)$$

If F_Y is monotonic, there will be a y value to minimize the expected loss, with $\{y : F_Y(y) = \tau\}$. In this case, when there is a single solution, the equation $\hat{y} = F^{-1}(\tau)$ is provided. Otherwise, for the quantile intervals, it is required to select the one with the minimum losses. It is important to consider the quantiles when estimating the optimum point for asymmetric and linear loss lines. It is known that the median is used for the absolute loss under the symmetrical condition. If $F(\cdot)$ is chosen as in Eq.(9) on the basis of the empirical distribution function,

$$F_n(y) = n^{-1} \sum_{i=1}^n I(Y_i \leq y_i) \quad (9)$$

a \hat{y} is selected to achieve the following equation in order to minimize the expected loss.

$$\int \rho_{\tau}(Y - \hat{y}) dF_n(y) = n^{-1} \sum_{i=1}^n \rho_{\tau}(Y - \hat{y}) \quad (10)$$

In this way, the τ^{th} sample quantile value is obtained. Since τ is more reliable at a solution range, some uncertainties about the solution are encountered when it is constant. A constant τ is used for ease in practice. The problem of finding the sample quantile, as a simple optimization problem, is directly relevant to the sequencing of the sample observations (Koenker, 2005; Koenker and Hallock 2001).

Given the classical regression model with the conditional mean function of $E(Y / X = x) = X^T \beta$, partial derivatives are taken on variables x_j when interpreting the β parameters of a model.

$$\frac{\partial E(Y / X = x)}{\partial x_j} = \beta_j \quad (11)$$

Similarly, when the logarithmic transformation is applied to the dependent variable ($h(Y) = \log(Y)$), the expected value in Eq.(12) and the partial differential equations in Eq.(14) are obtained.

$$E(h(Y) / X = x) = x^T \beta \quad (12)$$

$$\frac{\partial E(Y / X = x)}{\partial x_j} = \frac{\partial h^{-1}(x^T \beta)}{\partial x_j} \quad (13)$$

In QR model, the parameter estimation process resembles the method of mentioned above. If it is ensured that $Q_{h(Y)}(\tau / X = x) = x^T \beta(\tau)$ when the equation $Q_{h(Y)}(\tau / X = x) = h(Q_Y(\tau / X = x))$ is assumed for any monotonic $h(\cdot)$ transformation, the following equation is provided:

$$\frac{\partial Q_Y(\tau / X = x)}{\partial x_j} = \frac{\partial h^{-1}(x^T \beta)}{\partial x_j} \quad (14)$$

As a special case, when the same is considered for the logarithmic transformation, the following equation is obtained for $Q_{\log(Y)}(\tau / X = x) = x^T \beta(\tau)$:

$$\frac{\partial Q_Y(\tau / X = x)}{\partial x_j} = \exp(x^T \beta) \beta_j \quad (15)$$

Different from the parameter estimation method of the classical LSR, a function is obtained for the τ^{th} conditional quantile instead of a function of the conditional means of Y . Related estimations are given by Eq.(16).

$$\hat{Y}_\tau = \hat{\alpha}_\tau + \hat{\beta}_\tau X \quad (16)$$

As the quantiles range from 0 to 1, all information about the conditional distribution of Y could be obtained. As a solution to the optimization problem in the QR model, the estimations of the QR coefficients are investigated and $\hat{\alpha}_\tau$ and $\hat{\beta}_\tau$ are determined to minimize the expected losses. The difference between the LSR and the QR is the structure of the loss function.

$$\hat{F}^{-1}(\tau / X) = \hat{\alpha}_\tau + \hat{\beta}_\tau X + F_\varepsilon^{-1}(\tau) \quad (17)$$

Contrary to the LSR method, values of $\hat{\alpha}_\tau$ and $\hat{\beta}_\tau$ cannot be found analytically in the QR model. Some analytical and intuitive optimization methods are required for the solution. T Specialized numerical algorithms are proposed to obtain reliable results (Koenker and Hallock 2000).

In their study, Koenker and Bassett (1978) defined an empirical quantile function as the solution to the minimization problem as,

$$\hat{m}_\tau = \arg \min_b \left\{ \sum_{i: y_i \geq b} \tau |Y_i - b| + \sum_{i: y_i < b} (1 - \tau) |Y_i - b| \right\} = \arg \min_b \left\{ \sum_i \rho_\tau(u) (Y_i - b) \right\} \quad (18)$$

where $\rho_\tau(u)$ is the control function and $I(\cdot)$ is the indicator function in Eq.(19).

$$\rho_{\tau}(u) = \begin{cases} \tau(u), & u \geq 0 \\ u(\tau-1), & u < 0 \end{cases} = u(\tau - I(u < 0)) \quad (19)$$

Let x_i ($i=1, \dots, n$) denote the $K \times 1$ dimensional vector of independent variables. The minimization function provided by Eq.(18) is called the objective function. The linear QR function is expressed as follows:

$$y_i = x_i^T \beta_{\tau} + \varepsilon_{\tau_i} \quad (20)$$

In Eq.(20), ε_{τ_i} denotes the error term and the value of the τ^{th} quantile of the error terms is zero under the condition of independent variables.

$$m_{\tau}(\varepsilon_{\tau_i} / x_i) = 0 \quad (21)$$

Following the Eq.(18) and Eq.(19), the τ^{th} conditional quantile of y_i could be written as follows.

$$m_{\tau}(y_i / x_i) = x_i \beta_{\tau} \quad (22)$$

By making use of Eq.(22), the estimated QR parameter values ($\hat{\beta}_{\tau}$'s) are provided as follows:

$$\arg \min_{\beta_{\tau} \in \mathbb{R}^K} \left\{ \sum_{i: y_i \geq x_i^T \beta_{\tau}} \tau |y_i - x_i^T \beta_{\tau}| + \sum_{i: y_i < x_i^T \beta_{\tau}} (1-\tau) |y_i - x_i^T \beta_{\tau}| \right\} = \arg \min_{\beta_{\tau} \in \mathbb{R}^K} \left\{ \sum_i \rho_{\tau}(u) (y_i - x_i \beta_{\tau}) \right\} \quad (23)$$

The QR objective function is the weighted sum of absolute deviations. In Eq.(23), it is clearly seen that when the real observation values are below the estimated values, they are weighted with $(1-\tau)$ and that when the real observation values exceed the estimated values, they are weighted with (τ) . Alternative optimization techniques such as genetic algorithm could be applied in this step in order to see the success of the methods.

Application

All democratic systems and organizations in Europe basically aim at the survival of healthy individuals and the societies consisting of healthy individuals for long years. Individuals and societies living healthily underlie power, stability, security, welfare, and happiness. The human health is directly proportional to the lifetimes of individuals. It is thought that the higher the quality of life of societies is, the longer their lifetimes will be. Besides, there are many factors which affect the lifetime. These factors include socio-economic factors (the Gross National Product (GNP)), investments in health, epidemic diseases, demographic and geographical factors, consumption of energy resources, unemployment, nourishment, and education. The most fundamental goal of health services is to protect the human health, i.e. to try to enable people to maintain their lives. It is of great importance to provide the conditions necessary to protect and develop the health of human beings – also the basic element of economic development – and to treat diseases. The OECD

(Organization for Economic Co-operation and Development) member countries increase the share they allocate from their budgets in this field every passing year.

In the study, the most important factors that affect the expected lifetime in the OECD member countries in Europe have been determined. R package has been used in all parts of the model estimations. Alternatively, Koenker (2008) and Koenker and Ng (2003) also introduced other commands in their studies.

Data

The data used are obtained from the database of the World Bank for 25 OECD member countries in Europe and 125 observations between the years 2005 and 2009. The list of the concerned countries is given in Table 1.

Table 1. The OECD member countries in Europe

Germany (DEU)	France (FRA)	Ireland (IRL)	Poland (POL)
Austria (AUT)	Luxembourg (LUX)	Spain (ESP)	Slovakia (SVK)
Belgium (BEL)	Norway (NOR)	Sweden (SWE)	Turkey (TUR)
The Czech Republic (CZE)	Portugal (PRT)	Switzerland (CHE)	Greece (GRC)
Denmark (DNK)	Slovenia (SVN)	Italy (ITA)	
Estonia (EST)	The Netherlands(NLD)	Iceland (ISL)	
Finland (FIN)	England (GBR)	Hungary (HUN)	

Dependent variable (Y) is the “Lifetime”. The summary statistics of the dependent variable are presented in Table 2.

Table 2. The summary statistics of the lifetime variable

Min.	Max.	Mean	Q ₅	Q ₂₅	Q ₅₀	Q ₇₅	Q ₉₅
72.08	82.04	78.56	72.81	77.84	79.44	80.40	81.45
Std.Dev.	Kurtosis	Std.Err.	Skewness	Std.Err.			
2.67	-0.066	0.217	-1.054	0.430			

The coefficients of the skewness (-1.054) in Table 2 provide information about a left-skewed distribution of Y. This means that the data is accumulated to the right side of the distribution. Additionally, the coefficient of Kurtosis (-0.066) indicates the lower, wider peak around the mean and thinner tails. The fact that the dependent variable has a skewed distribution is a good reason for to use of the QR methods for analyzing the data. Figure 2 presents the quantile values for the variable of the expected lifetime.

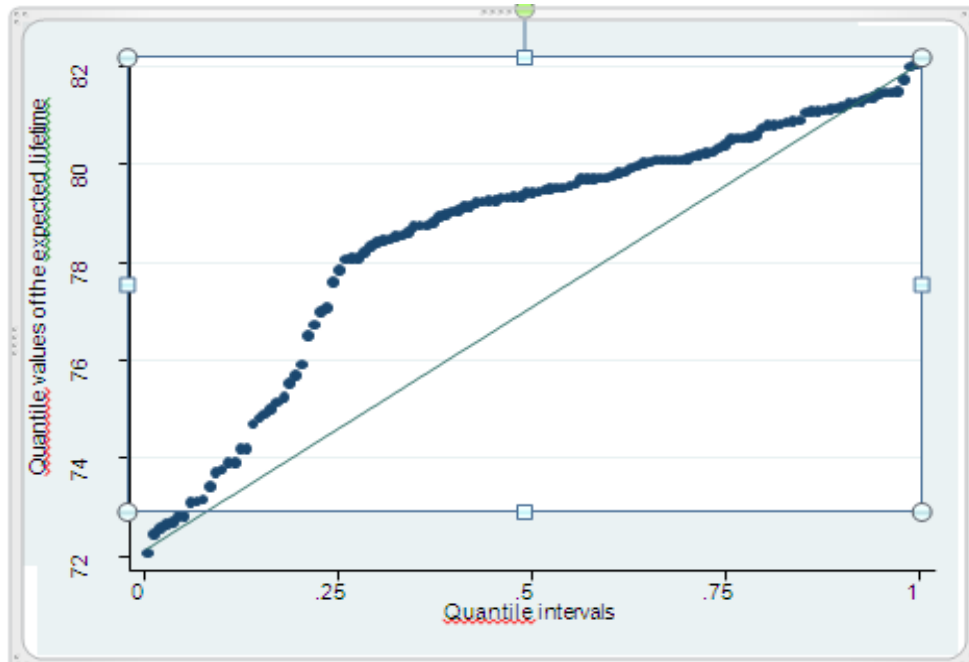


Figure 2: A Graph of the Quantile Values of the Lifetime Data

The following definitions are related to the independent variables in the model.

S_1 : The ratio of the total health expenses to the Gross National Product (GNP) (%)

S_2 : Prevalence of HIV (%)

K_1 : The net rate of enrollment in the elementary education (%)

K_2 : The percentage of fossil fuel energy consumption (%)

N_1 : The ratio of urban population to total population (%)

N_2 : Population growth (%) (as compared with the previous year)

I_1 : The rate of unemployment (%)

I_2 : The ratio of long-term unemployment to total unemployment

$\ln T_1$: GNP (\$) (logarithmic transformation was applied.)

T_2 : Inflation (the annual change in consumer prices) (%)

The summary statistics related to the explanatory variables are presented in Table 3.

Table 3. The summary statistics of the explanatory variables

Variable	Min.	Max.	Mean	Q ₅	Q ₂₅	Q ₅₀	Q ₇₅	Q ₉₅	S.D**
S ₁ *	4.99	11.98	8.87	5.97	7.79	9.00	10.04	11.21	1.60
S ₂	0.10	1.2	0.24	0.10	0.10	0.20	0.30	0.57	0.22
K ₁	89.59	99.82	96.62	92.89	95.56	97.37	98.66	99.21	2.17
K ₂	15.74	95.68	74.14	32.13	69.23	80.39	89.52	93.05	19.49
N ₁	48.30	97.38	72.65	56.27	62.70	73.40	81.82	92.27	11.87
N ₂	-0.25	2.53	0.65	-0.17	0.22	0.56	0.81	1.86	0.62
I ₁	2.30	18.00	7.08	3.00	4.70	7.10	8.40	13.65	3.08
I ₂	4.10	73.10	26.44	8.24	25.76	26.55	27.18	56.04	15.37
ln T ₁	23.22	28.92	26.44	23.59	25.76	26.55	27.18	28.66	1.41
T ₂	-4.48	12.68	2.93	-0.07	1.57	2.43	3.77	9.72	2.56

* S₁ has Normal distribution according to the Kolmogorov-Smirnov test; ** S.D: Standard Deviation

Depending on the explanatory variables, the lifetime data has been modeled by the methods of LSR and QR, respectively.

Least Square Regression (LSR)

Table 4 gives some basic statistics and F-test about the validity of the LSR model. Results indicate the validity of the model ($p = 0.00 < \alpha = 0.05$). This means that at least one explanatory variable has an effect on the dependent variable.

Table 4. ANOVA results of the LSR model

Source	SS	df	MSE	R ² = 0.827
Model	729.238	10	72.923	Adjusted R ² = 0.811
Errors	153.069	114	1.342	F(10;114) = 54.31
Total	882.307	124	7.115	p = 0.0000

SS: Sum of Squares; df: degrees of freedom; MSE: Mean Square Error

The estimated coefficients, Standard Error (S.E) of the coefficients, t values, probability (p) values and 95% confidence intervals are provided in Table 5.

Table 5. Results of the LSR model

Variable	Estimated coefficient	S.E	t	p	95% confidence int.
Constant	40.530	5.047	8.03	0.00	(30.531)_(50.528)
S_1^*	0.580	0.099	5.83	0.00	(0.383)_(0.778)
S_2	0.605	0.497	1.22	0.23*	(-0.379)_(1.589)
K_1	0.323	0.059	5.46	0.00	(0.206)_(0.440)
K_2	-0.029	0.007	-4.05	0.00	(-0.043)_(0.015)
N_1	-0.006	0.011	-0.55	0.59*	(-0.027)_(0.015)
N_2	1.215	0.204	5.95	0.00	(0.810)_(1.619)
I_1	-0.218	0.042	-5.13	0.00	(-0.302)_(0.133)
I_2	0.000	0.009	-0.03	0.98*	(-0.019)_(0.0185)
$\ln T_1$	0.215	0.105	2.06	0.04	(0.008)_(0.423)
T_2	-0.303	0.049	-6.18	0.00	(-0.400)_(0.206)

* Insignificant explanatory variable in the model; S.E: Standard Error of the estimated coefficient.

The estimated LSR model is given below.

$$E(Y / X) = 40.53 + 0.58S_1 + 0.605S_2 + 0.323K_1 - 0.029K_2 - 0.006N_1 + 1.215N_2 - 0.218I_1 + 0.00I_2 + 0.215 \ln T_1 - 0.303T_2 \quad (24)$$

Model results in Table 5 show that the contributions of prevalence of HIV (%) (S_2), the ratio of urban population to total population (N_1) and the ratio of long-term unemployment to total unemployment (I_2) are insignificant at the significance level of ($\alpha = 0.05$) ($p > \alpha$) and that those of the other variables are significant at the same significance level. In other words, prevalence of HIV, the ratio of urban population to total population, and the ratio of long-term unemployment to total unemployment have insignificant effect on lifetime of an individual.

All the interpretations made from the LSR model reflect the reality only if we study on the mean value of the dependent variable, linear correlation between the dependent and independent variables and Normality assumption of Y to test the significance of the estimated parameters. However, as mentioned above, the distribution of the expected lifetime variable is left-skewed and most of the data values are pooled on the right side of the curve.

At this point, the most proper approach for modeling the data will be to focus on the region where too much density is observed. QR models discussed below are good candidates that achieve these necessities.

Quantile Regression (QR)

QR models have been constructed as alternatives to the classical LSR models to investigate whether the explanatory variables have important effect on the extreme observations of the dependent variable. Thus, the QR models for the extreme quantile values of $\tau=0.05$ and $\tau=0.95$ have been set up.

Given the QR process, the interpretation of the tail values of the dependent variable allows obtaining important findings. It is of importance to interpret Q_5 and Q_{95} models as they are robust methods of inference when the tail of the distribution is addressed. Therefore, using the Q_5 and Q_{95} models to determine the expected lifetimes of Turkey and Switzerland, the lowest and topmost tails of the distribution of the dependent variable, respectively plays an essential role in obtaining significant results.

Findings for Turkey

When the statistics obtained from the real data set are examined, it is obviously seen that Turkey (TUR) remains at the lowest level in terms of the expected lifetime among the OECD countries in Europe.

Between 2005 and 2009, Turkey's:

Mean ratio of its total health expenses to the Gross National Product (GNP) is computed as 6.02%. With this ratio, it lags behind the OECD member countries in Europe with a mean ratio of the total health expenses to the GNP as 8.86%. Mean rate of unemployment is found as 11.22%. With this rate, it again lags behind the OECD member countries in Europe with a mean rate of unemployment as 7.08%. When the case is examined in terms of the ratio of long-term unemployment to total unemployment, it is pleasing that with a mean of 31.52%, it is lower than 35.13% – the mean of the OECD member countries in Europe.

Mean percentage of fossil fuel energy consumption is determined as 89.60%. It is seen that with this rate, it lags behind the mean of the OECD member countries in Europe with a mean percentage of fossil fuel energy consumption as 74.14%. Mean percentage of inflation is announced as 9.22%. With this rate, it lags far behind the mean of the OECD member countries in Europe with a mean percentage of inflation as 2.93%.

Mean percentage of population growth is determined as 1.33%. This rate is quite higher than the mean of the OECD member countries in Europe with a mean percentage of population growth as 0.65%.

After a brief illumination about some aspects of Turkey, Q_5 model results will lead us to come up to the more realistic interpretations related to the expected life time.

Table 6. Results of Q_5 model

Variable	Estimated coefficient	S.E ¹	t	p	[95% confidence interval]
Constant	40.688	2.862	14.22	0.00	35.019_46.358
S_1^*	0.513	0.050	10.21	0.00	0.414_0.613
S_2	1.228	0.191	6.42	0.00	0.849_1.606
K_1	0.272	0.023	11.87	0.00	0.227_0.317
K_2	-0.031	0.003	-10.92	0.00	-0.036_-0.025
N_1	-0.005	0.005	-1.13	0.26*	-0.014_0.004
N_2	1.449	0.096	15.09	0.00	1.259_1.639
I_1	-0.388	0.023	-16.17	0.00	-0.434_-0.342
I_2	0.024	0.004	5.26	0.00	0.015_0.033
$\ln T_1$	0.369	0.621	5.94	0.00	0.246_0.492
T_2	-0.401	0.050	-7.96	0.00	-0.501_-0.301

The estimated Q_5 model is presented by Eq.(25).

$$Q_5 = m_5 (Y_i / X_i) = 40.688 + 0.513S_1 + 1.228S_2 + 0.272K_1 - 0.031K_2 - 0.005N_1 + 1.449N_2 - 0.388I_1 + 0.024I_2 + 0.369\ln T_1 - 0.401T_2 \quad (25)$$

p values for the significance test of the parameters concerning the Q_5 model show that the contribution of the coefficient of the ratio of urban population to total population, variable N_1 , is insignificant at the significance level of ($\alpha = 0.05$) ($p = 0.26 > \alpha = 0.05$) and that the contributions of the coefficients of the other variables are significant at the same significance level ($p = 0.00 < \alpha = 0.05$).

When the signs of the coefficients of the Q_5 model have been examined, we see that the sign of the estimated coefficient of K_2 , N_1 , I_1 and T_2 are negative. This means that the percentage of fossil fuel consumption (K_2), the ratio of urban population to total population (N_1), the rate of unemployment (I_1) and inflation (T_2) have a reducing effect on the expected lifetime. Two important factors increasing the lifetime in Turkey seem primarily the 'Population growth (%)' as compared with the previous year with a positive coefficient of 1.449 and then the 'Prevalence of HIV' (K_2) with a coefficient of 1.228.

In general, it may be stated that it is failed to find some good results when most of the mean values of Turkey in terms of the explanatory variables under examination are compared with the mean values of the OECD member countries in Europe. With these results, it is inevitable that the

expected lifetimes of the inhabitants of Turkey are lower than those of the other people living in the OECD member countries in Europe.

Solutions for Turkey to reach or exceed the mean expected lifetimes of the inhabitants of the OECD member countries in Europe could be to share the increasing health expenses in the GNP and to reduce the rate of unemployment, the percentage of fossil fuel energy consumption and the percentage of inflation.

Table 7 provides the values of Residual Sum of Squares (RSS) calculated from the three alternative models set up for the 5% quantile slice.

Table 7. RSS values for the 5% slice of the distribution of the dependent variable

Method	RSS
LSR	20.38
Q_5	1.07
Q_{95}	166.88

When the estimated values of Q_5 models are examined for the 5% group including Turkey, it is seen that the estimation values obtained for the 5% group in terms of the expected lifetime are closer to the real values. That is, when the lower tail of the distribution is considered, we see that the regression models set up for the low quantile values are better (RSS value is 1.07 for Q_5 model). Moreover, that the finding of the model obtained with the LSR and Q_5 models to estimate the expected lifetime for Turkey significantly increase the estimation error is a crucial result.

Findings for Switzerland

Switzerland (CHE) is at the highest level in terms of the expected lifetime among the OECD countries in Europe. When the Q_{95} models are examined, we see that the estimation values obtained for high quantiles are closer to the real values. That is, when the upper tail of the distribution is considered, it is seen that the regression models set up for high quantile values are better.

The mean ratio of the total health expenses by Switzerland to the GNP between 2005 and 2009 is 10.94%. With this ratio, it is above the mean of the OECD member countries in Europe with a mean ratio of the total health expenses to the GNP as 8.86%.

Between 2005 and 2009, Switzerland's:

Mean rate of unemployment is determined as 3.90%. With this rate, it is far from the mean of the OECD member countries in Europe with a mean rate of unemployment as 7.08%. This is a quite positive result for Switzerland. When it is examined in terms of the ratio of long-term

unemployment to total unemployment, it is thought-provoking that with a mean of 36.66%, it is higher than 35.13% – the mean of the OECD member countries in Europe.

Mean percentage of fossil fuel energy consumption is determined as 53.34%. This rate is rather lower than the mean of the OECD member countries in Europe with a mean percentage of fossil fuel energy consumption as 74.14%.

Mean percentage of inflation is determined as 0.98%. It is seen that this rate is rather lower than the mean of the OECD member countries in Europe with a mean percentage of inflation as 2.93%.

Mean percentage of population growth is determined as 0.94%. This rate is higher than the mean of the OECD member countries in Europe with a mean percentage of population growth as 0.65%.

In general, it might be stated that in terms of the explanatory variables under examination, the significant majority of the mean values of Switzerland are better than the mean values of Turkey. According to these results, it is a natural outcome that the expected lifetimes of the people living in Switzerland are longer than those of the people living in Turkey. As in the case of Turkey, this explanation related to the Switzerland will be beneficial for us to compare the results of Q_{95} model under reality. Q_{95} model results are as in the following.

Table 8. Results of Q_{95} model

Variable	Estimated coefficient	S.E ¹	t	p	[95% confidence interval]
Constant	31.219	2.279	13.70	0.00	26.704_35.733
S_1^*	0.490	0.037	13.26	0.00	0.417_0.563
S_2	2.191	0.249	8.79	0.00	1.698_2.685
K_1	0.429	0.018	23.44	0.00	0.393_0.465
K_2	-0.051	0.002	-21.94	0.00	-0.056_-0.047
N_1	-0.025	0.005	-5.40	0.00	-0.034_-0.016
N_2	2.455	0.077	31.96	0.00	2.303_2.607
I_1	-0.094	0.015	-6.23	0.00	-0.124_-0.064
I_2	0.018	0.004	4.80	0.00	0.011_0.026
$\ln T_1$	0.263	0.081	3.25	0.00	0.103_0.424
T_2	-0.090	0.021	-4.35	0.00	-0.131_-0.049

¹ S.E: Standard Error of the estimated coefficient.

The Q_{95} model is presented by Eq.(26).

$$Q_{95} = m_{95}(Y_i / X_i) = 31.219 + 0.49S_1 + 2.191S_2 + 0.429K_1 - 0.051K_2 - 0.025N_1 + 2.455N_2 - 0.094I_1 + 0.018I_2 + 0.263\ln T_1 - 0.049T_2$$

(26)

When Table 8 is examined, we see that all the variables in Q_{95} model are statistically significant at a significance level of $\alpha = 0.05$ ($p = 0.00 < \alpha = 0.05$).

The signs of the coefficients of the Q_{95} model of variables K_2 , N_1 , I_1 and T_2 are negative. This means that according to this model, variables of the percentage of fossil fuel consumption (K_2), the ratio of urban population to total population (N_1), the rate of unemployment (I_1) and inflation (T_2) have a decreasing effect on the expected lifetime. The situation for the remaining explanatory variables is the direct contrary. Especially, the population growth (%) (as compared with the previous year) (N_2) with the estimated coefficient of 2.455 exceedingly has an increasing effect on lifetime and the second important factor to increase the lifetime seems the prevalence of HIV (%) (S_2) in Switzerland.

Table 9. RSS values for the 95% slice of the distribution of the dependent variable

Method	RSS
LS	24.04
Q_5	100.61
Q_{95}	4.18

As for the interpretation of Table 9, it is an expected result that the RSS value (4.18) of the Q_{95} model sharply decreases compared with other alternatives. Therefore, the use of the Q_{95} model in the estimations of lifetime for Switzerland will yield more significant results.

Concluding Remarks

The basic purpose of the regression analysis is to set up a statistically significant model which minimizes the difference between real and estimation values. This provides good prediction values about the new data which is out of the used in modeling. This procedure of minimization is carried out differently for different types of regression models. One of the classical methods is to use linear regression that models the mean of the dependent variable conditional on the explanatory variables. This method actually could yield good results under various assumptions such as linearity of the dependent and independent variables, Normality condition for the distribution of the dependent variable Y , etc. However, data structures do not mostly provide all these assumptions. For instance the distribution of Y may not be symmetrical and some outliers may be observed in the data set after applying the outlier detection rules. All these will lead to the deviations from the normality of the distribution of Y .

Therefore, robust methods have been developed as alternatives to the linear regression methods that use the LSR technique in the parameter estimation. The robust methods including the QR reduce the importance of the extreme observations by giving lower weighing values to them. The QR

provides information on the entire conditional distribution of Y according to the explanatory variables X but not conditional mean. Methods developed on quantile values are widely preferred since they are more robust against the extreme values in the dependent variable. In this study, it is concluded that the QR models yield better estimation results when the extreme values of the dependent variable are addressed.

Factors influencing the expected lifetime in OECD member countries in Europe including Turkey are discussed in the application part of the study. When we examine the data structure, we realize that the right-skewed distribution of the concerned variable, lifetime data is observed. Therefore, we have decided to apply QR models for %5 and %95 quantiles.

When the results obtained from the Q_5 model are interpreted for the data structure for Turkey, we conclude that the percentage of fossil fuel consumption, the ratio of urban population to total population, the rate of unemployment and inflation have a reducing effect on the expected lifetime. Two important factors increasing the lifetime in Turkey seems the population growth (%) as compared with the previous year and 'Prevalence of HIV' with positive coefficients of 1.449 of 1.228, respectively. With these results, it is obvious that the expected lifetimes of the inhabitants of Turkey are lower than those of the other people living in the OECD member countries in Europe.

As for Q_{95} model for the data structure for Switzerland, percentage of fossil fuel consumption, the ratio of urban population to total population, the rate of unemployment and inflation have a decreasing effect on the expected lifetime. On the other hand, it is concluded that the population growth (as compared with the previous year) and the prevalence of HIV have an increasing effect on lifetime for Switzerland.

References

- Barrodale, I., Roberts, F.D.K. (1974). Solution of an over determined system of equations in the 1 norm. *Communications of the ACM*. 17, 319-320.
- Bassole, L. (2007). Child Malnutrition in Senegal: Does Access to Public Infrastructure Really Matter? A Quantile Regression Analysis, *CERDI-CNRS*, Université d'Auvergne, Jobmarket Paper.
- Bedi, A., Edwards, J. (2002). The impact of school quality on earnings and educational returns evidence from a low-income country. *J Dev Econ*. 68, 157-185.

- Belmaker, J., Jetz, W. (2012). Regional pools and environmental controls of vertebrate richness. *The American Naturalist* 179(4), 512-523.
- Budd, J.W., Mc, Call, B.P. (2001). The grocery stores wage distribution: A semi-parametric analysis of the role of retailing and labor market institutions. *Industrial and Labor Relations Review* 54, 484-501.
- Buchinsky, M. (1994). Changes in the U.S wage structure 1963-1987: Application of quantile regression. *Econometrica* 62(2), 405-458.
- Buchinsky, M., Hahn, J. (1998). An alternative estimator for censored quantile regression. *Econometrica* 66, 653-710.
- Buchinsky, M., Hahn, J., Powell, J. (2001). Two step estimation of semiparametric censored regression models. *J Econometrics* 103, 73-110.
- Chen, C., Yu, K. (2009). Automatic bayesian quantile regression curve fitting. *Statistics and Computing* 19, 271-281.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning* 20(3), 273-297.
- Hao, L., Naiman, D.Q. (2007). *Quantile Regression, Quantitative Applications in the Social Sciences*. Sage Publications, California.
- Hyndman, R.J., Bashtannyk, D.M., Grunwald, G.K. (1996) Estimating and visualizing conditional densities. *J Comput Graph Stat.* 5, 315-336.
- Keskin, B., Akkuş, Ö. (2014). OECD'ye Üye Ülkelerde Beklenen Yaşam Süresinin Kantil Regresyon Yöntemi ile Analizi, 15. *Uluslararası Ekonometri, Yöneylem Araştırması ve İstatistik Sempozyumu*, 22-25 Mayıs 2014, Süleyman Demirel University, Isparta.
- Khmaladze, E.V. (1981). Martingale approach in the theory of goodness of fit tests. *Theory of Probability and Its Applications* 26(1), 240-257.
- Koenker, R., Bassett, G. (1978). Regression quantiles. *Econometrica* 46, 33-50.
- Koenker, R., d'Orey, V. (1987). Computing regression quantiles. *Appl Statist.* 36, 383-393.
- Koenker, R., Hallock, F.K. (2000). *Quantile Regression: An Introduction*. Urbana-Champaign, University of Illinois, 24p.
- Koenker, R., Hallock, F.K. (2001). Quantile regression. *J Econ Perspect.* 15, 143-156.

- Koenker, R., Xiao, Z. (2002). Inference on the quantile regression processes, *Econometrica* 70, 1583-1612.
- Koenker, R., Ng, P. (2003). Sparse M: A sparse linear algebra package for R. *J Stat Softw.* 8, 1-9.
- Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, New York, 349p.
- Koenker, R. (2008). Censored quantile regression redux. *J Stat Softw.* 27, 1-24.
- Vapnik, V., (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc.
- Koenker, R. (2011). Quantile Regression in R: A vignette. technical Reportç. Cran.r-project.org. 20p.
- Kuan, C.M., Lin, H.Y. (2010). An encompassing test for non-nested quantile regression models. *Economics Letters* 107, 257-260.
- Manning, W.G., Blumberg, L., Moulton, L.H. (1995). The demand for alcohol: The differential response to price. *J Health Econ.* 14, 123-148.
- Portnoy, S., Koenker, R. (1997). The gaussian hare and the laplacian tortoise: Computability of squared-error vs. absolute-error estimators. *Statistical Science* 12, 279-300.
- Powell, J.L. (1983). The asymptotic normality of two-stage least absolute deviations estimators. *Econometrica* 51,1569–1576.
- Scaccia, L., Green, P.J. (2003). Bayesian growth curves using normal mixtures with nonparametric weights. *J Comput Graph Stat.* 12, 308-331.
- Schaeck, K. (2008). Bank liability structure, FDIC loss, and time to failure: A quantile regression approach. *J Financ Serv Res.* 33, 163-79.
- Shapiro, I., Friedman, J. (2001). *Income Tax Rates and High-Income Taxpayers: How Strong Is the Case For Major Rate Reduction?*, Washington. Center for Budget and Policy Priorities. 23p.
- Taddy, M., Kottas, A. (2007). A Nonparametric Model-Based Approach to Inference for Quantile Regression. Department of Applied Mathematics and Statistics. University of California. Santa Cruz. 34p.

Appendix: R Commands for QR Model

R commands constructed for Quantile Regression Model are introduced with their explanations below.

Step 1. Data set named “uyver” in “.xls” format is copied to the file “lifetime” in C:\lifetime\uyver.csv

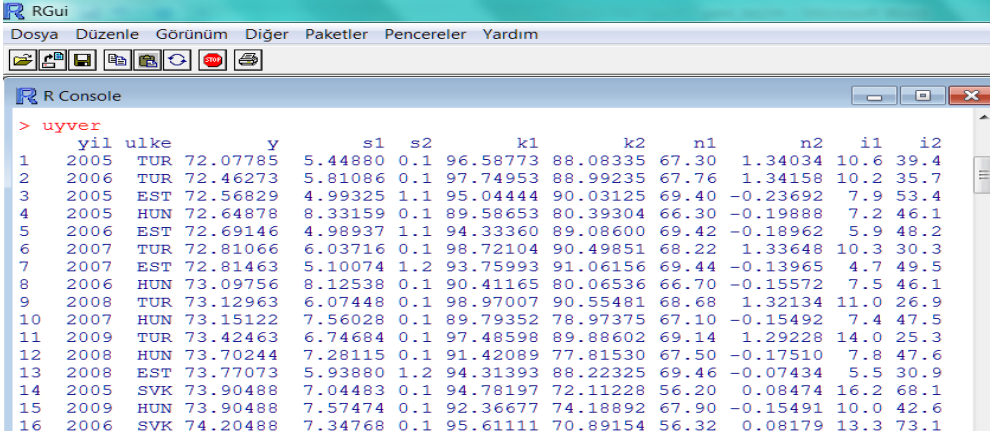
Step 2. R programme is opened and the following commands are run to transfer the data to R.

```
>setwd("C:/lifetime")
```

```
>getwd() # Introduce the data bus to R.
```

```
>uyver <- read.table("uyver.csv", header=T, sep=",") #Calls the Excel file saved by the name “uyver”.
```

```
>uyver # Displays the numerical values of “uyver” in the R screen.
```



	yil	ulke	Y	s1	s2	k1	k2	n1	n2	i1	i2
1	2005	TUR	72.07785	5.44880	0.1	96.58773	88.08335	67.30	1.34034	10.6	39.4
2	2006	TUR	72.46273	5.81086	0.1	97.74953	88.99235	67.76	1.34158	10.2	35.7
3	2005	EST	72.56829	4.99325	1.1	95.04444	90.03125	69.40	-0.23692	7.9	53.4
4	2005	HUN	72.64878	8.33159	0.1	89.58653	80.39304	66.30	-0.19888	7.2	46.1
5	2006	EST	72.69146	4.98937	1.1	94.33360	89.08600	69.42	-0.18962	5.9	48.2
6	2007	TUR	72.81066	6.03716	0.1	98.72104	90.49851	68.22	1.33648	10.3	30.3
7	2007	EST	72.81463	5.10074	1.2	93.75993	91.06156	69.44	-0.13965	4.7	49.5
8	2006	HUN	73.09756	8.12538	0.1	90.41165	80.06536	66.70	-0.15572	7.5	46.1
9	2008	TUR	73.12963	6.07448	0.1	98.97007	90.55481	68.68	1.32134	11.0	26.9
10	2007	HUN	73.15122	7.56028	0.1	89.79352	78.97375	67.10	-0.15492	7.4	47.5
11	2009	TUR	73.42463	6.74684	0.1	97.48598	89.88602	69.14	1.29228	14.0	25.3
12	2008	HUN	73.70244	7.28115	0.1	91.42089	77.81530	67.50	-0.17510	7.8	47.6
13	2008	EST	73.77073	5.93880	1.2	94.31393	88.22325	69.46	-0.07434	5.5	30.9
14	2005	SVK	73.90488	7.04483	0.1	94.78197	72.11228	56.20	0.08474	16.2	68.1
15	2009	HUN	73.90488	7.57474	0.1	92.36677	74.18892	67.90	-0.15491	10.0	42.6
16	2006	SVK	74.20488	7.34768	0.1	95.61111	70.89154	56.32	0.08179	13.3	73.1

Figure 3: Data Input Screen of R

After the completion of the data inputation step, analyses could be made by calling the related libraries.

Descriptive Statistics

The following command written in R console table gives the summary statistics such as max, min, mean and quantile values of the variable “uyver”.

```
>summary(uyver)
```

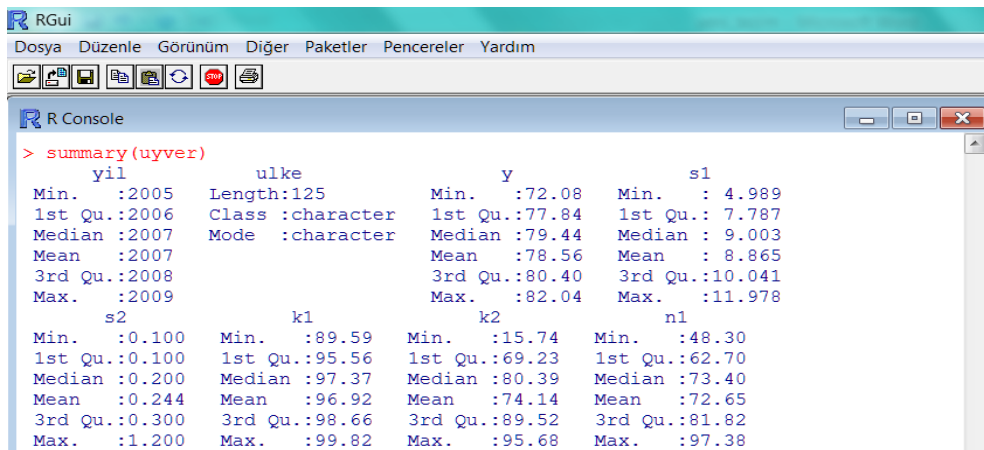


Figure 4: Summary Statistics

Step 3. Application of the Quantile Regression (QR) in R

1. Installing the Packages

First, required package for the application of the QR model called “quantreg” must be installed as,

```
>install.packages("quantreg")
```

After that, the following command is run to reach all libraries in the installed package.

```
>library(quantreg)
```

If any problem is occurred in this step, we could call the commands help(“package=”quantreg”) or help(rq) for the alternative solutions to the problem.

2. Application

The general structure of the application of the QR models in R is as the following.

```
rq(formula, tau=0.5, data, subset, weights,na.action,method="br", model = TRUE, contrasts),
```

Descriptions of the statements inside the function of “rq” are given below.

Formula: An expression that regresses the dependent and independent variables each other with the symbol “~” like $(y \sim x + z)$

Tau: The quantile value that will be estimated in interval (0,1).

Data: Describes the data set that is used in the analysis.

Subset: A vector constituted by the subset of the observations used in the estimation process.

Weights: A vector that shows the weights of observations. The length of the weights should be the same as the number of observation and must be positive.

Na.action: A function that filters the missing data. If only one missing value is observed, use **na.fail** for constituting the error and use **na.omit** for deleted one or more than one observation.

Method: An algorithmic method used for the model fitting.

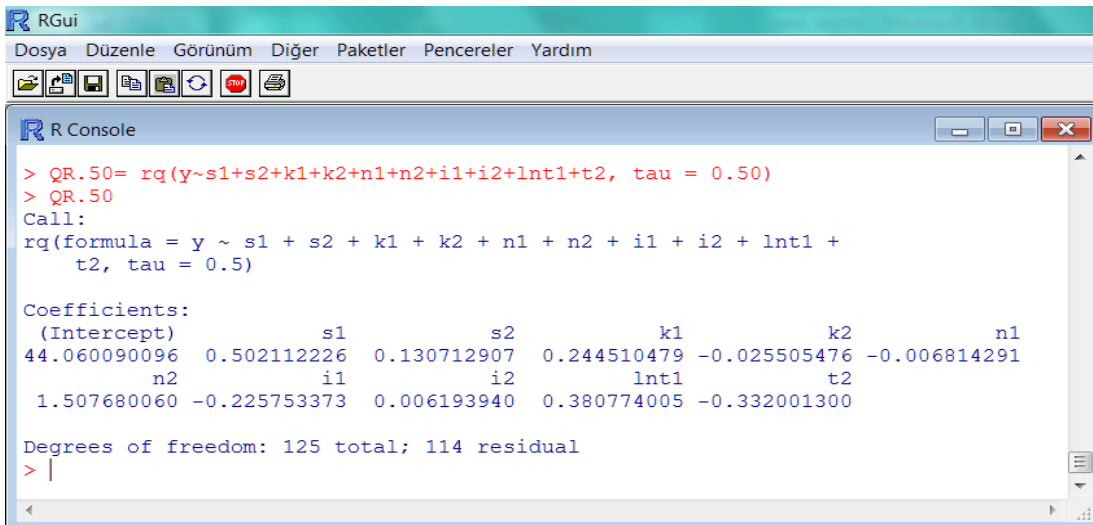
Contrasts: List of oppositions for some or all of the assumed factors. Elements of this list should have the same name as the variables.

QR models could be applied for different tau values. For example, for tau:0.50, the median regression is obtained and should be run the following commands for the results.

```
>QR.50= rq(y~s1+s2+k1+k2+n1+n2+i1+i2+lnt1+t2, tau = 0.50)
```

#QR.50 is the name that we assigned to the model.

>QR.50 command is written in R console table to see the obtained coefficients.



```
RGui
Dosya Düzenle Görünüm Diğer Paketler Pencereleer Yardım

R Console

> QR.50= rq(y~s1+s2+k1+k2+n1+n2+i1+i2+lnt1+t2, tau = 0.50)
> QR.50
Call:
rq(formula = y ~ s1 + s2 + k1 + k2 + n1 + n2 + i1 + i2 + lnt1 +
t2, tau = 0.5)

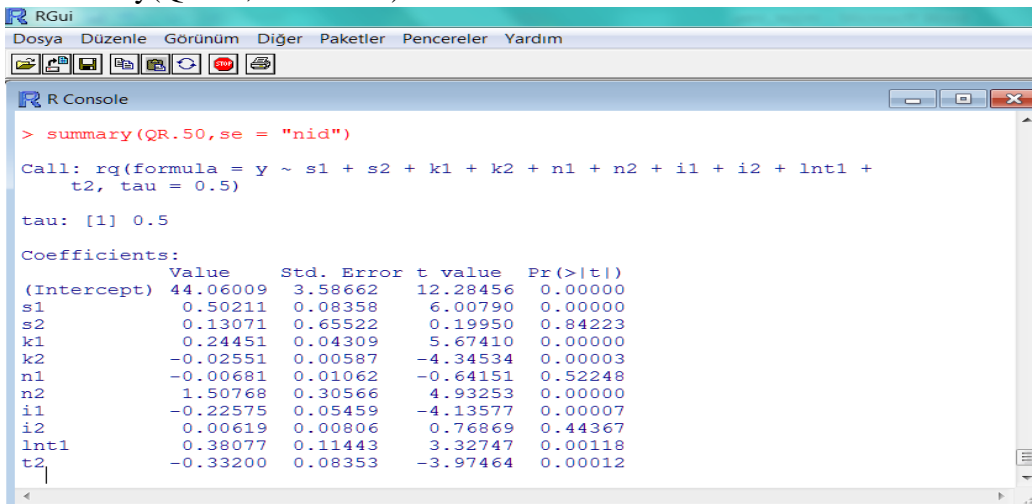
Coefficients:
(Intercept)          s1          s2          k1          k2          n1
44.060090096  0.502112226  0.130712907  0.244510479 -0.025505476 -0.006814291
          n2          i1          i2          lnt1          t2
1.507680060 -0.225753373  0.006193940  0.380774005 -0.332001300

Degrees of freedom: 125 total; 114 residual
> |
```

Figure 5: QR Estimation Results for Tau=0.50

The following command generally gives the summary statistics such as the coefficients and the confidence intervals of the coefficients of the estimated model.

```
> summary(QR.50, se = "nid")
```



```
RGui
Dosya Düzenle Görünüm Diğer Paketler Pencereleer Yardım

R Console

> summary(QR.50, se = "nid")

Call:
rq(formula = y ~ s1 + s2 + k1 + k2 + n1 + n2 + i1 + i2 + lnt1 +
t2, tau = 0.5)

tau: [1] 0.5

Coefficients:
      Value   Std. Error t value Pr(>|t|)
(Intercept) 44.06009  3.58662  12.28456 0.00000
s1           0.50211  0.08358   6.00790 0.00000
s2           0.13071  0.65522   0.19950 0.84223
k1           0.24451  0.04309   5.67410 0.00000
k2          -0.02551  0.00587  -4.34534 0.00003
n1          -0.00681  0.01062  -0.64151 0.52248
n2           1.50768  0.30566   4.93253 0.00000
i1          -0.22575  0.05459  -4.13577 0.00007
i2           0.00619  0.00806   0.76869 0.44367
lnt1        0.38077  0.11443   3.32747 0.00118
t2          -0.33200  0.08353  -3.97464 0.00012
```

Figure 6: QR Summary Statistics for Tau=0.50

In the estimation of the standard errors of the estimated model coefficients, several alternative techniques could be used. The default method for estimation of the standard errors in “Quantreg” package is called “sandwich” formula that uses the Hall-Sheather bandwidth rule with the commands `se = “nid”` in “summary” Alternatives are `se = “ker”` based on Powell kernel estimator and `se = “boot”` for estimating bootstrap standard errors.

Another important part of the algorithm is the determination of the “method”. Simplex algorithm of Barrodale and Roberts (1974) is used as default in R console. This algorithm expressed by (`method = “br”`) was defined by Koenker and d’Obey (1987). Alternatively, Frisch-Newton algorithm introduced by Portnoy and Koenker (1997) and denoted by the expression (`method = “fn”`) in R could be used if the number of observations is considerably large. Various other methods for different conditions were developed for this step. You can use “Help” command for any support in R. The following commands calculate and display the estimated model residuals in R.

```
>rs50=resid(QR.50)
```

```
>rs50
```

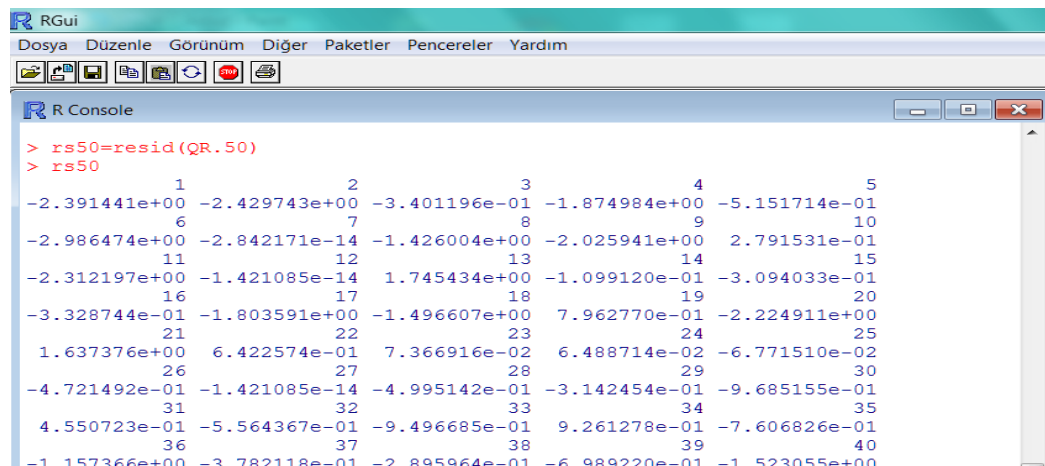


Figure 7: QR Residuals for Tau=0.50

`>fitted(QR.50)` command provides and displays getting the estimated values.

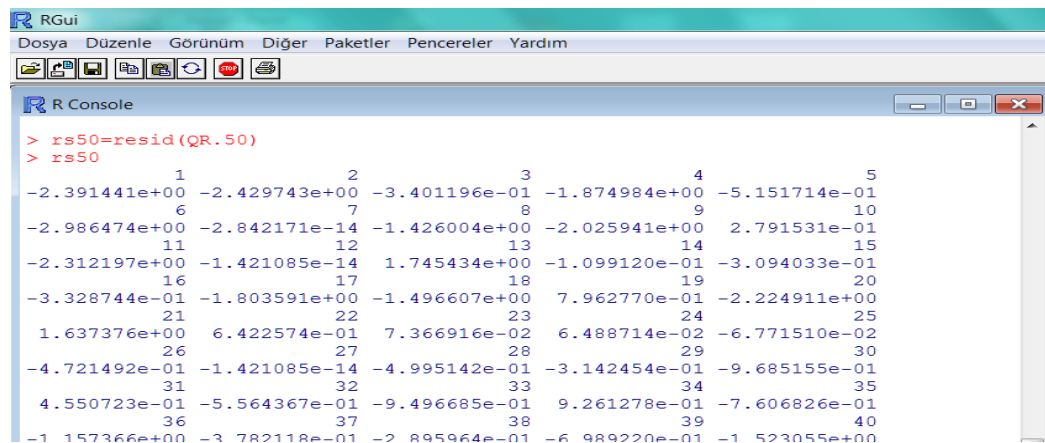
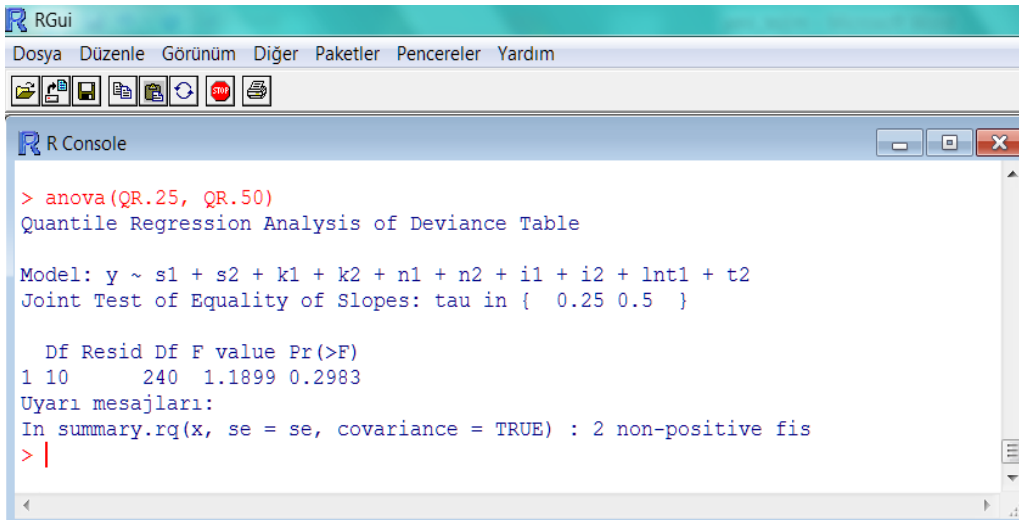


Figure 8: Estimated Values for Tau=0.50 in QR

ANOVA table for the fitted models could be obtained by running the commands,
>anova(QR.50,QR.25)



```

RGui
Dosya Düzenle Görünüm Diğer Paketler Pencereleer Yardım

R Console
> anova(QR.25, QR.50)
Quantile Regression Analysis of Deviance Table

Model: y ~ s1 + s2 + k1 + k2 + n1 + n2 + i1 + i2 + lnt1 + t2
Joint Test of Equality of Slopes: tau in { 0.25 0.5 }

  Df Resid Df F value Pr(>F)
1 10      240 1.1899 0.2983
Uyarı mesajları:
In summary.rq(x, se = se, covariance = TRUE) : 2 non-positive fis
> |

```

Figure 9: ANOVA Tables of QR Models For Tau=0.25 and Tau=0.50

Equality of the parallel slopes throughout the quantiles could be tested by using the “**anova**” commands but it is generally known that a goodness of fit test expressed in R as **Khmaladze.test** introduced by Khmaladze (1981) gives better results. This test is developed by Koenker and Xiao (2002). If required, detailed information related to this test could be accessed by the following link. <http://astrostatistics.psu.edu/datasets/R/html/quantreg/html/khmaladze.test.html>

Graphical representation for the model estimated over only two variables could be presented by using the commands “**plot(.)**” and the “**abline(.)**” for determining the colour and shape of the graphs. Here, “**data**” represents the name of the data set.

```
>attach(data)
```

```
>plot(x, y, cex= 0.25, type= "n", xlab= "...", ylab= "...")
```

```
>points (x, y, cex = 1, col = "blue")
```

```
>abline(DR, lty= 3, col= "blue")
```

#lty commands determined the structure of the line (discontinuous, dotted etc.)

```
>abline(QR.05, col= "red")
```

```
>abline(QR.25, col= "pink")
```

```
>abline(QR.50, lty=2, col= "green")
```

```
>abline(QR.75, col= "black")
```

```
>abline(QR.95, col= "orange")
```