Study important variables that effecting in obesity by Bayesian and non-Bayesian methods

Assist. Lecturer :shatha Awwad AL-Fatlawy
University of Al-Qadisiyah
College of Administration and Economics
Department of Statistics
satha.awwad@qu.edu.iq

Abstract

Regression models are important tools in estimating the effect relationships between the dependent variable and a set of independent variables. To estimate the parameters of the studied model, there are various estimation methods Bayesian and non-Bayesian are will used . The obesity affected with many independent variables some these variables have direct effecting and non- direct, therefor choosing optimal variables that effecting in our model considered give us high explanatory power for the studied phenomenon, in this variable there are subset variables effecting on obesity via many different quantiles level.

Keywords: variable selection , Bayesian and non-Bayesian methods , obesity problem, quantiles level

Introduction

Obesity is defined as a high percentage of total body fat, which leads to an increase in body weight at a rate 20% or more Of normal weight for age and height. Obesity usually occurs due to the consumption of more calories than they are burned over time, which leads to the accumulation of fat. The diagnosis of obesity is based on the body mass index (BMI), which is link to the health of the appearance of the body depending on the length and weight and calculated as in the following equation

BMI= Body weight / Square of length by meter

If the value of BMI is (less than 20) is considered Less than normal weight, but when the value of BMI is (20-25) is considered normal weight , when the value of BMI is (25-30) is considered overweight , when the value of BMI is (30-35) is considered Light obesity (first degree) , when the value of BMI is (35-40) is considered middle obesity (second degree) , when the value of BMI is (more than 40) is considered excessive obesity (third degree) . there are many methods avoid the obesity where many local and international institutions are making a number of efforts aimed at combating obesity, such as spreading awareness about obesity and its dangers, and providing a number of ways to help people with obesity to overcome it. Despite the effectiveness of these efforts, combating obesity radically requires a lot of effort, and most importantly, increasing the extent of cooperation between the various institutions to combat this problem, which is responsible for raising the odds of infection with a huge number of diseases among people of all ages, including among children. It is worth mentioning here that each of us has a fundamental role to play in combating obesity, and this role is represented in protecting ourselves from it by adopting healthy eating habits, the most prominent of which is reducing the rate of consumption of canned food and fast-prepared food as much as possible, and most importantly, committing to physical activity in a moderate amount periodically; By doing so, we significantly reduce our own risk of obesity, while we encourage others to follow healthy habits. It is important to note that parents' adherence to healthy eating habits is the most effective way to protect their children from obesity. Although prevention is better than cure, treatment is also important and an absolute necessity to combat obesity. The first step in treating people with obesity is to consult a specialist doctor to identify all the factors contributing to their obesity in order to find the best ways to combat each of these causes. Here, it should be noted that the methods of combating obesity differ relatively from one case to another. However, all methods depend mainly on increasing the rate of physical activity in addition to following a healthy diet that is low in calories and integrated with nutrients. You can learn more information about that by reading my article (What are the proven obesity solutions) and (how to lose weight in a natural way). There are many variables effected on obesity . in this paper , we provide Bayesian and non-Bayesian method to selection optimal variables that effecting in obesity problem .This paper is organized by four section , first section general information about quintile regression model, second section focus on non-Bayesian quantile regression , and third section focus on Bayesian quantile regression, forth section study the data of obesity via estimation of model parameters . fifth section focus on conclusion and recommendations.

1-Quantile Regression

The relationship affecting between the mean of a dependent variable (outcome variable) Y in a set of independent variable Xs may be is not effective . In many real data the mean regression may be not a good tools to describing the relationship between dependent variables Y and some independent variables Xs Especially when the data is so spread out . For example, the effect of demographic properties and maternal conduct on the weight of infant born was studied by (Abrevaya (2008)). in the United States. This study was focused on low birth weight for infants which causes many health problems. This data was analyzed by mean regression, the conditional mean was not a good approach for low tail distribution. Quantile regression that proposed by (Koenker and

Bassett (1978)) as an more extension compared classical mean regression in conditional different levels of quantiles for a response variables is consider a good tools with low tail distribution. Quantile regression model is able of give us Full image about different quantiles of the relationships between dependent variables and independent variables. Recently, Quantile regression model has got much interest in theoretical and application aspects . Quantile regression model is become effective tools in many fields of such as:, , growth chart (Wei et al.,(2006)) agricultural economics (Kostov and Davidova, (2013) and others. The Quantile regression models have a Attractive features compared with classical regression model. Quantile regression model is a robust against outliers and outliers (Koenker and Geling, (2001)). Quantile regression effective with any supposition about the random error . Quantile regression give us good inference at when the random error is non-normal, Quantile regression model is effective with economic problems. All quantile regression have a good features and informative model in application fields..

$$y_i = \beta_{0(\tau)} + \beta_{1(\tau)}X_{1i} + \cdots + \beta_{p(\tau)}X_{pi} + \epsilon_i, \qquad \tau \in (0,1), \qquad [1\ ]$$

, where y_i is the dependent variable Xs is a p-dimensional vector of independent variables, β_t is a coefficients vector of Quantile regression model.

There are infinite regression lines belong to the interval $t \in (0,1)$). Each regression line belong to quantile level, as the following figure

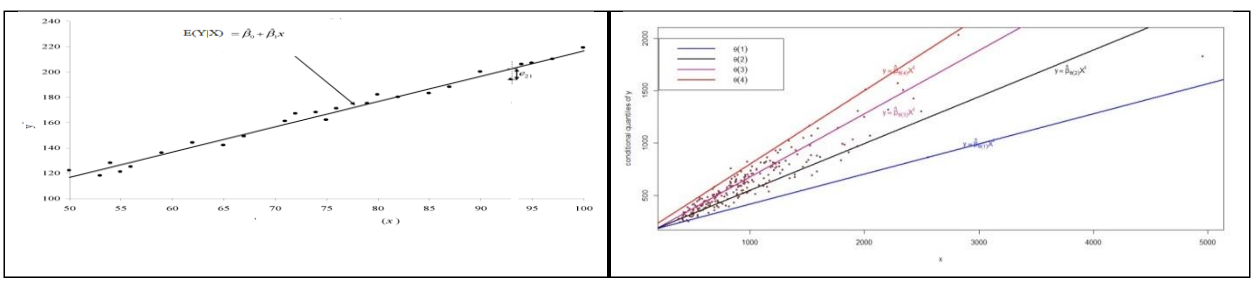

Figure 1: regression lines belong to classical and quantile regression (alhuseni ,fadel 2018)

In the right figure -1- estimation of line classical regression model, it not ability covered all data . But the left figure is estimated four quantile regression lines at four different quantile levels. The quantile regression model give us covered to majority of data. To estimating the parameters of quantile models , we can use two types of functional estimation

2-Variable selection of non- Bayesian Method

The model in equation (1), we can rewrite in a new formula by using matrix approaches

$y_i = x_i^T \beta_\tau + \epsilon_i, \qquad \tau \in (0,1),$ [2]

To estimation the parameters of above model using the following check function

$\mathrm{Eargmin}\left[\rho_\tau(y_i - x_\imath \acute{\beta})\right]$ [3]

We can rewrite the equation (3) as following

$\rho_\tau(\epsilon) = \epsilon(\tau - I), (\epsilon < 0)$

$= \begin{cases} \tau|\epsilon|, & \epsilon > 0 \\ (1-\tau)|\epsilon|, & \epsilon < 0 \end{cases}$ ………………….. …………….[4]

We can estimation the above check function by using ($arg_\beta min$)

$\hat{Q}_\tau(y|x) = argmin \sum_{i=1}^{n} \rho_\tau(\epsilon)$

Then $\epsilon = y_i - \acute{x}_i \beta_{(\tau)}$ therefor ($\hat{\beta}_{(\tau)}$) is

$$\hat{\beta}_{(\tau)} = \arg_\beta \min \sum_{i=1}^{n} \rho(y_i - \acute{x}_i \beta_{(\tau)})$$

To estimation the parameters of lasso quantile regression model via the following equation

$$\hat{\beta} = arg_\beta min \sum_{i=1}^{n} (y_i - x_\imath \hat{\beta}_\tau)^2 + \lambda \sum_{i=1}^{p} |\beta_j| \qquad (5)$$

Unfortunately, the equation (5) is not differentiable at zero, (Koenker, (2005)) shows the minimization of (2) can by a linear programming method which is proposed (Koenker and D'Orey, 1987)) .This equation existing in package –R- within function rq()

3-Variable selection of Bayesian quantile regression model

The distribution of dependent variable correspond with random error distribution in regression models. The random error that belong to quantile regression model distributed according to skew-Laplace distribution (SLD), with probability density function

$f(\epsilon_i|\mu,\sigma,\tau) = \frac{\tau(1-\tau)}{\sigma} \exp{-\rho_\tau}\left\{\left(\frac{\epsilon_i - \mu}{\sigma}\right)\right\}$ [6]

When the $\mu = 0$

$$f(\epsilon_i|\mu = 0, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp{-\rho_\tau}\left\{\left(\frac{\epsilon_i}{\sigma}\right)\right\} \quad [7]$$

The equation (7) is skew-Laplace distribution (SLD) with mean and variance as follows

$$\mathrm{E}(\epsilon) = \frac{\sigma(1-2\tau)}{\tau(1-\tau)} \qquad , \nu(\epsilon) = \frac{\sigma^2(1-2\tau+2\tau^2)}{\tau^2(1-\tau)^2}$$

More detail see (Yu and Zhang,2005). Yu and Moyeed, (2001) mentioned using skew-Laplace distribution (SLD) for random error distribution with Bayesian quantile regression . But in this paper ,we will used Kozumi and Kobayashi (2011)Transformation for skew-Laplace distribution (SLD) via reformulating the SLD as a scale mixture normal distribution . According to Mixture representation for scale parameter and location parameter. The random error defined as following

$$\epsilon_i = \theta z_i + \tau\sqrt{\sigma z_i}\ m_i$$

where m_i is distributed standard normal distribution and z_i is distributed standard exponential distribution . Therefore, the model in equation (1) is equal

$$y_i = x_i^T\beta_\tau + \theta z_i + \tau\sqrt{\sigma z_i}\ m_i \quad [8]$$

The dependent variable (y_i) is distributed normal distribution with mean ($x_i^T\beta_\tau + \theta z_i$) and variance ($2\tau^2\sigma\ z_i$)

The probability density function of dependent variable y_i is

$$f(\,y_i|\beta_\tau\,, z_i) \propto \left(\prod_{i=1}^{n}(4\pi\sigma^2 z_i)^{-\frac{1}{2}}\right)\exp\left[-\frac{\sum_{i=1}^{n}(y_i - x_i^T\beta_\tau - \theta z_i)^2}{4\tau^2\sigma z_i}\right] \quad (9)$$

According to (Park and Cassella ,2008) and Kozumi and Kobayashi (2011) the variable selection via Bayesian lasso quantile regression is defined by hierarchical model as following

$y_i = x_i^T\beta_\tau + \theta z_i + \tau\sqrt{\sigma z_i}\ m_i$,

$\beta/\emptyset_j \sim \mathbb{N}(\beta_{j0}, \vartheta)$,

$\pi(\sigma^2) \sim \mathit{Inverse\ Gamma}$,

$z_i \sim \exp(\sigma^2)$,

$\emptyset_j \sim \exp(2/\lambda^2)$

where ϑ is diagonal matrix its main diagonal is equal $(\emptyset_j)$,therefore the posterior distribution of this method is defined.

1- The Conditional Distribution of y_i

$$f(y_i|\beta_\tau, z_i) = \left(\prod_{i=1}^{n}(4\pi\sigma^2 z_i)^{-\frac{1}{2}}\right)\exp\left[-\frac{\sum_{i=1}^{n}(y_i - x_i^T\beta_\tau - \theta z_i)^2}{4\tau^2\sigma z_i}\right]$$

2 − The full Conditional Posterior Distribution of β

$$f\left(\beta\middle|\sigma, z_i\right) = \frac{1}{\sqrt{2\pi\gamma}}e^{-\frac{(\beta-\delta)^2}{2\gamma}}$$

$$where\ \gamma^{-1} = \sum_{i=1}^{n}\frac{x_i\acute{x}}{\tau^2\sigma z_i} + \emptyset^{-1} \quad and\ \delta = \gamma\left[\sum_{i=1}^{n}(x_i(y_i - \theta z_i)/\tau^2\sigma z_i) + \emptyset^{-1}\beta_0\right]$$

3-The full conditional posterior of $\emptyset_j$

$$\propto (\emptyset_j)^{-\frac{1}{2}}exp\left[-\frac{1}{2}\left\{(B_j - B_{jo})^2\emptyset_j^{-1} + \lambda^2\emptyset_j\right\}\right]$$

$$\emptyset_j|y, \beta_\tau, \sigma, z \sim GIG\left[(B_o - B_{jo})^2, \lambda\right]$$

4-The full conditional posterior of σ is inverse gamma with shape parameters $\frac{n}{2}$+a and scale parameter $(y - x\beta - \theta z)^t(y - x\beta - \theta z)\tau^{-2}z^{-1} + b$

5- the full conditional posterior of σ Generalized inverse Gaussion

$$z_i|y, \beta_\tau, \sigma^2 \sim GIG(1/2\varphi, \delta)$$

The Bayesian estimators are equal to the arithmetic mean to conditional Posterior distributions.

4- Real Data

In this section we will focus on important variables that effecting on obesity problem ,the response variable is represented the weight by kilograms, there are set of independent variables have direct and indirect effecting on the obesity these variables is

1- X1 is the age
2- X2 is working hours
3- X3 is sleeping hours
4- X4 is Physical exercise
5- X5 is total number of meals
6- X6 is type working
7- X7 is food metabolism rate
8- X8 is dietary habits
9- X9 is Obesity the hereditary
10- X10 is drinking alcohol
11- X11 is smoking
12- X12 TV watching hour
13-X13 Internet sitting hours

In this study, the sample size is equal 110 observations , we will using four quantile levels as following $\tau_1 = 0.17, \tau_2 = 0.47 , \tau_3 = 0.77\ and\ \tau_4 = 0.97$ and using the Bayesian and non-Bayesian to estimation the parameters of these models as following

$at\ \tau_1 = 0.17$ the quantile regression model is defined as following

$$y_i = \beta_{1(\tau=0.17}X1 + \beta_{2(\tau=0.17}X2 + \cdots + \beta_{13(\tau=0.17}X13 + \epsilon_i$$

$at\ \tau_2 = 0.47$ the quantile regression model is defined as following

$$y_i = \beta_{1(\tau=0.47}X1 + \beta_{2(\tau=0.47}X2 + \cdots + \beta_{13(\tau=0.47}X13 + \epsilon_i$$

$at\ \tau_3 = 0.77$ the quantile regression model is defined as following

$$y_i = \beta_{1(\tau=0.77}X1 + \beta_{2(\tau=0.77}X2 + \cdots + \beta_{13(\tau=0.77}X13 + \epsilon_i$$

$at\ \tau_4 = 0.97$ the quantile regression model is defined as following

$$y_i = \beta_{1(\tau=0.97}X1 + \beta_{2(\tau=0.97}X2 + \cdots + \beta_{13(\tau=0.97}X13 + \epsilon_i$$

Table (1) show point estimation and confidence interval for the parameters of quantile regression model at ($\tau_1 = 0.17$)

Variable	Name of variables	Non-Bayesian method	Bayesian method

s		Point Estimation	Confidence interval estimation		Point Estimation	Confidence interval estimation	
			Low	Upper		Low	Upper
X_1	the age	1.261	0.641	1.587	1.064	0.571	1.243
X_2	working hours	0.413	0.280	0.971	0.356	0.197	0.765
X_3	sleeping hours	0.401	0.168	0.623	0.297	0.132	0.693
X_4	Physical exercise	0.817	0.681	1.471	0.424	0.281	0.781
X_5	total number of meals	0.261	0.196	0.574	0.301	0.165	0.483
X_6	type working	0.000	-0.131	0.236	0.000	-0.176	0.563
X_7	food metabolism rate	0.351	0.237	0.523	0.611	0.577	0.881
X_8	dietary habits	1.052	0.876	1.353	1.341	0.852	1.502
X_9	Obesity the hereditary	0.317	0.186	0.562	0.273	0.175	0.655
X_{10}	drinking alcohol	0.000	-0.321	0.433	0.000	-0.121	0.363
X_{11}	smoking	0.051	0.028	0.361	0.246	0.124	0.442
X_{12}	TV watching hour	0.471	0.256	0.547	0.427	0.170	0.563
X_{13}	Internet sitting hours	0.262	0.292	0.558	0.242	0.182	0.582

From the result in table (1) there are independent variables have positive and negative effecting on obesity . we see the Bayesian method agree with non- Bayesian method to achieving variable selection .we see the variables (type working and drinking alcohol) are not effecting on response variable ,we can exclude these variables from our model because parameters estimation of these variables is equal zero exactly .

Table (2) show point estimation and confidence interval for the parameters of quantile regression model at ($\tau_2 = 0.47$)

Variables	Name of variables	Non-Bayesian method			Bayesian method		
		Point Estimation	Confidence interval estimation		Point Estimation	Confidence interval estimation	
			Low	Upper		Low	Upper
X_1	the age	0.160	0.150	0.387	0.465	0.370	0.853
X_2	working hours	0.503	0.180	0.970	0.336	0.097	0.763
X_3	sleeping hours	0.240	0.068	0.613	0.197	0.131	0.783
X_4	Physical exercise	0.807	0.680	1.570	0.515	0.180	0.780

X_5	total number of meals	0.160	0.096	0.375	0.300	0.063	0.583
X_6	type working	0.000	-0.030	0.036	0.000	-0.076	0.363
X_7	food metabolism rate	0.330	0.137	0.613	0.600	0.377	0.880
X_8	dietary habits	0.031	0.876	0.333	0.350	0.231	0.701
X_9	Obesity the hereditary	0.307	0.086	0.661	0.173	0.073	0.333
X_{10}	drinking alcohol	0.000	-0.210	0.433	0.000	-0.110	0.463
X_{11}	smoking	0.030	0.018	0.560	0.156	0.105	0.551
X_{12}	TV watching hour	0.570	0.236	0.757	0.517	0.170	0.763
X_{13}	Internet sitting hours	0.161	0.121	0.338	0.151	0.081	0.381

From the result in table (2) there are independent variables have positive and negative effecting on obesity . We see the Bayesian method agree with non- Bayesian method to achieving variable selection .We see the variables (type working and drinking alcohol) are not effecting on response variable ,we can exclude these variables from our model because parameters estimation of these variables is equal zero exactly . From table (2) the results of parameters estimation at second quantile level similarity of parameters estimation at first quantile level

Table (3) show point estimation and confidence interval for the parameters of quantile regression model at ($\tau_3 = 0.77$)

Variables	Name of variables	Non-Bayesian method			Bayesian method		
		Point Estimation	Confidence interval estimation		Point Estimation	Confidence interval estimation	
			Low	Upper		Low	Upper
X_1	the age	0.560	0.250	0.987	0.467	0.170	0.874
X_2	working hours	0.304	0.280	0.670	0.446	0.197	0.664
X_3	sleeping hours	0.370	0.168	0.674	0.397	0.247	0.884
X_4	Physical exercise	0.207	0.180	7.862	0.512	0.380	0.980
X_5	total number of meals	0.262	0.196	0.562	0.436	0.164	0.784
X_6	type working	0.000	-0.040	0.026	0.000	-0.176	0.264
X_7	food metabolism rate	0.440	0.747	0.674	0.600	0.477	0.880
X_8	dietary habits	0.047	0.876	0.444	0.470	0.447	0.707
X_9	Obesity the hereditary	0.407	0.086	0.567	0.154	0.074	0.315
X_{10}	drinking alcohol	0.000	-0.070	0.144	0.000	-0.170	0.064
X_{11}	smoking	0.000	-0.078	0.060	0.000	-0.104	0.174
X_{12}	TV watching hour	0.210	0.146	0.781	0.271	0.175	0.563

X_{13}	Internet sitting hours	0.167	0.047	0.248	0.263	0.187	0.487

From the result in table (3) there are independent variables have positive and negative effecting on obesity . We see the Bayesian method agree with non- Bayesian method to achieving variable selection .We see the variables (type working , drinking alcohol and smoking) are not effecting on response variable ,we can exclude these variables from our model because parameters estimation of these variables is equal zero exactly.

Table (4) show point estimation and confidence interval for the parameters of quantile regression model at ($\tau_4 = 0.97$)

Variables	Name of variables	Non-Bayesian method			Bayesian method		
		Point Estimation	Confidence interval estimation		Point Estimation	Confidence interval estimation	
			Low	Upper		Low	Upper
X_0	the age	0.620	0.467	0.972	0.656	0.423	0.372
X_2	working hours	0.792	0.572	0.892	0.370	0.233	0.482
X_3	sleeping hours	0.562	0.367	0.780	0.178	0.060	0.238
X_4	Physical exercise	0.207	0.180	0.571	0.156	0.274	0.529
X_5	total number of meals	0.019	0.002	0.140	0.075	0.030	0.270
X_6	type working	0.000	-0.078	0.372	0.000	-0.176	0.652
X_7	food metabolism rate	0.277	0.156	0.374	0.136	0.124	0.409
X_8	dietary habits	0.240	0.199	0.477	0.170	0.133	0.306
X_9	Obesity the hereditary	0.421	0.247	0.699	0.350	0.307	0.566
X_{10}	drinking alcohol	0.000	-0.342	0.185	0.000	-0.172	0.085
X_{11}	smoking	0.000	-0.109	0.180	0.000	-0.023	0.268
X_{12}	TV watching hour	0.503	0.189	0.423	0.279	0.127	0.280
X_{13}	Internet sitting hours	0.040	0.020	0.340	0.000	-0.101	0.264

From the result in table (4) there are independent variables have positive and negative effecting on obesity . We see with non-Bayesian method the variables (type working , drinking alcohol and smoking) are not effecting on response variable ,we can exclude these variables from our model because parameters estimation of these variables is equal zero exactly. But in Bayesian method the variables (type working , drinking alcohol ,smoking , Internet sitting hours) are not effecting on response variable ,we can exclude these variables from our model because parameters estimation of these variables is equal zero exactly.

5- Conclusions and Recommendations

5-1: Conclusions

There are some conclusions are

1- The values of parameters estimation not much different between Bayesian and non-Bayesian methods via different quantile levels.

2- In quantile regression model at quantile levels ($\tau_1 = 0.17\,, \tau_2 = 0.47 \; and \; \tau_3 = 0.77$ the Bayesian and non-Bayesian methods agree for exclude these variables (type working and drinking alcohol) from our model.

3- Bayesian estimation with quantile regression model at quantile levels ($\tau_4 = 0.97$) will removed these variables (type working , drinking alcohol ,smoking , Internet sitting hours) from our model. But non-Bayesian estimation with quantile regression model at quantile levels ($\tau_4 = 0.97$) remove these variables (type working , drinking alcohol ,smoking) from our model.

4- There are 11 variables have significant effecting on the obesity problem via Bayesian and non-Bayesian methods through four quantile levels.

5- The best quantile regression line at level ($\tau_4 = 0.97$) in represented the data of our study.

5-2 : Recommendations

There are some recommendations as following

1- Expanding the current study by increasing the number of independent variables affecting the quantile regression model, as well as increasing the quantile levels to greater levels and observing the process of selecting the variables.

2- Using new Bayesian lasso method which proposed by other researchers in quantile lasso regression filed to achieving coefficient's estimation and variables selection

3- Using another penalty function which have good features for achieving variable selection.

Bibliography

1- Abrevaya, J., & Dahl, C. M. (2008). The effects of birth inputs on birthweight: evidence from quantile estimation on panel data. *Journal of Business & Economic Statistics*, *26*(4), 379-397.

2- Alhuseni ,fadel (2018) "Bayesian variables selection and coefficients estimation in tobit quantile regression model " thesis doctoral university of Craiova

3- Koenker, R. and G. J. Bassett (1978). Regression quantiles. *Econometrica 46*, 33–50.

4- Kostov, P., & Davidova, S. (2013). A quantile regression analysis of the effect of farmers' attitudes and perceptions on market participation. *Journal of Agricultural Economics*, *64*(1), 112-132.

5- Koenker, R., & Geling, O. (2001). Reappraising medfly longevity: a quantile regression survival analysis. *Journal of the American Statistical Association*,*96*(454), 458-468.

6- Koenker, R. (2005). *Quantile Regression*. Cambridge Books. Cambridge University Press.

7- Koenker, R. and V. D'Orey (1987). Algorithm AS 229: Computing regression quantiles.

8- Journal of the Royal Statistical Society: Series C (Applied Statistics) 36, 383–393.

9- Kozumi, H. and Kobayashi, G. (2011). "Gibbs sampling methods for Bayesian quantile regression." Journal of statistical computation and simulation, 81(11): 1565–1578. 3

10- Park, T. and Casella, G. (2008). The Bayesian lasso. Journal of the American Statistical Association 103, 681–686.

11- Yu, K. and R. A. Moyeed (2001). Bayesian quantile regression. *Statistics* & *Probability Letters 54*, 437–447.

12- Yu, K. and Zhang, J. (2005). "A three-parameter asymmetric Laplace distribution and its extension, "*Communications in Statistics - Theory and Methods*, **34**, 1867–1879.

13- Wei, Y., Pere, A., Koenker, R., and He, X. (2006). "Quantile regression methods for reference growth charts." *Statistics in medicine*, 25(8): 1369–1382. 1