Quantile regression analysis and the estimation of baroreflex sensitivity

S Gouveia, C Rocha, AP Rocha, and ME Silva

October, 2010

Abstract Quantile regression is a well established statistical methodology for estimating conditional quantile functions in a regression setting. In comparison to classical regression, quantile regression is a more robust procedure and allows a more complete characterization of a set of distributions. This work applies classical and quantile regression to the estimation of baroreflex sensitivity (BRS), which is a clinically accepted method for the assessment of the integrity of the autonomic nervous system. The BRS estimation approaches are compared using experimental data of the EuroBaVar dataset.

Sónia Gouveia
Centro de Matemática da Universidade do Porto (CMUP) and Departamento de Matemática, Universidade de Aveiro, Portugal, e-mail: sonia.gouveia@fc.up.pt

Conceição Rocha
Departamento de Matemática, Faculdade de Ciências, Universidade do Porto and Centro I&D em Matemática e Aplicações, Universidade de Aveiro (CIDMA), Portugal, e-mail: mnrocha@fc.up.pt

Ana Paula Rocha
Departamento de Matemática, Faculdade de Ciências, Universidade do Porto and UI&D CMUP, Portugal, e-mail: aprocha@fc.up.pt

M. Eduarda Silva
Faculdade de Economia, Universidade do Porto and UI&D CIDMA, Portugal, e-mail: mesilva@fep.up.pt

1 Introduction

Many methods in applied statistics can be regarded as a regression model leading to least squares estimation methods. In the classical methodology of least squares regression, the relationship between a response variable Y and a set of regressors $\mathbf{X}$ is described solely by the conditional mean function. However, as Mosteller and Tukey [10] remark: "*just as the mean gives an incomplete picture of a single distribution, so the regression curve gives a correspondingly incomplete picture for a set of distributions*". In fact, when analyzing a single sample, measures of spread, skewness and kurtosis are employed to characterize the data beyond the mean. The quantile regression introduced by Koenker and Bassett in the seminal paper of 1978 [7] extends this notion to regression by estimating conditional quantile functions and, thus allowing the estimation of the entire distribution of the response variable conditionally on a set of regressors. The quantiles are linked to ordering and sorting the sample observations. However, just as the sample mean can be defined as the solution of the problem of minimizing a sum of squared residuals, the quantiles can be defined as the solution of the problem of minimizing a weighted sum of residuals, the solution being that of a linear programming problem. These methods were introduced in the seventies, and since then a practical statistical methodology for estimating and doing inference about conditional quantile functions has been developed. It has been used by econometricians after the nineties and is now also being used in the analysis of geophysical and climatologic data [1].

In this paper, the problem of estimating spontaneous baroreflex sensitivity (BRS) is considered. This index has been shown useful in the study of cardiac-pathological states, with lower BRS values being associated with increased morbidity and mortality [8]. It is accepted that the BRS can be quantified from the joint analysis of systolic blood pressure (SBP) and RR intervals and, using time domain methods, the BRS is estimated by the slope between SBP and RR values using least squares approach [5]. Here, the aim is to investigate whether quantile regression is able to provide new insights into BRS characterization. This paper is organized as follows: section 2 presents the basic principles of quantile regression and inferential procedures; BRS estimation steps are described in section 3, followed by the comparison between BRS estimated from classical and quantile regressions, in section 4. Finally, section 5 presents the conclusions of the study.

2 Quantile Regression

Let X be a real-valued random variable with distribution function $F(x) = P(X \leq x)$, then $F^{-1}(\tau) = \inf\{x : F(x) \geq \tau\}$ is said the τth quantile of X, $0 < \tau < 1$. Just as the mean may be seen as the solution of the problem of minimizing the expected quadratic loss function, the quantiles may be seen as the problem of minimizing the expected loss for the asymmetric linear loss function $\rho_\tau(u) = u\,(\tau - I(u < 0))$ where $I(A)$ is the indicator function of the set A. In other words, the solution of the minimization problem

$$\min_k \mathrm{E}(\rho_\tau(X-k)) = (\tau-1)\int_{-\infty}^{k}(x-k)\,dF(x) + \tau\int_{k}^{+\infty}(x-k)\,dF(x), \quad (1)$$

is $k = F^{-1}(\tau)$, the τth quantile (or an interval of τth quantiles from which the smallest element must be chosen), [6]. Function $\rho_\tau(.)$ is represented in figure 1. For $\tau = 0.5$, $\rho_\tau(.)$ is the absolute value, a symmetric linear function, and it is well known that then (1) produces the median.

Now, given a sample $X_1, \ldots, X_n$, define the empirical distribution function as $F_n(x) = n^{-1}\sum_{i=1}^{n} I(X_i \leq x)$. The empirical quantiles may be obtained by replacing $F(x)$ by $F_n(x)$ and minimizing the loss $n^{-1}\sum_{i=1}^{n}\rho_\tau(x_i - k)$, thus replacing *sorting*

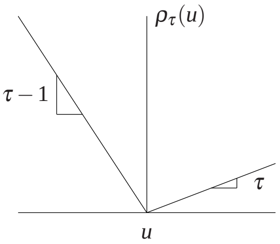


Fig. 1 Quantile regression ρ function. Figure reproduced from [6].

by *optimizing*. The optimization procedure for determining sample quantiles has the advantage of allowing the estimation of models of conditional quantile functions.

In the classical (simple linear) regression setting, the conditional mean of Y (the dependent variable) given X (the independent variable or regressor) is expressed as $\mathrm{E}[Y|X=x]=\beta x$. Then, given a sample (x_i,y_i), $i=1,\ldots,n$, β is estimated as the solution of the least squares problem $\min_{\beta\in\mathbf{R}}\sum_{i=1}^{n}(y_i-x_i\beta)^2$. It is well known that inference on the estimators is dependent on the assumptions of homoscedasticity, Gaussianity and independence. Suppose now, that instead of specifying the conditional mean of Y, one specifies the τth conditional quantile function of Y given X, $Q_y(\tau|x)=x\beta_\tau$. Then β_τ may be estimated by solving

$$\min\sum_i \rho_\tau(y_i-x_i\beta_\tau). \tag{2}$$

The quantile regression (2) may be formulated as a linear programming problem as

$$\min_{(\beta,u,v)\in\mathbf{R}\times\mathbf{R}^2}\{\tau\mathbf{1}_n^{'}u+(1-\tau)\mathbf{1}_n^{'}v|\mathbf{X}\beta+u-v=y\}, \tag{3}$$

where $\mathbf{1}_n$ represents a vector of ones and $\mathbf{X}$ represents the usual regression design matrix ($n\times 2$ in the simple regression case). The solution of this linear function on a polyhedral constraint set yields $\hat{\beta}_\tau$, which is called the τth regression quantile, with properties that follow from well-know properties of linear programming. For a detailed account of quantile regression refer to [6].

There is an extensive literature and several approaches to statistical inference (estimation and testing) for quantile regression. The most usual test regards the location-shift hypothesis of equality of slopes across quantiles. In this work, a Wald approach is used to compare different slopes, based on the joint asymptotic covariance matrix estimated by bootstrapping the (x_i, y_i) pairs [6].

3 Estimation of baroreflex sensitivity (BRS)

The first step for BRS quantification is the acquisition of the ABP and ECG signals and the extraction of the SBP and RR time series from the acquired signals (Section 3.1). Then, a BRS estimate is obtained as a slope computed from SBP and RR series, only considering the pairs of values identified in baroreflex related segments, here referred to as baroreflex events – BEs (Section 3.2).

3.1 Experimental protocol and data: EuroBaVar dataset

The EuroBaVar dataset is available for the comparison of BRS estimation procedures [9]. It consists of 46 paired records of spontaneous ECG and ABP recordings, acquired from 21 subjects in *Lying* (*L*) and *Standing* (*S*) positions. For each subject and position, the ABP and ECG signals were recorded non invasively, respectively with the use of skin electrodes and a Finapres™ finger/arm cuff device [9]. The signals were acquired in stationary conditions during 10 minutes and at a sampling frequency of 500 Hz. Each subject was first recorded in *S* position and the recording started after 5 min standing. Afterwards, followed the *L* position and the recording started after 5 min supine. In between conditions, there was a 10 minutes rest period, when the ABP finger cuff was removed and patients could speak. This dataset is also provided in beat-to-beat series, namely RR (sec) and SBP (mmHg) series extracted, respectively, from the acquired ECG and ABP signals (see figure 2). The length of these series ranges from 553 to 1218 beats and, to set comparable results for all recordings, BRS analysis was based on the first 512 beats of each file.

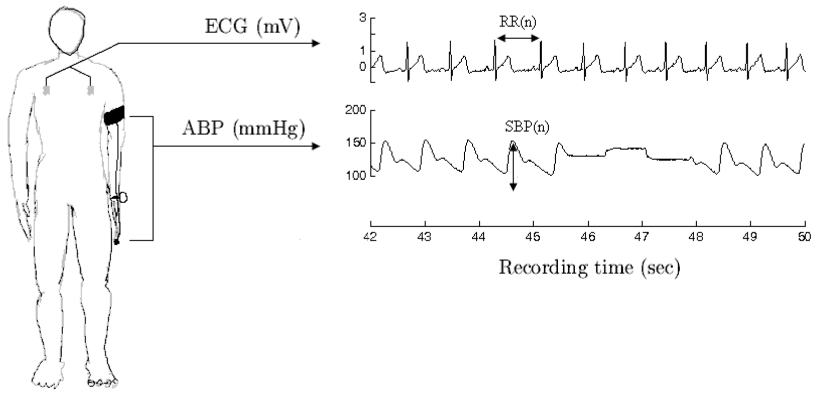


Fig. 2 Setup for ECG and ABP acquisition, showing anatomical position of the ECG electrodes and ABP finger/arm cuff for ABP acquisition. The figure also illustrates how the SBP and RR time series, used for BRS estimation, can be extracted from the acquired signals.

The EuroBaVar dataset is composed of paired *L* and *S* recordings from non-homogenous subjects. This dataset includes two subjects with autonomic dysfunction, which are expected to exhibit lower BRS estimates in comparison with those of the remaining subjects.

3.2 Identification of baroreflex events (BEs) and slope estimation

The methods for BRS estimation have been previously detailed [5]. BRS estimation is performed over SBP and RR series, here denoted $x_{\text{SBP}}(n)$ and $x_{\text{RR}}(n)$, respectively, with $n = 1, 2, ..., N_{max}$ indicating the beat number. In concordance with previous studies, these series are considered with one beat delay, i.e., $x_{\text{SBP}}(n-1)$ is paired with $x_{\text{RR}}(n)$ [2].

Each baroreflex event BE_k, $k = 1, 2, \ldots, K$ is identified as a segment with N_k pairs of values $(\mathbf{x}^k_{\text{SBP}}, \mathbf{x}^k_{\text{RR}})$ beginning at index n_k, i.e.,

$$\mathbf{x}^k_{\text{SBP}} = \left[x_{\text{SBP}}(n_k - 1)\ x_{\text{SBP}}(n_k)\ \cdots\ x_{\text{SBP}}(n_k + N_k - 2) \right]$$
$$\mathbf{x}^k_{\text{RR}} = \left[x_{\text{RR}}(n_k)\ x_{\text{RR}}(n_k + 1)\ \cdots\ x_{\text{RR}}(n_k + N_k - 1) \right],$$

that exhibit a minimum beat length ($N_k \geq 3$) and a minimum correlation between the x_{SBP} and x_{RR} values in that segment ($r_k \geq 0.8$).

After the segments identification, the mean is removed from x_{SBP} and x_{RR} values at each segment, by performing the operation $\mathbf{d}^k_\vartheta = \mathbf{x}^k_\vartheta - \vec{x}^k_\vartheta\, \mathbf{1}_{N_k},\ \vartheta \in \{\text{SBP,RR}\}$, where $\vec{x}^k_\vartheta$ represents the mean of the $\mathbf{x}^k_\vartheta$ values. The detrended values from all segments are then concatenated in $\mathbf{d}_\vartheta = \left[\mathbf{d}^1_\vartheta\ \ \mathbf{d}^2_\vartheta\ \ \ldots\ \ \mathbf{d}^K_\vartheta\right],\ \vartheta \in \{\text{SBP,RR}\}$ vectors, respectively. Finally, the BRS estimate is the slope $\hat{\beta}$ obtained from the regression analysis

$$\mathbf{d}_{\text{RR}} = \beta\, \mathbf{d}_{\text{SBP}} + c\mathbf{1}_N + \varepsilon, \tag{4}$$

where c is an unknown constant and ε is a noise vector. In this work, the usual estimate for β (obtained by ordinary least squares minimization [5]) is compared with that estimated from quantile regression (see section 2). Figure 3 illustrates the BRS estimation in two EuroBaVar records showing that, due to the data characteristics, there are cases in which the estimates $\hat{\beta}_{\text{OLS}}$ and $\hat{\beta}_{0.5}$ seem to differ.

Quantile regression provides a more complete characterization of the data than OLS regression, by simply considering other quantile values besides the median. Therefore, the baroreflex estimation was further explored in this work, concerning

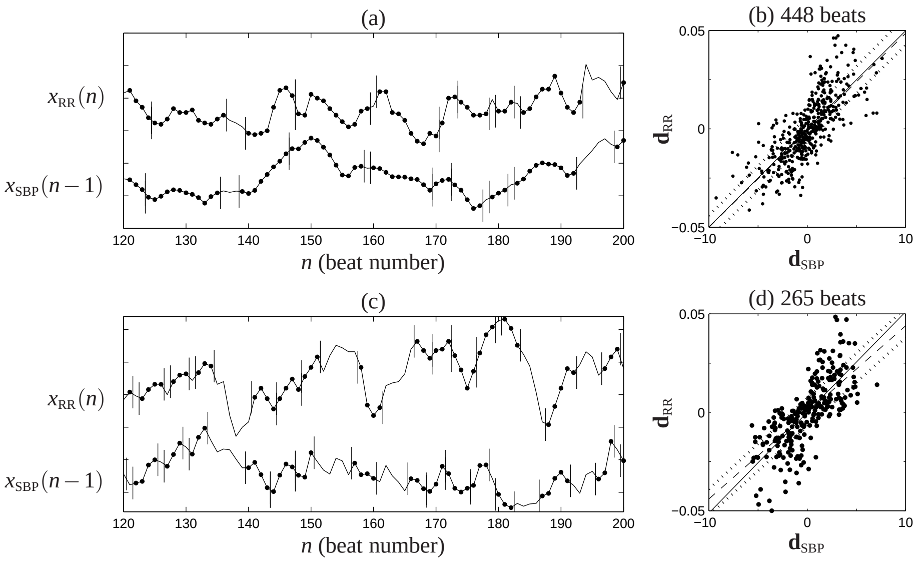


Fig. 3 BRS estimation in two EuroBaVar files "A001LB" (a,b) and "B012SB" (c,d). After the identification of BEs in the $x_{SBP}(n-1)$ and $x_{RR}(n)$ series (a,c), the dispersion diagrams are obtained for slope computation (b,d). Solid line has OLS slope $\hat{\beta}_{OLS}$, dashed line has slope estimated by quantile regression $\hat{\beta}_{0.5}$ and dotted lines have slope $\hat{\beta}_\tau$ for $\tau \in \{0.25, 0.75\}$.

the behavior of the tails of the data distribution. Figures 3(b,d) also show the lines with slopes obtained for other quantiles besides the median.

4 Results

For each record, the comparison between different slopes was performed by means of Wald test, where $\sigma_{\hat{\beta}}$ was estimated via Bootstrap simulation [4]. For each record, 1000 bootstrap replicas of the same length as the original set were generated by re-sampling with replacement the original $\mathbf{d}_{SBP}$ and $\mathbf{d}_{RR}$ pairs of values. This procedure allowed to keep the heteroscedasticity pattern in the data (see figures 3(b,d)) and, consequently, to obtain an adequate $\hat{\sigma}_{\hat{\beta}}$. The slopes comparisons indicate that only 7 out of the 46 records exhibit significant differences between β_{OLS} and $\beta_{0.50}$. More-over, significant differences between pairs ($\beta_{0.25}$,$\beta_{0.75}$) and ($\beta_{0.05}$,$\beta_{0.95}$) were found in 3 and 2 records out of 46, respectively. These results indicate that, although the data exhibits a heteroscedasticity pattern, its dispersion around the OLS/median line is fairly symmetric (see figures 3(b,d)). As a consequence, in only 4 and 6 records out of the 46, the equalities $\beta_{0.25}$=$\beta_{0.50}$=$\beta_{0.75}$ and $\beta_{0.05}$=$\beta_{0.25}$=$\beta_{0.50}$=$\beta_{0.75}$=$\beta_{0.95}$ were rejected, at 5% significance level.

The $\hat{\beta}$ obtained for all records are represented in figure 4(a), and illustrates the similarity between the distributions of the different slopes and the high inter-subject dispersion between the $\hat{\beta}$. The latter is in accordance with the fact that the EuroBaVar records were collected from heterogeneous subjects and, therefore, it was expected to include a wide range of $\hat{\beta}$ values. Because of this, the intra-subject dispersion was quantified from the coefficient of variation $\delta_{\hat{\beta}}$, which measures the dispersion $\hat{\sigma}_{\hat{\beta}}$ as a percentage of $\hat{\beta}$. In 36/46 records it was found the relation $\delta_{\hat{\beta}_{0.50}} > \delta_{\hat{\beta}_{\mathrm{OLS}}}$. Nevertheless, as illustrated in figure 4(b), $\delta_{\hat{\beta}_{\mathrm{OLS}}}$ and $\delta_{\hat{\beta}_{0.5}}$ exhibit similar distributions, with around 75% of the records presenting both $\delta_{\hat{\beta}_{\mathrm{OLS}}}$ and $\delta_{\hat{\beta}_{0.5}}$ below 10% of the corresponding $\hat{\beta}$ values. Finally, the $\delta_{\hat{\beta}_{\tau}}$ evaluated for $\tau \in \{0.05, 0.95\}$ are higher than those evaluated for $\tau \in \{0.25, 0.5, 0.75\}$.

Subjects with autonomic dysfunction are expected to present lower BRS estimates in comparison to those of normal subjects [9]. For the EuroBaVar subjects with autonomic dysfunction (open circles in figure 4), the $\hat{\beta}$ values are lower than the 5th percentile of the $\hat{\beta}$ distribution for the remaining subjects. The records with the lowest $\hat{\beta}$ exhibit similar δ values in comparison with that of the remaining.

For the discrimination between L and S positions, it is expected that the L to S ratio of $\hat{\beta}$ ($\hat{R}_{LS}$) is above 1 [9]. As shown in figure 5, there is strong evidence that both mean and median of $\hat{R}_{LS}$ are above 1 for all approaches, being approximately twice greater in L than in S. The $\hat{\beta}_{0.05}$ and $\hat{\beta}_{0.95}$ values are also able to distinguish the

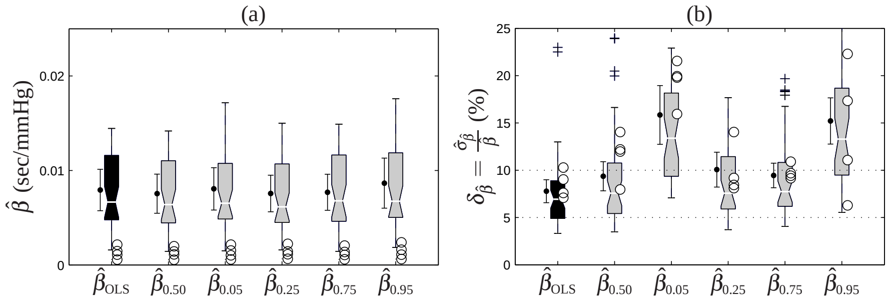
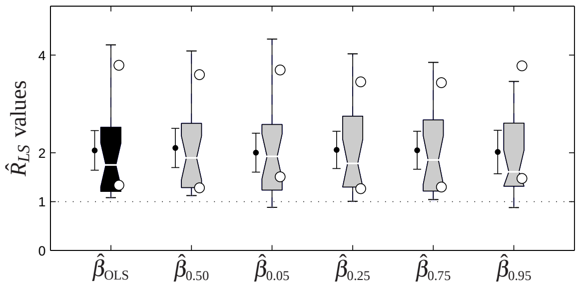


Fig. 4 Boxplots of (a) $\hat{\beta}$ and (b) $\delta_{\hat{\beta}}$ values evaluated for all 46 EuroBaVar files. Median and mean 95% confidence intervals represented by the notch and by the interval displayed at the left of each boxplot. The circles localize the 4 paired files from the 2 subjects with autonomic dysfunction.

different positions for almost all subjects, although exhibiting the largest dispersions when comparing the different slope approaches (see figure 4(b)).

Figure 5 also highlights that it is not possible to differentiate the dysfunction cases from the remaining. The location of $\hat{R}_{LS}$ for these files in separate tails of the overall distribution could be explained by the fact that the ratio of two small $\hat{\beta}$ values is more sensitive to a small variation in one of the values. Another explanation could be the different origins of the baroreflex failure (one diabetic with cardiac neuropathy and another after heart transplantation). This work suggests that clinical interpretation studies facing pathological/control cases should be carried out in order to further investigate this behavior.

Fig. 5 Boxplots of the ratio between the $\hat{\beta}$ obtained in L and S positions ($\hat{R}_{LS}$), for the 46 paired recordings of the EuroBaVar dataset. Same graphical display as in figure 4.



5 Conclusions

In this work, quantile regression (QR) is considered for estimating baroreflex sensitivity (BRS). The results from experimental data indicate that OLS slope and QR slope at quantile 0.5 do not exhibit significant differences. In spite of QR having the advantage over OLS to provide a slope for any quantile, the EuroBaVar slopes at other quantiles besides 0.5 do not provide different information.

In BRS analysis, occasional very large errors can occur (e.g., in nonstationary records). Because QR estimation is based on robust measures of location (quantiles) [6], it is expected to outperform OLS estimation in BRS assessment.

Acknowledgements This work was supported by the European Regional Development Fund (FEDER) through the programme COMPETE and by the Portuguese Government through the FCT, Fundação para a Ciência e a Tecnologia, under the projects PEst-C/MAT/UI0144/2011 (CMUP, Porto, www.fc.up.pt/cmup/) and PEst-C/MAT/UI4106/2011 (CIDMA, Aveiro, http://ma.mat.ua.pt). Sónia Gouveia also acknowledges the Postdoctoral grant by CMUP. Conceição Rocha also acknowledges the grant SFRH/BD/61781/2009 by FCT/ESF.

References

1. Barbosa, S.M.: Quantile trends in Baltic sea-level. Geophys. Res. Lett. **35**, L22705 (2008).
2. Bertineri, G., Di Rienzo, M., Cavallazzi, A.: Evaluation of baroreceptor reflex by blood pressure monitoring in unanesthetized cats. Am. J. Physiol. **254**, H377-H383 (1988).
3. Di Rienzo, M., Bertineri, G., Mancia, G., Pedotti, A.: A new method for evaluating the baroreflex role by a joint pattern. Analysis of pulse interval and systolic blood pressure series. Med. Biol. Eng. Comput. **23**, 313-314 (1985).
4. Efron B., Tibshirani, R.J.: An introduction to the Bootstrap. Chapman & Hall (1993).
5. Gouveia, S., Rocha, A.P., Laguna, P., Lago, P.: Time domain baroreflex sensitivity assessment by joint analysis of spontaneous SBP and RR series. Biomed. Signal. Process. Control **4**(3), 254-261 (2009).

6. Koenker, R.: Quantile Regression. Cambridge University Press (2005).
7. Koenker, R., Basset, G.: Quantile Regression. Econometrica. **46**(1), 33-50 (1978).
8. LaRovere, M.T., Pinna, G.D., Raczak, G.: Baroreflex sensitivity: measurement and clinical implications. Ann. Noninvasive Electrocardiol. **13**, 191-207 (2008).
9. Laude, D. *et al*: Comparison of various techniques used to estimated spontaneous baroreflex sensivity. (the EuroBaVar study). Am. J. Physiol. Regul. Integr. Comp. Physiol. **286**, R226-R231 (2004).
10. Mosteller, F., Tukey, J.: Data analysis and regression: A second course in Statistics. Addison-Wesley (1977).