The Use of Nonparametric Quantile Regression and Least Median of Squares Regression for Construction of Growth Curves of Weight

Ağırlıkça Büyüme Eğrilerinin Oluşturulmasında Nonparametrik Kantil Regresyon ve En Küçük Medyan Kareler Regresyonunun Kullanımı

Handan ANKARALI,[a]
Özge YILMAZ,[a]
Münevver KIZILAY,[b]
İlknur ARSLANOĞLU,[b]
Duygu AYDIN[a]

Departments of
[a]Biostatistics,
[b]Pediatrics,
Düzce University Faculty of Medicine,
Düzce

Geliş Tarihi/*Received:* 15.05.2012
Kabul Tarihi/*Accepted:* 04.12.2012

The abstract of this study was published in the Proceedings Book of XII. National Biostatistics Congress, 28 June-1 July 2010, Van, Turkey.

Yazışma Adresi/*Correspondence:*
Handan ANKARALI
Düzce University Faculty of Medicine,
Department of Biostatistics, Düzce,
TÜRKİYE/TURKEY
hankarali@yahoo.com

ABSTRACT Objective: This study aimed to investigate the use of the Least Median Squares (LMS) regression and nonparametric quantile regression model comparatively to describe children's weight growth. **Material and Methods:** Two different models were used to obtain the percentile curves to identify the weight growth in girls. The first model was obtained by LMS regression, which is a member of the family of nonlinear parametric quantile regression. In addition, in this model percentile curves used to define growth were generated using the Box-Cox transformation and the cubic spline. The second model was obtained by nonparametric quantile regression that did not require the assumption of a normal distribution for construction of percentile curves. This method is a flexible approach, as well as being computationally simple. The weight values obtained from 1771 healthy girls aged between 6 and 14 years were used in both methods. The data were collected from the cross-sectional study conducted in schools in Düzce city. **Results:** The distributions of weight measurements according to ages revealed that there were deviations from normality at some ages, there were deviated values in the tail regions of the distribution, and the variances changed according to ages. Using both methods, growth curves were constructed separately for each age group. Predicted values of the LMS and the non-parametric quantile regression models were similar for each age. In addition, the error sum of squares derived from non-parametric quantile regression was lower than that derived from LMS regression for each percentile curve. Moreover, the estimations obtained from both methods were highly correlated with the estimation values of the province İstanbul, which was considered the reference. **Conclusion:** When the assumptions about the distribution and variances of the data are violated and these assumptions cannot be achieved with the transformation, nonparametric quantile regression method gives more reliable results for the creation of percentile curves.

Key Words: Growth&development; growth charts

ÖZET Amaç: Bu çalışmada, çocuklardaki ağırlıkça büyümeyi tanımlamada, En Küçük Medyan Kareler (LMS) regresyonu ve nonparametrik kantil regresyon modellerinin karşılaştırmalı olarak incelenmesi amaçlanmıştır. **Gereç ve Yöntemler:** Çalışmada, kız çocuklarının ağırlıkça büyümesini tanımlamak için persentil eğrilerinin elde edilmesinde iki farklı büyüme modeli kullanılmıştır. Bu yöntemlerden birisi olan LMS regresyonu, doğrusal olmayan parametrik kantil regresyon ailesinden olup, Box-Cox transformasyonu ve kübik eğri yardımıyla persentil eğrilerini oluşturur. İkinci yöntem, LMS yöntemine alternatif olabilecek ve dağılım ön şartı gerektirmeyen nonparametrik kantil regresyondur. Bu yöntem, hesaplama kolaylıklarının yanı sıra esnek bir yaklaşımdır. Her iki yöntemin uygulamasında, yaşları 6 ile 14 arasında değişen toplam 1771 sağlıklı kız çocuğundan elde edilen ağırlık değerleri kullanılmıştır. Bu veriler, Düzce ilindeki okullarda yürütülen kesitsel bir çalışmaya aittir. **Bulgular:** Yaşlara göre ağırlık ölçümlerinin dağılımı incelendiğinde, bazı yaşlarda normal dağılımdan sapmaların gözlendiği, dağılımın kuyruk bölgelerinde sapan değerlerin olduğu ve varyansların yaşlara göre değiştiği belirlenmiştir. Her iki yöntem yardımıyla, her bir yaş grubu için ayrı ayrı büyüme eğrileri oluşturulmuştur. Bu eğrilerden elde edilen tahmini ağırlık değerleri birbirine benzer çıkmıştır. Bunun yanı sıra oluşturulan her bir persentil eğrisi için parametrik olmayan kantil regresyon tahminlerinin hatası, LMS yöntemine göre daha küçük bulunmuştur. Ayrıca her iki yöntem sonucunda elde edilen tahminlerin, referans olarak kabul edilen İstanbul ili tahmin değerleri ile hayli kuvvetli bir ilişki içinde olduğu belirlenmiştir. **Sonuç:** Verilerin dağılımı ve varyanslar ile ilgili varsayımların bozulduğu durumlarda ve transformasyonla bu varsayımların sağlanamadığı koşullarda, nonparametrik kantil regresyon metodu, persentil eğrilerinin oluşturulmasında daha güvenilir sonuçlar vermektedir.

Anahtar Kelimeler: Büyüme ve gelişme; büyüme gözlem kartları

doi: 10.5336/medsci.2012-30442

Turkiye Klinikleri J Med Sci 2013;33(3):692-701

Growth is expressed as a structural increase and is defined as an increase of body volume and mass with the increase of number and size of cells.[1] Various growth models have been used to describe the growth occurring in different physical dimensions or organs of the human body and to reveal whether the current growth is healthy growth. Some of them focus on the mean and the others are able to describe the entire conditional distribution of the dependent variable (logistic nonlinear mixed model). Moreover, these models are classified as linear and nonlinear models or parametric, semi-parametric, and non-parametric growth models.[1,2]

Quantile regression (QR) techniques are widely used in preliminary medical diagnosis to identify unusual subjects in the sense that the value of some particular measurement lies in one or another tail of the appropriate reference distribution. QR can therefore help us to obtain a more complete picture of the underlying relationship between outcome such as weight and covariate such as age. QR results are characteristically robust to outliers and heavy-tailed distributions.[3-5]

Least Median Squares (LMS) method belongs to the family of nonlinear semi-parametric quantile regression and is a generalized form for the determined quantiles of median regression. Recently some alternative methods have been developed to LMS, which have insufficient results, and one of those is the nonparametric quantile regression (NPQR) model.[6,7]

This study aimed to compare the use of LMS regression model and nonparametric quantile regression model to describe the weight growth of a healthy child.

MATERIAL AND METHODS

LMS METHOD

The distribution of the outcome variable changes according to age is shown by the reference centile curves. The changing distribution of three curves representing the median (M), coefficient of variation (S), and skewness (L or λ), which is expressed as a Box-Cox power, are summarized by the LMS method. Three curves (L, M, and S) can be fitted as cubic splines with non-linear regression using penalized likelihood, and the amount of smoothing needed can be given in terms of smoothing parameters or equivalent degrees of freedom. These parameters can be interpreted as the dimensionality of the fitted function and are measured by computing the trace of the pseudo-projection matrix defining the estimator.[6,8] In the selection of appropriate effective degrees of freedom (edf) value, the changes in the Deviance and Akaike Information Criteria (AIC) values are considered. The curves are created by taking combinations of L, M, S where Deviance or AIC values have the smallest number. The edf values that give good results in many circumstances for L, M, S parameters, are 3, 5, 3 respectively.

LMS method selects appropriate λ, applies Box-Cox transformation according to this lambda and transforms Y (t) ‘s measurements to standardized Z (t) values to ensure normality. The transformed observations are independent and normally distributed with constant variance.[9]

$$Z(t)=\frac{\left[\frac{Y(t)}{\mu(t)}\right]^{\lambda(\tau)}-1}{\lambda(\tau)\sigma(t)}$$

After transformation, quantile curve for α ∈ [0,1] is estimated by the following model.

$$Q\,(\alpha It)=\mu(t)[1+\lambda(t)\sigma(t)\phi^{-1}(\alpha)]^{1/\lambda(\tau)}$$

In this equation;

α : The lower tail area of the centile,

z_α : The normal equivalent deviate of size α.[9,10]

A main assumption of the LMS method is that the data are normally distributed. The main problem about the assumption may be the presence of kurtosis, which could not be adjusted with transformation. However, kurtosis tends to be less important than skewness as a contributor to non-normality.[6]

When the assumptions of normality and the constancy of the variance of outcome variable are not valid, NPQR method is preferred instead of LMS for construction of growth charts.[11]

NONPARAMETRIC QUANTILE REGRESSION

Various non-parametric quantile regression NPQR tools have been proposed in the literature.[11-14] Non-parametric spline-based quantile regression is a flexible approach, as well as being computationally simple, allowing a different grade of curvature for each conditional quantile.[13]

Kernel, Local polynomials or smoothing splines are used for smoothing the percentile curves predicted in NPQR method. We used series estimators, which are constructed based on cubic *B*-splines. These splines directly extend linear and low order polynomial models. To generate *B-splines*, n, m, and p values are needed. Here, p is the degree of B-spline, n +1 is the number of control points, m is the number of basic functions and m= n+p+1. In general, the lower the degree, the closer a *B-spline* curve follows its control polyline.[2,15,16] The approximation abilities of *B-splines* are well known from the theoretical standpoint; spline models retain the advantages of algebraic polynomials. Iterative steps are used for smoothing and appropriate smoothing coefficient is determined by error of model in each iteration and proximity to each other of coefficients estimated for each smoothing parameter. Smoothing parameters are estimated by reweighted least square method iteratively.

In the NPQR method, the quantiles are estimated as a linear combination of multiple basis function.[16]

DATA

Weight is one of the most important indicators for growth of children and adolescents. The distribution of weight and changes in variances of weights are important issues for the selection of suitable growth curve in each age group.[16]

The data were obtained from a cross-sectional study carried out in Düzce city, located in the northwest of Turkey with a heterogeneous socio-cultural structure. Therefore weights were measured only once from each subject in each age group. Schools were selected by stratified sampling method according to income level. Weights were measured twice in 770 girls from high-income schools, 288 girls in middle income schools and 713 girls in low-income schools. The arithmetic mean of the two measurements were recorded. Weights were measured with precision digital scales (Felix brand), which are sensitive to 0.1 kg. In summary, the data set consisted of the body weight measurements of 1771 healty Turkish girls aged between 6 and 14 years, in Düzce city between 2009 and 2010. The study was approved by the ethical committee of the university.

In this study, unconditional growth curves were constructed for each age by using LMS and NPQR. However, other covariates (other social-demographic features) were not considered to build these curves.

Fitting of the percentile curves was performed using the LMS Chart Maker Ligth software program (version 2.3; The Institute of Child Health, London) and ***R*** [is a public domain language for data analysis sustained by the R Development Core Team (2004)]. ***R*** commands for LMS and NPQR methods were given in the Appendix section.

RESULTS

DESCRIPTIVE VALUES

Mean, median, standard deviation (SD), minimum ,and maximum values were given in Table 1. The proportional distribution of age groups were as follows: 6.21% of subjects were 6 years-old, 13.15% were 7 years-old, 12.93% were 8 years-old, 12,87% were 9 years-old, 11.85% were 10 years-old, 14.54% were 11 years-old, 13.21% were 12 years-old, 11.57% were 13 years-old, and 3.71% were 14 years-old. Although the rate of 6 and 14 years-old children was smaller than the rate of other age groups, the sample sizes of those two age groups was not too small. The standard deviation seemed to increase with age (Table 1).

Percentile values of this study were frequently used to identify human growth percentile values (Table 2).

The distribution of weights according to ages, revealed that the weight values of the girls who were 8, 9 and 11 years old were close to the deviation from normality. In contrast, the values in the other age groups were distributed normally (Table 2).

APPENDIX

R commands for Parametric Quantile Regression Model (LMS)

```
> library (lmsqreg)
> centiles <- c(0.03,0.05,0.10,0.25,0.5,0.75,0.9,0.95,0.97)
> lms.fit <- lmsqreg.fit (weight, age, edf=c(3,5,3),pvec=centiles)
> plot (lms.fit)
> points (age, weight, col="blue")
> print (lms.fit)
```

R command for Nonparametric Quantile Regression Model (NPQR)

```
> library (MASS)
> plot (age, weight, xlab = "age", ylab = "weigth (kg)")
> library (splines)
> plot (age, weight, xlab = "age", ylab = " weigth (kg)",
+ type = "n")
> points(age, weight, cex = 0.75)
> X <- model.matrix (kilo ~ bs(age, df = 5))
> library (quantreg)
> for (tau in c (0.03,0.05,0.10,0.25,0.5,0.75,0.9,0.95,0.97)) {
+ fit <- rq(weight ~ bs(age, df = 5), tau = tau)
+ weight.fit <- X %*% fit$coef
+ lines(age, weight.fit)
+ }
```

Table 3 showed that skewness and kurtosis coefficients of weight values of the girls between 7 and 11 years-old showed more deviation than the normal distribution values. These results were compatible with the normality test in Table 2. Moreover, the variances at different ages were heterogeneous, leading to the conclusion that variances are not stable (Figure 1).

On the other side, the raw percentile curves, which were not transformed and not smoothed according to age were given in Figure 2. The percentile values that were taken into consideration, are important in health literature. Because of violations of the assumptions for parametric tests, firstly Box-Cox transformation was applied to the data. And then percentile curves were constructed with the LMS method.

RESULTS OF LMS METHOD

Weight values deviated from normal distribution at some ages and deviated values existed in tails of distribution. In addition, the variances changed according to age. Since these assumptions were damaged, the LMS method that includes Box-Cox transformation was used. The corrected percentile curves, that were predicted according to ages after the LMS method was applied, and the L, M, S estimation curves were given in Figure 3.

TABLE 1: Descriptive statistics of weights.

Age	N	Mean	Median	SD	Minimum	Maximum
6	110	22,248	21,800	2,752	16,400	30,000
7	233	24,913	24,800	3,816	17,200	38,800
8	229	27,303	26,600	4,874	18,600	43,800
9	228	30,937	30,200	5,585	21,400	47,400
10	210	35,554	34,800	6,804	23,200	57,200
11	256	39,324	38,400	7,778	25,200	66,400
12	234	46,523	45,500	8,565	28,800	77,000
13	205	49,908	49,200	8,648	30,600	76,800
14	66	53,324	51,700	8,109	39,200	75,200

N: Number; SD: Standard deviation.

TABLE 2: Raw percentile values of weight.

	Percentile values									Normality test
Age	3	5	10	25	50	75	90	95	97	(P)
6	17,600	17,820	18,620	20,350	21,800	23,850	25,580	27,870	28,734	0,587
7	19,004	19,570	20,200	22,200	24,800	26,800	29,800	32,530	33,396	0,169
8	20,360	21,000	21,600	24,000	26,600	29,400	34,600	37,500	40,220	0,05
9	23,148	23,690	24,400	26,600	30,200	34,500	39,200	42,110	43,478	0,066
10	25,798	26,600	27,400	30,350	34,800	39,000	45,100	48,360	52,268	0,108
11	28,342	28,600	30,140	33,600	38,400	44,150	50,400	53,010	57,000	0,073
12	32,620	33,800	36,000	40,350	45,500	52,050	58,700	61,650	63,580	0,318
13	34,636	36,520	40,440	43,900	49,200	55,100	61,800	65,600	68,476	0,269
14	41,208	42,140	44,340	46,750	51,700	57,550	66,080	69,190	69,994	0,633

TABLE 3: Shape parameters of weight distribution and homogeneity test results.

.			Levene test for Homogenity of variance (df1=8, df2=1762)	
Age	Skewness±S.E.	Kurtosis±S.E	Test statistics	P value
6	0,474 ± 0,230	0,151 ± 0,457	33,292	0,0001
7	0,792 ± 0,159	1,000 ± 0,318		
8	1,022 ± 0,161	1,074 ± 0.320		
9	0,729 ± 0,161	0,030 ± 0,321		
10	0,728 ± 0,168	0,315 ± 0,334		
11	0,697 ± 0,152	0,274 ± 0,303		
12	0,478 ± 0,159	0,054 ± 0,317		
13	0,413 ± 0,170	0,111 ± 0,338		
14	0,637 ± 0,295	-0,194 ± 0,582		

SE: Standard error.

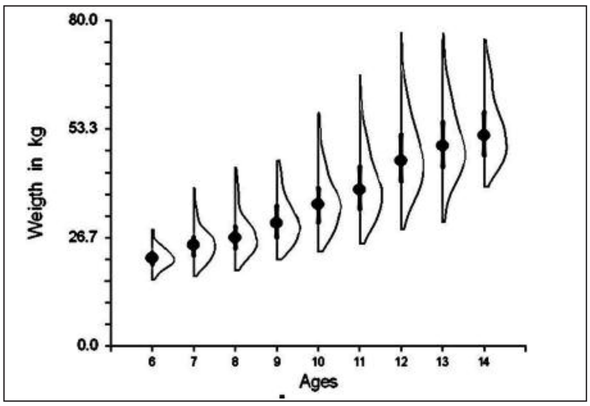

FIGURE 1: Distribution of weight values of girls at each age.

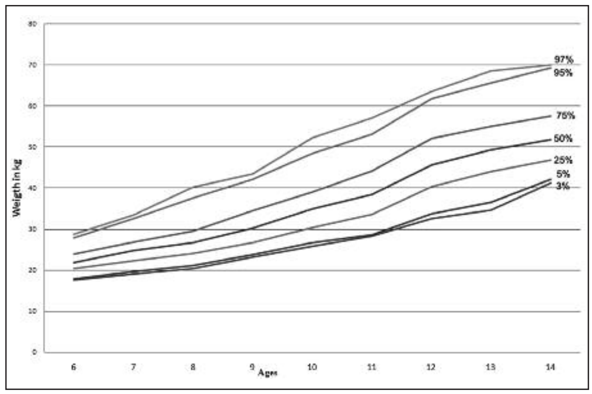

FIGURE 2: Raw percentile values in each group.

While there was a linear increase in L and M values in Figure 3, S value seemed to increase until the age 10-11 years followed by a decreasing trend later. The blue spots in the graphic show real weight values and the discontinuous lines shows the percentile curves that are estimated with LMS.

The percentile values that were estimated with LMS method were given in Table 4. These values were obtained after both transformation and smoothing.

In Table 5, the estimated values of L, M, and S parameters according to ages were shown. Percentile values were estimated by using these values.

RESULTS OF NONPARAMETRIC QUANTILE REGRESSION

Percentile curves were reconstructed by using nonparametric quantile regression on the same data be-

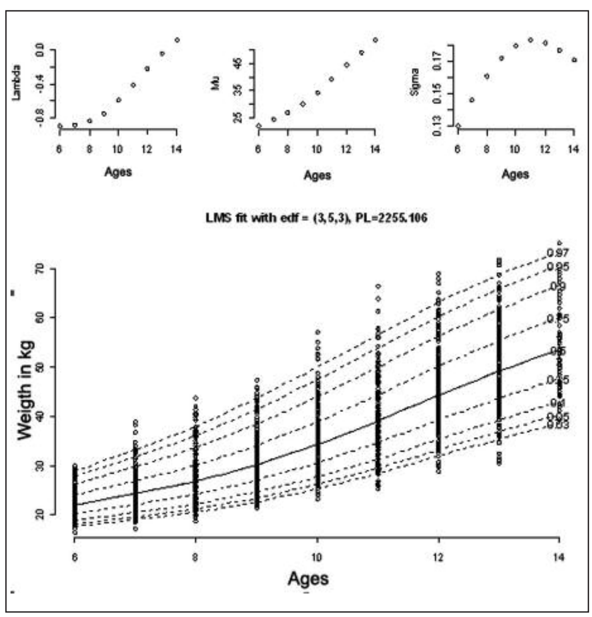

FIGURE 3: Predicted percentile curves according to the LMS method.

TABLE 4: Smoothing percentile curves obtained from LMS.

	LMS								
	Predicted Percentile Values								
Age	3	5	10	25	50	75	90	95	97
6	17,627	18,080	18,823	20,202	21,979	24,077	26,320	27,861	28,958
7	18,999	19,539	20,428	22,099	24,285	26,920	29,797	31,813	33,266
8	20,468	21,102	22,152	24,141	26,774	29,994	33,567	36,105	37,952
9	22,562	23,319	24,577	26,964	30,137	34,030	38,360	41,440	43,682
10	25,209	26,124	27,640	30,510	34,301	38,908	43,965	47,517	50,077
11	28,309	29,409	31,225	34,631	39,067	44,350	50,012	53,901	56,659
12	31,923	33,225	35,360	39,318	44,371	50,245	56,367	60,470	63,332
13	35,369	36,854	39,271	43,687	49,209	55,467	61,813	65,971	68,827
14	38,694	40,341	42,999	47,790	53,662	60,168	66,612	70,755	73,567

LMS: Least median of squares.

cause there was deviation in tails of distribution even if Box-Cox method was applied. The percentile values that were estimated by the NPQR method were given in Table 6.

The percentile curves for the NPQR method were illustrated in Figure 4. In the figure, symmetric percentile curves were shown with the same color. The effect of covariate (age) occured differently on different percentile curves in both LMS and NPQR methods because the estimated percentile curves were not exactly parallel to each other.

The results of symmetric percentiles were given in Tables 7-11, which involves NPQRL model coefficients for each percentile values. 'bs' is the base function and all the models constitute of 5 base functions. "B" is the regression coefficient of the base function and in the next column standard

TABLE 5: L, M and S values at each age.

Age	L	M	S (Coeff. Var.)
6	-0,903	21,978	0,130
7	-0,891	24,285	0,146
8	-0,852	26,774	0,160
9	-0,755	30,138	0,172
10	-0,602	34,302	0,180
11	-0,414	39,067	0,183
12	-0,226	44,372	0,182
13	-0,048	49,209	0,177
14	0,112	53,662	0,171

TABLE 6: Smoothing percentile values obtained from NPQR model.

	NPQR								
	Predicted Percentile Values								
Age	3	5	10	25	50	75	90	95	97
6	17,60	18,40	18,80	20,40	22,00	24,00	25,40	27,60	28,60
7	18,85	19,21	20,00	22,00	24,40	26,80	29,80	32,50	33,25
8	20,80	21,20	22,00	24,00	26,99	29,80	34,27	37,08	39,20
9	23,13	23,68	24,40	26,50	30,20	33,65	39,20	42,07	44,80
10	25,60	26,20	27,03	29,72	34,34	38,80	44,84	48,00	50,75
11	28,40	29,20	30,60	34,00	39,34	44,91	50,94	54,53	57,00
12	31,80	33,13	35,62	39,40	44,80	51,20	57,00	61,00	63,20
13	36,00	37,60	40,80	44,28	49,40	55,85	62,00	66,70	67,91
14	41,20	42,00	44,40	46,60	51,60	56,80	64,80	68,80	69,40

NPQR: Nonparametric quantile regression.

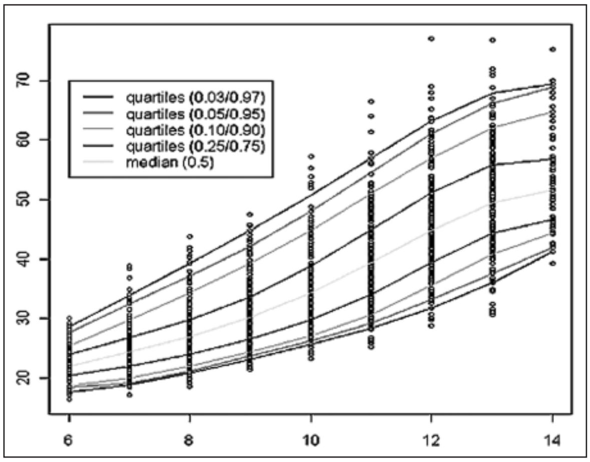

FIGURE 4: Predicted percentile curves for the nonparametric quantile regression (NPQR) method.

error of that coefficient. Next column of the table indicates standart error of B coefficient. t value shows the hypothesis test of the B coefficient. The intercept value is equal to the minimum weight value in the percentile. Percentile value of a new subject is estimated by using the intercept value and the regression coefficients.

Table 12 includes the mean error of the models, and standard deviation of those errors with minimum and maximum values. Errors were computed by summing the squares of the differences between the estimated values from the LMS and NPQR models and their observed weight values. This table revealed that the sum of error squares of the NPQR estimates was lower than the LMS estimates in all percentile values. Especially this difference was highest at a percentile value of 10%. This result indicated that the estimations of NPQR method were more successful (Table 12).

The studies that are conducted in İstanbul constitute the basis of growth curves that are developed in Turkish children. Neyzi et al.[17] described the reference values for various percentile values of girls aged between 12-14 years (Table 13). These values were obtained via the LMS method. When the reference values described by Neyzi et al.[17] were compared with the LMS estimations (Table 4) and NPQR estimations (Table 6), similar results were obtained.

TABLE 7: NPQR model coefficients for 3[th] and 97[th] percentiles.

	Model coefficients for 3. percentile				Model coefficients for 97. percentile			
Model coefficients	**B**	**Std. Error**	**t value**	**Pr(>|t|)**	**B**	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	17.6	0.2653	66.33881	0	28.6	0.40026	71.45442	0
bs(age, df = 5)1	0.78279	0.76576	1.02224	0.30681	5.23904	2.05577	2.54846	0.0109
bs(age, df = 5)2	4.68415	1.16665	4.01504	0.00006	13.89144	3.04651	4.55979	0.00001
bs(age, df = 5)3	11.20669	1.7237	6.50151	0	30.30452	3.19347	9.48952	0
bs(age, df = 5)4	17.83722	2.68745	6.63724	0	41.48217	3.89841	10.64079	0
bs(age, df = 5)5	23.6	2.33766	10.09558	0	40.8	1.10715	36.85129	0

B: Regression coefficient of the base function; NPQR: Nonparametric quantile regression; Pr(>|t|): Actual p value.

TABLE 8: NPQR model coefficients for 5[th] and 95[th] percentiles.

	Model coefficients for 5. percentile				Model coefficients for 95. percentile			
Model coefficients	**B**	**Std. Error**	**t value**	**Pr(>|t|)**	**B**	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	18.4	0.46745	39.36284	0	27.6	1.17017	23.58634	0
bs(age, df = 5)1	0.0028	0.98809	0.00283	0.99774	5.30435	2.48221	2.13695	0.03274
bs(age, df = 5)2	4.62928	0.90399	5.12095	0	11.51232	2.44075	4.71672	0
bs(age, df = 5)3	10.81077	1.27702	8.46561	0	29.08565	3.3393	8.7101	0
bs(age, df = 5)4	19.4362	1.75624	11.06693	0	40.34435	2.8138	14.33805	0
bs(age, df = 5)5	23.6	1.16034	20.33878	0	41.2	2.04663	20.13061	0

B: Regression coefficient of the base function; bs: Base function; df: Difference; NPQR: Nonparametric quantile regression; Pr(>|t|): Actual p value.

TABLE 9: NPQR model coefficients for 10th and 90th percentiles.

	Model coefficients for 10. percentile				Model coefficients for 90. percentile			
Model coefficients	B	Std. Error	t value	Pr(>|t|)	B	Std. Error	t value	Pr(>|t|)
(Intercept)	18.8	0.491	38.2889	0	25.4	0.70229	36.16737	0
bs(age, df=5)1	0.66805	0.98684	0.67695	0.49852	4.49177	2.35082	1.91072	0.0562
bs(age, df=5)2	4.77571	0.81647	5.84919	0	11.1996	2.51131	4.45967	0.00001
bs(age, df=5)3	11.36023	1.20989	9.3895	0	27.52696	2.2648	12.15425	0
bs(age, df=5)4	23.36828	1.73096	13.50019	0	38.08063	2.59948	14.64933	0
bs(age, df=5)5	25.6	1.36196	18.79646	0	39.4	2.58122	15.26411	0

B: Regression coefficient of the base function; bs: Base function; df: Difference; NPQR: Nonparametric quantile regression; Pr(>|t|): Actual p value.

TABLE 10: NPQR model coefficients for 25th and 75th percentiles.

	Model coefficients for 25. percentile				Model coefficients for 75. percentile			
Model coefficients	B	Std. Error	t value	Pr(>|t|)	B	Std. Error	t value	Pr(>|t|)
(Intercept)	20.4	0.22347	91.28936	0	24	0.37096	64.69689	0
bs(age, df = 5)1	1.43838	0.64853	2.21791	0.02669	2.91646	0.93504	3.11907	0.00184
bs(age, df = 5)2	4.54785	0.93105	4.88462	0	6.65378	1.26115	5.27598	0
bs(age, df = 5)3	13.72472	1.0978	12.50207	0	22.62521	1.8715	12.08937	0
bs(age, df = 5)4	25.85018	1.10155	23.4672	0	34.38789	2.06874	16.62264	0
bs(age, df = 5)5	26.2	0.70188	37.32829	0	32.8	2.06976	15.84727	0

B: Regression coefficient of the base function; bs: Base function; df: Difference; NPQR: Nonparametric quantile regression; Pr(>|t|): Actual p value.

The relationship between the reference values[17] and the estimated values obtained in the present study was given in Table 14 collectively. Table 14 revealed that there was a significant and very strong relationship between reference values [17] and the estimation values obtained by the LMS and NPQR methods in the present study.

TABLE 11: NPQR model coefficients for 50th percentile.

	Model coefficients for 50. percentile			
Model coefficients	B	Std. Error	t value	Pr(>|t|)
(Intercept)	22	0.35468	62.02722	0
bs(age, df = 5)1	2.4435	0.86432	2.82707	0.00475
bs(age, df = 5)2	5.94821	0.9659	6.15819	0
bs(age, df = 5)3	18.4484	1.22	15.12163	0
bs(age, df = 5)4	29.11412	1.42472	20.43498	0
bs(age, df = 5)5	29.6	1.17999	25.08489	0

B: Regression coefficient of the base function; bs: Base function; df: Difference; NPQR: Nonparametric quantile regression; Pr(>|t|): Actual p value.

DISCUSSION

Many studies have defined human growth in the medical research literature.[6,8,17-21] Observing growth stages is a very important process for understanding whether the development of children is healthy or not. Assessment of children's growth is monitored by using percentile curves that are developed for weight, height and head circumference measurements according to the age and gender.

In this study, percentile curves were constructed using weight data of healthy girls by using the LMS and NPQR methods where the distribution of weight values was skewed and the variance of the weights varied according to age. Some researchers have emphasized that the quantile regression model gives better results in such circumstances.[22] The results of the Monte Carlo simulation study showed that non-parametric quantile regression methods might provide better and more robust estimation results especially when the underlying model was non-linear and/or the error term followed a non-normal distribution compared to their parametric counterparts.[23]

TABLE 12: Goodness of fit results of LMS ve NPQR methods.

	Percentile	N	Mean Error	Std. Dev.	Sum of error square	Minimum	Maximum
Evaluation of the LMS estimates	3	9	-0.325	0.734	5.264	-1.881	0.733
	5	9	-0.177	0.704	4.246	-1.645	0.809
	10	9	-0.067	0.794	5.082	-1.282	1.085
	25	9	-0.052	0.593	2.84	-1.032	1.031
	50	9	-0.131	0.518	2.298	-1.129	0.667
	75	9	-0.021	0.753	4.534	-1.805	0.674
	90	9	-0.411	1.08	10.843	-2.333	1.282
	95	9	-0.212	1.015	8.648	-1.395	1.645
	97	9	-0.28	1.284	13.891	-2.268	1.881
Evaluation of the NPQR estimates	3	9	-0,074	0,586	2,793	-1,364	0,820
	5	9	-0,098	0,567	2,661	-1,080	0,670
	10	9	-0,057	0,322	0,858	-0,460	0,380
	25	9	0,133	0,436	1,679	-0,400	0,950
	50	9	-0,008	0,498	1,987	-0,940	0,700
	75	9	0,066	0,645	3,370	-0,760	0,850
	90	9	0,334	0,713	5,069	-0,540	1,700
	95	9	-0,051	0,746	4,476	-1,520	0,650
	97	9	0,337	0,785	5,949	-1,322	1,518

N: Number; LMS: Least median of squares; NPQR: Nonparametric quantile regression.

LMS method is used mostly for construction of percentile curves. However, this method was suggested to give good results when assumptions like the homogenity of the variances and normal distribution were met.[6,8] In addition, the NPQR method is more flexible than the LMS about taking different covariates to the model and it is more suitable when the parametric model assumptions are not valid.[16]

Our results suggest that the sum of error squares of the NPQR estimates was lower than LMS estimates in all percentile values. Especially this difference was highest at a percentile value of

TABLE 13: Weight values of Turkish girls estimated from Istanbul reference values (Neyzi et al 2008).

	Girls' weight percentiles (reference values)/İstanbul						
Age	3%	10%	25%	50%	75%	90%	97%
6	15,7	17,0	18,6	20,6	22,9	25,3	27,9
7	17,2	18,7	20,6	22,9	25,7	28,6	31,9
8	18,9	20,8	22,9	25,7	28,9	32,4	36,5
9	20,9	23,1	25,6	28,9	32,8	37,0	41,8
10	23,0	25,6	28,7	32,6	37,3	42,3	48,0
11	26,4	29,6	33,4	38,2	43,7	49,5	55,9
12	32,0	35,8	39,9	45,1	50,9	56,8	63,1
13	37,4	41,4	45,1	50,0	55,5	60,8	66,6
14	41,6	45,0	48,8	53,3	58,3	63,2	68,5

TABLE 14: The relationships between estimates and reference values.

	The relation between LMS predicted values in this study and reference values in Table 13			The relation between NPQR predicted values in this study and reference values in Table 13		
Percentiles	r	R-sqr (%)	p	r	R-sqr (%)	p
3th	0.995	99.0	<0.0001	0.996	99.1	<0.0001
10th	0.997	99.4	<0.0001	0.999	99.9	<0.0001
25th	0.999	99.7	<0.0001	0.999	99.9	<0.0001
50th	0.999	99.8	<0.0001	0.999	99.8	<0.0001
75th	0.998	99.5	<0.0001	0.999	99.7	<0.0001
97th	0.997	99.4	<0.0001	0.998	99.6	<0.0001

LMS: Least median of squares; NPQR: Nonparametric quantile regression.

10%, which led to the conclusion that the NPQR method was less affected by outliers. In addition, the reference values in the Istanbul study and estimates of our models were highly correlated.[17]

REFERENCES

1. Gonzalez X, Miles D. Wage inequality in a developing country: decrease in minimum wage or increase in education returns. In: Fitzenberger B, Koenker R, Machado JAF, eds. Economic Applications of Quantile Regression. 1st ed. Heidelberg: Physica-Verlag; 2002. p.135-45.
2. Koenker R. Quantile treatment effects. In: Chester A, Jackson M, eds. Quantile Regression. 1st ed. Cambridge: Cambridge University Press; 2005. p.26-42.
3. Yu K, Lu Z, Stander J. Quantile regression: applications and current research areas. The Statistician 2003;52(3):331-50.
4. Koenker R, Hallock K. Quantile regression: an introduction. Journal of Economic Perspectives 2001;15(4):143-56.
5. Cole TJ, Cortina-Borja M, Sandhu J, Kelly FP, Pan H. Nonlinear growth generates age changes in the moments of the frequency distribution: the example of height in puberty. Biostatistics 2008;9(1):159-71.
6. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 1992;11(10):1305-19.
7. Chen L. An introduction to quantile regression and the QUANTREG Procedure. Statistics and Data analysis 2005; 213–230. SUGI30 Proceedings, available at children and adolescents. Conference on Quantitative Social Science Research Using R. available at http://www.cis.fordham.edu/QR2009/presentations/RefGrowthCharts.pdf
8. Wei Y, Pere A, Koenker R, He X. Quantile regression methods for reference growth charts. Stat Med 2006;25(8):1369-82.
9. Chuchana P, Marchand D, Nugoli M, Rodriguez C, Molinari N, Garcia-Sanz JA. An adaptation of the LMS method to determine expression variations in profiling data. Nucleic Acids Res 2007;35(9):e71.
10. Koenker R, Basset G. Regression quantiles. Econometrica 1978;46(1):33-50.
11. Koenker R, Ng P, Portnoy S. Quantile smoothing splines. Biometrika 1994;81(4): 673-80.
12. Yu K, Jones MC. Local linear quantile regression. The Journal of the American Statistical Association 1998;93(441):228-37.
13. Bosch RJ, Ye Y, Woodworth GG. A convergent algorithm for quantile regression with smoothing splines. Comput Stat Data Anal 1995;19(6):613-30.
14. Thompsona P, Caia Y, Moyeeda R, Reeveb D, Standera J. Bayesian nonparametric quantile regression using splines. Commput Stat Data Anal 2010;54(4):1138-50.
15. He X. Quantile curves without crossing. The American Statistician 1997;51(2):186-92.
16. Takeuchi I, Le QV, Sears T, Smola AJ. Nonparametric quantile regression. Journal of Machine Learning Research 2006;7(7):1231-64.
17. Neyzi O, Günöz H, Furman A, Bundak R, Gökçay G, Darendeliler F, et al. [Body weight, height length, head circumference and body mass index reference values for Turkish children]. Turkish Pediatric Journal 2008;51(1):1-14.
18. Royston P. A parametric model for ordinal response data, with application to estimating age-specific reference intervals. Biostatistics 2000;1(3):263-77.
19. Gasser T, Ziegler P, Seifert B, Prader A, Molinari L, Largo R. Measures of body mass and of obesity from infancy to adulthood and their appropriate transformation. Ann Hum Biol 1994;21(2):111-25.
20. Royston P, Wright EM. A method for estimating age-specific reference interval ('normal ranges') based on fractional polynomials and exponential transformation. Journal of the Royal Statistical Society, Series A, General 1998;161(1):79-101.
21. Heagerty P, Pepe MS. Semi-parametric estimation of regression quantiles with application to standardizing weight for height and age in US children. Applied Statistics 1999;48(4): 533-51.
22. Landajo M, Andrés J, Lorca P. Measuring firm performance by using linear and non-parametric quantile regressions. Applied Statistics 2008;57(2):227-50.
23. Min I, Kim I. A Monte Carlo comparison of parametric and non-parametric quantile regressions. Applied Economic Letters 2004; 11(2):71-4.