Time Domain BRS Estimation: Least Squares versus Quantile Regression

S Gouveia[1,2], C Rocha[3,4], AP Rocha[3,1], ME Silva[3,4]

[1] Centro de Matemática da Universidade do Porto (CMUP), Portugal
[2] Departamento de Matemática da Universidade de Aveiro, Portugal
[3] Departamento de Matemática, Faculdade de Ciências, Universidade do Porto, Portugal
[4] Centro de I&D em Matemática e Aplicações, Universidade de Aveiro (CIDMA), Portugal

Abstract

The BRS can be quantified as the slope between SBP and RR values identified in baroreflex events, estimated by ordinary least squares (OLS) minimization. Quantile regression (QR) is a more robust procedure than OLS and allows a more complete characterization of the data, by estimating conditional functions for different quantiles of interest. In this work, OLS and QR for BRS estimation are compared regarding slope estimates and dispersion.

The EuroBaVar results indicate that OLS slope and QR slopes at different quantiles do not exhibit significant differences. Also, OLS and QR slopes require similar number of beats to achieve a given BRS precision in stationary recordings. Finally, BRS estimated with OLS exhibit relative dispersion lower than 10% and 5% when computed from stationary recordings of approximately 3 and 9 minutes length, respectively.

1. Introduction

Lower baroreflex sensitivity (BRS) estimates are associated with increased morbidity and mortality [1]. Time domain methods quantify spontaneous BRS as a slope between systolic blood pressure (SBP) and RR interval values, estimated by ordinary least squares (OLS) minimization [2]. Using OLS, the relationship between a response variable Y and a set of regressors X is described solely by the conditional mean function. The quantile regression (QR) extends this description by estimating the entire distribution of Y conditionally on X, using conditional quantile functions [3]. Because QR estimation is based on robust measures of location (quantiles), it is expected to outperform OLS estimation in terms of robustness. Additionally, QR provides a more complete characterization of the data than OLS regression, by simply considering other quantiles of interest besides the median. In this work, BRS estimation from OLS and QR approaches are compared regarding the BRS estimates and their precision. Additionally, indicative recording lengths are provided to achieve 10% and 5% precision on time domain BRS estimates.

2. Methods for BRS estimation

BRS estimation from the events technique is based on the identification of baroreflex events followed by the computation of a slope [2]. BRS is estimated from SBP and RR series assuming one beat delay, i.e., $x_{\text{SBP}}(n-1)$ paired with $x_{\text{RR}}(n)$ where n indicates the beat number.

2.1. Identification of baroreflex events (BE)

As illustrated in Fig. 1(a), each baroreflex event BE_k, $k = 1, 2, \ldots, K$ is identified as a segment with N_k pairs of values $(\mathbf{x}^k_{\text{SBP}}, \mathbf{x}^k_{\text{RR}})$ 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 BE identification, the mean is subtracted from x_{SBP} and x_{RR} values at each segment k

$$\mathbf{d}^k_\vartheta = \mathbf{x}^k_\vartheta - \bar{\mathbf{x}}^k_\vartheta \, \mathbf{1}_{N_k}, \; \vartheta \in \{\text{SBP,RR}\}, \tag{1}$$

where $\bar{\mathbf{x}}^k_\vartheta$ represents the mean of the $\mathbf{x}^k_\vartheta$ values in the segment. The detrended SBP and RR values from all segments are then concatenated in vectors

$$\mathbf{d}_\vartheta = \left[\mathbf{d}^1_\vartheta \; \mathbf{d}^2_\vartheta \; \ldots \; \mathbf{d}^K_\vartheta\right], \; \vartheta \in \{\text{SBP,RR}\} \tag{2}$$

and, finally, the BRS estimate is taken as the OLS slope $\hat{\beta}$ obtained considering the linear model

$$\mathbf{d}_{\text{RR}} = \beta \, \mathbf{d}_{\text{SBP}} + c\mathbf{1}_N + \boldsymbol{\epsilon}, \tag{3}$$

where c is an unknown constant and $\boldsymbol{\epsilon}$ is a noise vector [2]. In this work, β estimated from QR was also considered [3]. Figure 1(b) suggests that, due to the heteroscedastic pattern observed in the data, there might be cases for which $\hat{\beta}_{\text{OLS}}$

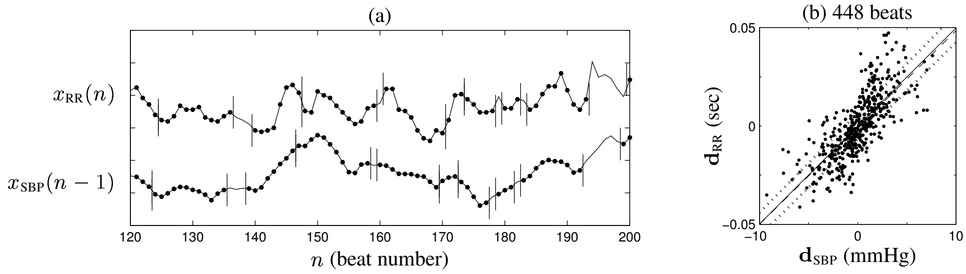

Figure 1. BRS estimation in EuroBaVar file "A001LB". After the identification of BEs (a), SBP and RR dispersion diagrams are obtained for slope computation (b). Solid line has OLS slope $\hat{\beta}_{OLS}$, dashed line has slope estimated by quantile regression $\hat{\beta}_{0.5}$ and dotted lines have slope $\hat{\beta}_\tau$ for $\tau \in \{0.25, 0.75\}$.

and $\hat{\beta}_{0.5}$ differ substantially. Also, the lines with slopes for quantiles 0.25 and 0.75 illustrate how QR can provide a more complete characterization of the data than OLS.

2.2. Estimation of the regression slope

Lets consider the linear relationship

$$Y = \beta X + c\mathbf{1}_N + \boldsymbol{\epsilon}, \tag{4}$$

where β and c are the parameters of the model and $\boldsymbol{\epsilon}$ is a vector of errors. In OLS regression, the conditional mean of Y given X is expressed as $\mathrm{E}[Y|X = x] = \beta x + c$. Then, for a sample (x_i, y_i), $i = 1, \ldots, N$, the parameters are estimated as the solution of the least squares problem

$$\min_{(\beta,c)\in\mathbf{R}^2} \sum_{i=1}^{N} (y_i - (\beta x_i + c))^2. \tag{5}$$

Instead of considering the conditional mean function, QR specifies the τth conditional quantile function of Y given X, i.e. $Q_\tau(Y|X = x) = \beta_\tau x + c_\tau$ with $0 < \tau < 1$. The parameters β_τ and c_τ may be estimated by solving

$$\min_{(\beta_\tau,c_\tau)\in\mathbf{R}^2} \sum_{i=1}^{N} \rho_\tau(y_i - (\beta_\tau x_i + c_\tau)), \tag{6}$$

where

$$\rho_\tau(\epsilon_i) = \begin{cases} -(1-\tau)\,\epsilon_i, & \epsilon_i < 0 \\ \tau\,\epsilon_i, & \epsilon_i \geq 0 \end{cases} \tag{7}$$

is a linear loss function that weights the model residuals depending on their sign (Fig. 2(a)). Following [3], the minimization procedure in Equation (6) can be reformulated as the linear programming problem

$$\begin{array}{ll} \displaystyle\min_{(\beta_\tau,c_\tau)\in\mathbf{R}^2} & \tau\mathbf{1}_N' U + (1-\tau)\mathbf{1}_N' V \\ \text{subject to} & Y = \beta_\tau X + c_\tau \mathbf{1}_N + U - V, \end{array} \tag{8}$$

where the error vector $\boldsymbol{\epsilon}$ is split into vectors U and V with elements containing respectively the positive and negative parts of the residuals ϵ_i for $i = 1, \ldots, N$, i.e.

$$u_i = \begin{cases} \epsilon_i, & \epsilon_i \geq 0 \\ 0, & \epsilon_i < 0 \end{cases} \text{ and } v_i = \begin{cases} 0, & \epsilon_i \geq 0 \\ -\epsilon_i, & \epsilon_i < 0 \end{cases}.$$

The linear programming problem in Equation (8) can be efficiently solved using simplex-type algorithms leading to the β_τ and c_τ estimates [3]. As illustrated in Fig. 2(b), the regression line with estimated parameters $\hat{\beta}_\tau$ and $\hat{c}_\tau$ divides the data in two sets, with the set below the line containing $\tau \times 100$ % of the data. The setting of $\tau = 0.5$ leads to the estimation of the conditional median function and of the slope $\hat{\beta}_{0.5}$, which is comparable to $\hat{\beta}_{OLS}$.

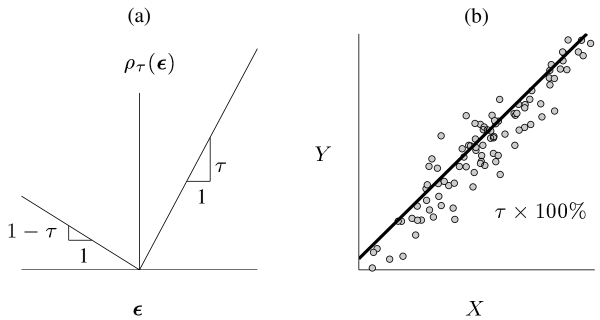

Figure 2. QR for a given τ: (a) linear loss function ρ_τ and (b) dispersion diagram superimposing line with slope β_τ. The set below the line contains $\tau \times 100\%$ of the data.

2.3. Estimation of the slope dispersion

The estimation of the slope dispersion $\sigma_{\hat{\beta}}$ is needed to compare the different approaches for slope estimation. On one hand, the equality of slopes for each record was tested with Wald statistical test, which makes use of the slope dispersion [3]. On the other hand, as heterogenous subjects are expected to exhibit different BRS estimates [4], the dispersions were compared using the coefficient of variation

$$\delta_{\hat{\beta}} = \hat{\sigma}_{\hat{\beta}}/\hat{\beta} \times 100 \ (\%). \tag{9}$$

Because the data exhibit heteroscedasticity (Fig. 1(b)), the joint asymptotic covariance matrix was estimated by bootstrap [5]. For each record, B bootstrap replicas of the same length as the original set were generated by resampling with replacement the original $\mathbf{d}_{SBP}$ and $\mathbf{d}_{RR}$ pairs. A slope estimate was then computed for each replica and the slope dispersion of each record $\hat{\sigma}_{\hat{\beta}}$ was estimated as the standard deviation of the bootstrapped slopes (Fig. 3(a)). In this work, B=1000 since $\delta_{\hat{\beta}}$ tends to stabilize for B>1000 for all recordings (Fig. 3(b)).

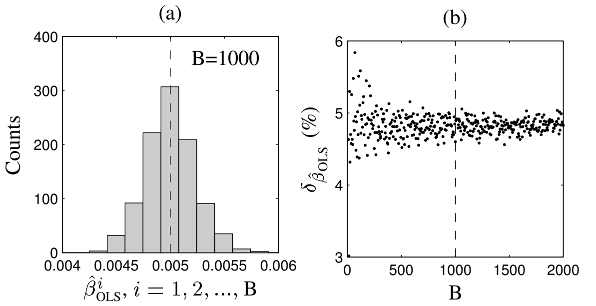

Figure 3. Bootstrap computation of δ for EuroBaVar file "A001LB": (a) histogram of bootstrapped $\hat{\beta}$ and (b) δ as a function of B$\in \{20, 25, ..., 2000\}$.

3. Results and Discussion

The OLS and QR estimation approches were compared using the 46 ABP and ECG recordings of the EuroBaVar dataset [4]. These data was collected from non-homogenous subjects, including two subjects with autonomic dysfunction that are expected to have the lowest BRS estimates [4]. Each subject was monitored in Lying (*L*) and Standing (*S*) positions. The signals were recorded non invasively, in stationary conditions during 10 minutes and at a sampling frequency of 500 Hz. The lengths of the corresponding SBP and RR series range from 553 to 1218 beats and, to set comparable results, BRS estimation was based on the first 512 beats of each recording.

Wald statistical testing indicates that only 7/46 recordings exhibit significant differences between β_{OLS} and $\beta_{0.50}$ at 5% level. Significant differences between pairs ($\beta_{0.25}$,$\beta_{0.75}$) and ($\beta_{0.05}$,$\beta_{0.95}$) were found in 3/46 and 2/46 recordings, respectively. Finally, in only 4/46 and 6/46 records, 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. These results indicate that for most of the EuroBaVar files the dispersion of the data around the OLS/median line is fairly symmetric (Figs. 1(b,d)).

Figure 4(a) illustrates the similarity between the distributions of the different slopes, which is in accordance with Wald testing results. The high inter-subject dispersion of $\hat{\beta}$ was expected as the EuroBaVar records were collected from heterogeneous subjects. The results concerning the slope dispersion are displayed in Fig. 4(b). The central slopes $\hat{\beta}_{OLS}$ and $\hat{\beta}_{0.5}$ exhibit similar dispersions, with around 75% of the records presenting δ below 10% of the corresponding $\hat{\beta}$ values. The dispersion of a quantile estimate reflects the density of observations near the quantile of interest. Therefore, the slope estimates for $\tau \in \{0.05, 0.95\}$ exhibit the highest variances, due to the low density of observations below/above these quantiles. Although exhibiting lower precision, the Wald testing indicated that only in 6/46 recordings there are significant differences between slopes across different quantiles. Finally, the subjects with autonomic dysfunction (i.e., with lower BRS estimates) exhibit similar δ in comparison with that of the remaining subjects (open circles in Fig. 4).

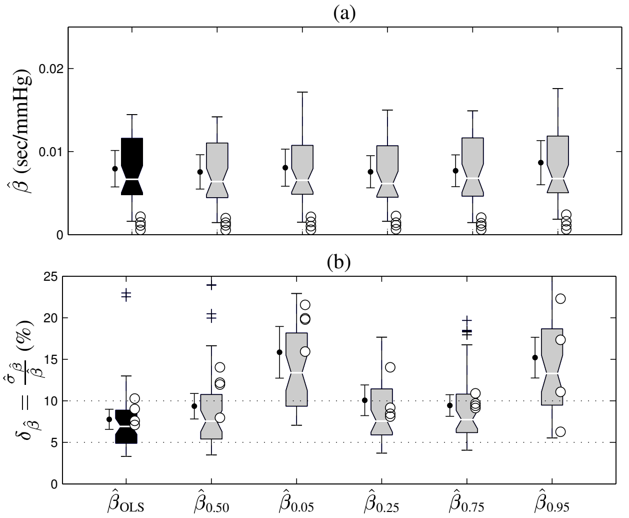

Figure 4. Boxplots of (a) $\hat{\beta}$ and (b) $\delta_{\hat{\beta}}$ obtained for all 46 EuroBaVar files, exhibiting median and mean 95% confidence intervals. The circles localize the 4 paired files from the 2 subjects with autonomic dysfunction.

Figure 5 shows that δ decreases with increasing N and r values. In order to set δ as a linear function of N and r rescalling using log function should be used. Multivariate regression analysis provided the following models

$$\log(\delta) = \begin{cases} 6.66 - 0.50 \log(N) - 2.66\, r \ \ \text{(OLS)} \\ 6.86 - 0.71 \log(N) - 1.15\, r \ \ \text{(QR)}, \end{cases} \qquad (10)$$

with parameters estimated by OLS regression. Statistical inference could not be carried out as Lilliefors and Jarque-Bera test rejected the hypothesis that the residuals have a normal distribution (5%). Alternatively, it was observed that the percentage of δ variance around its mean value is mostly due to $\log(N)$ variable (Table 1). Comparing approaches, the percentage of δ variance due to r is higher in OLS than in QR. However, r is a significative variable in both models (10), since the 95% confidence intervals over the corresponding coefficients do not include the value 0.

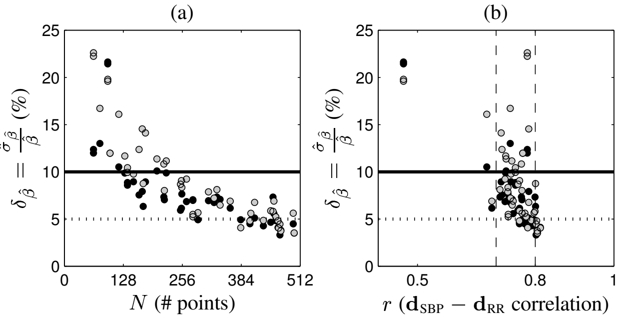

Figure 5. Plot of δ evaluated for $\hat\beta_{OLS}$ (black) and $\hat\beta_{0.5}$ (grey) as a function of N and r, considering all EuroBaVar files.

Table 1. Proportion of δ variability explained by different linear models (coefficient of determination).

	Variables in the model		
Regression	$\log(N)$ and r	$\log(N)$	r
OLS	0.91	0.76	0.52
QR	0.86	0.84	0.28

Increasing N requires the acquisition of longer recordings while increasing r is obtained if higher $\mathbf{d}_{SBP}$ and $\mathbf{d}_{RR}$ correlation is observed. In spontaneous recordings, $r \leq 0.8$ due to the apriori restriction r_{min}=0.8 in BE identification [2] and typically $0.7 < r \leq 0.8$ in stationary conditions (Fig. 5(b)). Values of $r > 0.8$ can only be observed with invasive BRS stimulation [6]. Therefore, δ can only decrease in spontaneous and stationary recordings by increasing N. Equation (10) allows to obtain indicative N values to achieve a target δ value, given the r observed in the data. Figure 6 shows that OLS and QR require similar N to achieve target $\delta = \{5, 10\}$ in stationary recordings (i.e., $0.7 < r \leq 0.8$). Considering $r = 0.7$, OLS approach requires $N = \{608, 151\}$ whereas QR approach needs $N = \{524, 198\}$. These results were corroborated by recomputing $\hat\beta$ considering long enough recordings to identify the indicative N values and observing that the obtained δ is below the target δ for each case.

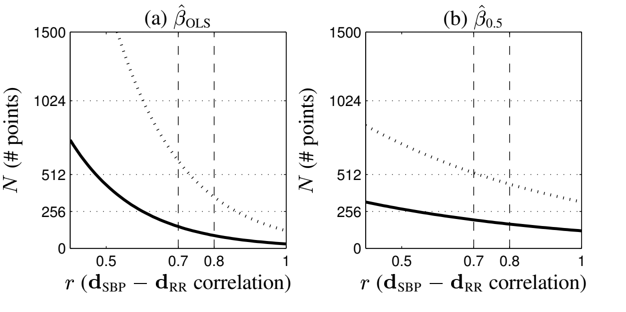

Figure 6. Indicative N values as a function of r considering the targets $\delta = 5\%$ (dotted) and $\delta = 10\%$ (full line).

The longest EuroBaVar recordings to achieve $\delta = \{5, 10\}$ were $\{8.5, 2.7\}$ minutes long for OLS and $\{7.7, 3.6\}$ minutes long for QR estimation.

4. Conclusions

In this work, OLS regression is compared to quantile regression (QR) for BRS estimation. The results from EuroBaVar 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 different quantiles do not provide different information. In stationary recordings acquired in spontaneous condition, OLS and QR approaches have shown to require a similar number of beats to achieve the same target precision. The results indicate that BRS estimates with OLS have relative dispersion lower than 10% and 5% when computed from 3 and 9 minutes long recordings, respectively.

Acknowledgements

S Gouveia acknowledges the PosDoc grant by CMUP. C Rocha acknowledges the grant SFRH/BD/61781/2009 by FCT/ESF. This work was partially supported by FCT (Portugal) through CMUP (Porto, http://www.fc.up.pt/cmup) and CIDMA (Aveiro, http://ma.mat.ua.pt).

References

[1] La Rovere MT, Pinna GD, Raczak G. Baroreflex sensitivity: Measurement and clinical implications. Ann Noninvasive Electrocardiol 2008;13(2):191–207.

[2] Gouveia S, Rocha AP, Laguna P, Lago P. Time domain baroreflex sensitivity assessment by joint analysis of spontaneous SBP and RR series. Biomed Signal Process Control 2009;4(3):254–261.

[3] Koenker R. Quantile Regression. Cambridge University Press, 2005.

[4] Laude D *et al*. Comparison of various techniques used to estimate spontaneous baroreflex sensitivity. The EuroBaVar study. Am J Physiol Regul Integr Comp Physiol 2004; 286(1):R226–R231.

[5] Efron B, Tibshirani RJ. An introduction to the Bootstrap. Chapman & Hall, 1993.

[6] Gouveia S, Rocha AP, Laguna P, Gujic M, Beloka SP, Van de Borne P, Lago P. BRS analysis from baroreflex sequences and baroreflex events compared using spontaneous and drug induced data. Comput Cardiol 2008;35:737–740.

Address for correspondence:

S´onia Gouveia
Centro de Matem´atica da Universidade do Porto
Rua do Campo Alegre, 687; 4169-007, Porto, Portugal.
E-mail address: sonia.gouveia@fc.up.pt.