Supplementary Material for: Threshold detection under a semiparametric regression model

Graciela Boente[1], Florencia Leonardi[2], Daniela Rodriguez[3] and Mariela Sued[4]

Abstract

In this Supplement, we include some additional numerical results corresponding to the simulation study reported in the main document. We also present the results of a simulation study conducted to compare our proposal with the method described in Zeileis et al. (2003) and a numerical experiment with Laplace distributed errors. Along the supplement, sections and figures are numbered S.1, S.2, ... References to equations or sections in the main body of the paper are indicated without the capital "S".

S.1 Monte Carlo Study

This section presents additional simulation results that complement the findings reported in Section 4 of the main document. Specifically, we provide detailed analyses for scenarios involving different combinations of threshold values (u_0), smoothness parameters (δ), error standard deviations (σ), and sample sizes (n). These results further illustrate the finite-sample behavior of the proposed estimators under various conditions. We also include some numerical results when the penalizing function equals $f(u) = \exp(u)$.

As mentioned in the main document, the family of regression functions considered in our numerical study, denoted as $r_{u_0,\delta} : [0,1] \to \mathbb{R}$, is labelled according to the threshold value u_0 and to the smoothing parameter $\delta \in \mathbb{R}$ and is defined as

$$r_{u_0,\delta}(x) = \begin{cases} g(x) & \text{if } 0 \le x \le u_0 \,, \\ \\ \{g'(u_0) + \delta\}(x - u_0) + g(u_0) & \text{if } u_0 \le x \le 1 \,, \end{cases}$$

where $g : [0,1] \to \mathbb{R}$ is the non-linear function given by

$$g(x) = \begin{cases} 4x^2(3 - 4x) & \text{if } 0 \le x \le 0.5 \,, \\ \\ \dfrac{4}{3}\, x\,(4x^2 - 10x + 7) - 1 & \text{if } 0.5 \le x \le 1 \,. \end{cases}$$

[1]Universidad de Buenos Aires and CONICET
[2]Universidade de São Paulo
[3]Universidad Torcuato Di Tella and CONICET
[4]Universidad de San Andrés and CONICET

The selected values of u_0 and δ to be described below lead to six continuous regression functions $r_{u_0,\delta}$ but with varying smoothness.

For completeness, we recall the scenarios considered in the numerical study reported in the main document The independent samples are such that $(X_i, Y_i) \sim (X, Y)$, $i = 1, \dots, n$, where $X \sim \mathcal{U}(0, 1)$ and $Y = r_{u_0,\delta}(X) + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, independent of X. In this way, for any value of δ, the observations satisfy the postulated model (1) with $b = g'(u_0) + \delta$, $a = g(u_0) - b\, u_0$ and threshold at u_0 as defined in (2). Sample sizes vary from $n = 100$ to $n = 2000$ with step 100. The analysis of the finite sample behavior of the threshold estimator $\widehat{u}_n$ considers the following combinations of the parameters u_0, δ, σ and for the penalty parameter:

- Threshold values: $u_0 = 0.5$ and $u_0 = 0.75$,
- Smoothness parameters: $\delta = -1, 0, 1$,
- Error standard deviations: $\sigma = 0.01, 0.05, 0.1$,
- Penalty constants:$c \in \{0.0001, 0.0005, 0.001, 0.002, 0.004, 0.008, 0.01, 0.02, 0.04\}$.

For each scenario, we performed $N_{\text{rep}} = 1000$ replications.

S.1.1 Additional numerical results when $f(u) = u$

To visualize the finite sample performance of the threshold estimates $\widehat{u}_n$, we provide parallel boxplots of the obtained estimates for the selected scenarios and for different choices of the penalizing constant c, when $n = 500$. Similar conclusions are obtained for other values of the sample size n and for that reason, we omit the plots. Figures S.1 to S.3 display the boxplots for $u_0 = 0.5$ and 0.75, $\delta = -1, 0, 1$ and $\sigma = 0.01, 0.05$ and 0.1 with varying penalty constants c. Note that we also present the results when $u_0 = 0.5$ and $\sigma = 0.01$ already shown in Figure 8 to facilitate comparisons. The true parameter is indicated with the horizontal solid red line. As mentioned in the main document, these boxplots lead to similar conclusions to obtained when considering Figure 8. Indeed, the threshold estimates are larger than the target u_0 when the penalty parameter c is small, while for large values of c the boxplots of $\widehat{u}$ lie below the true threshold. However, for the considered scenarios, the boxplot of the estimates is centered at u_0 at least for one of the selected values of c.

We also examine the estimation accuracy of the parameters α_0 and β_0 involved in the simple linear regression model defined beyond the threshold. Figures S.4 and S.5 present the boxplots of the estimators $\widehat{\alpha}_{\widehat{u}}$ and $\widehat{\beta}_{\widehat{u}}$, respectively, when $\sigma = 0.01$. The boxplots of the estimates of α_0 when $\sigma = 0.05$ and 0.10 are displayed in Figures S.6 and S.8, respectively, while those of $\widehat{\beta}_{\widehat{u}}$ are given in Figures S.7 and S.9. In each Figure, the horizontal solid red line corresponds to the true parameters α_0 and β_0, respectively. Taking into account that for values of c smaller than 0.001, the threshold estimates are mainly larger or equal than the target, the estimates of the slope and intercept accomplish the goal since their boxplots are centered around the true values. On the contrary, for large values of c some of the observations used to estimate the slope and intercept correspond to data generated using the function $g(x)$, that is, with the nonparametric model component and for that reason, they fail in their purpose.

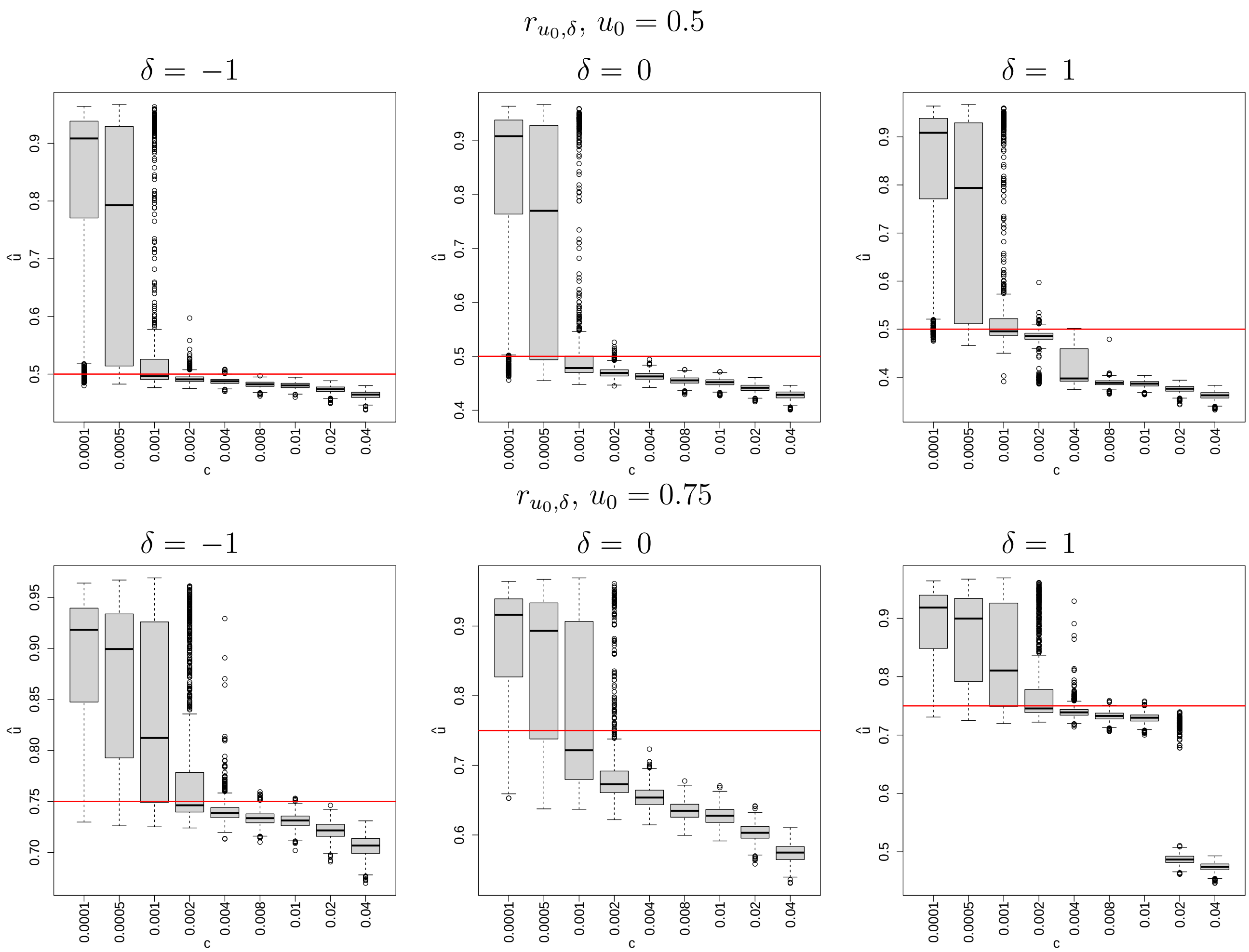

Figure S.1: Boxplots of the estimators $\widehat{u}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.01$. The horizontal red line corresponds to the true value u_0.

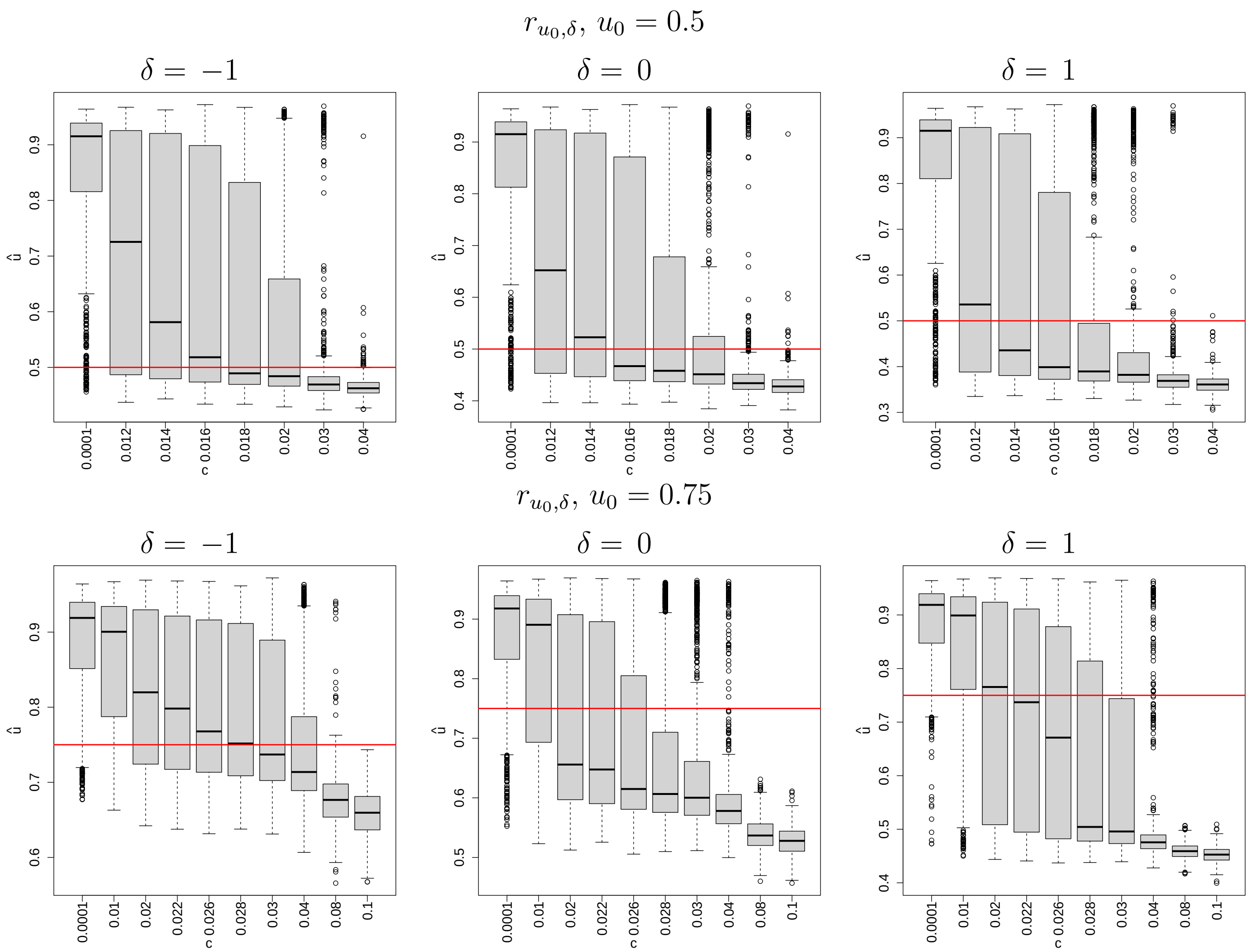

Figure S.2: Boxplots of the estimators $\widehat{u}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.05$. The horizontal red line corresponds to the true value u_0.

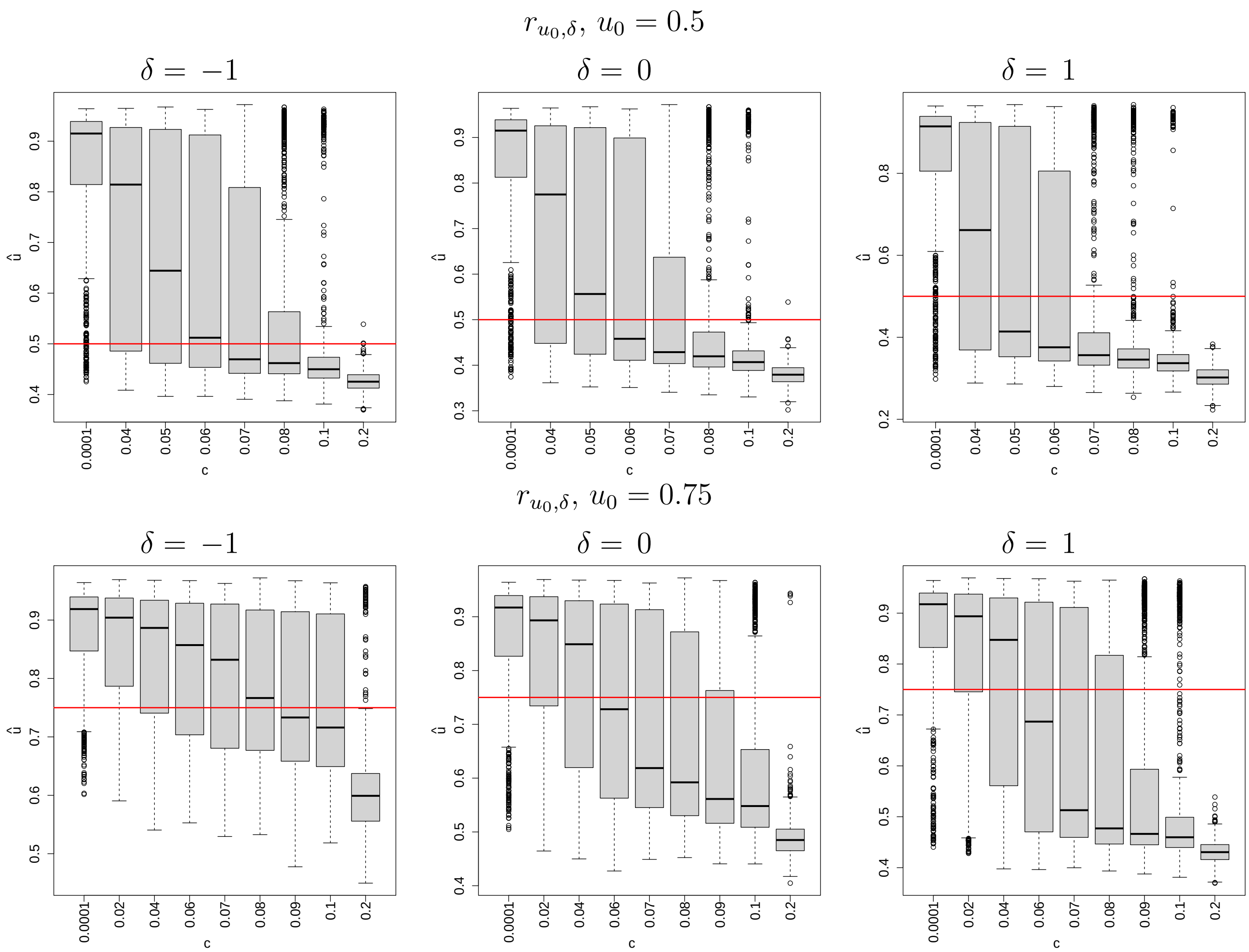

Figure S.3: Boxplots of the estimators $\widehat{u}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.10$. The horizontal red line corresponds to the true value u_0.

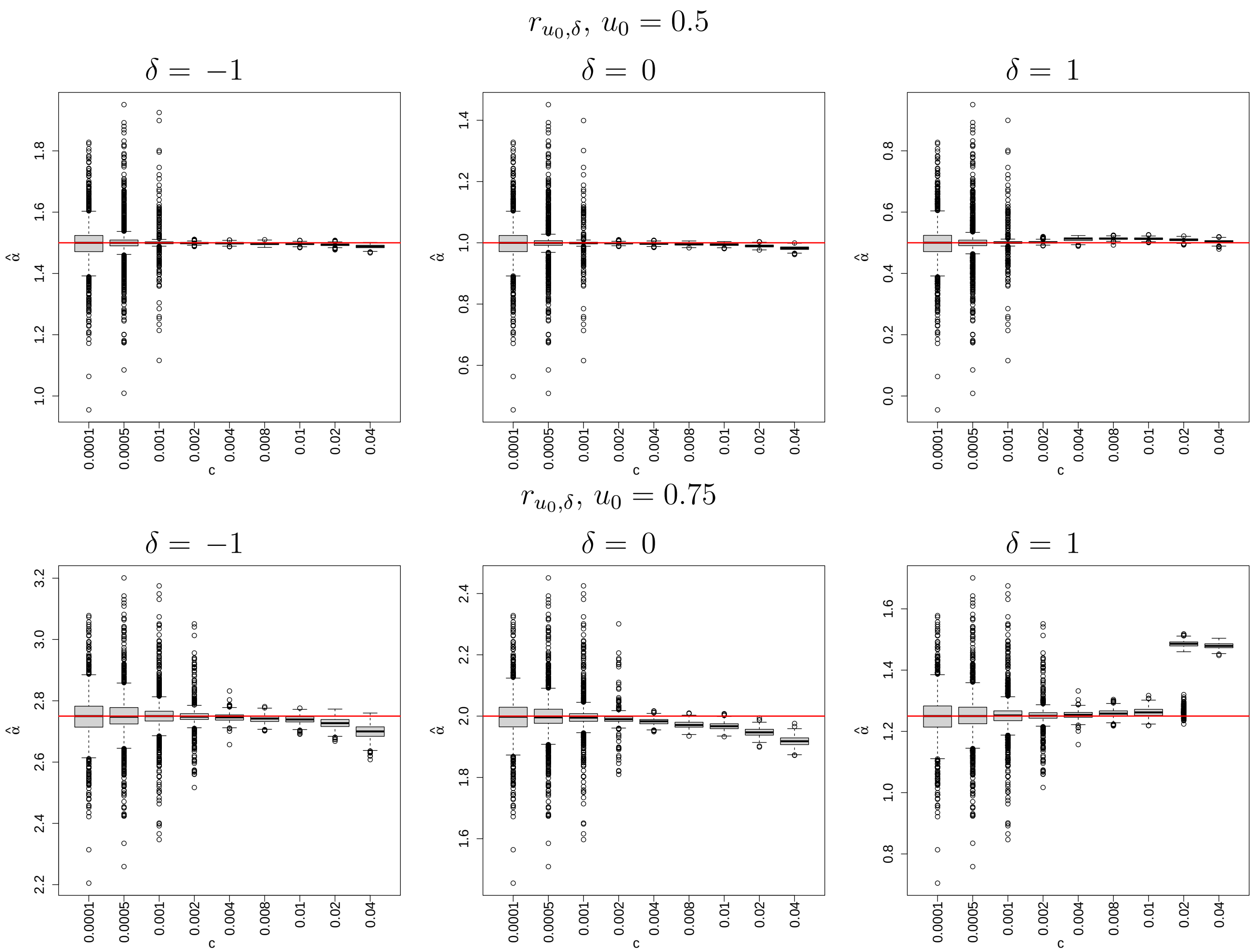

Figure S.4: Boxplots of the estimators $\widehat{\alpha}_{\widehat{u}}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.01$. The horizontal red line corresponds to the value α_0.

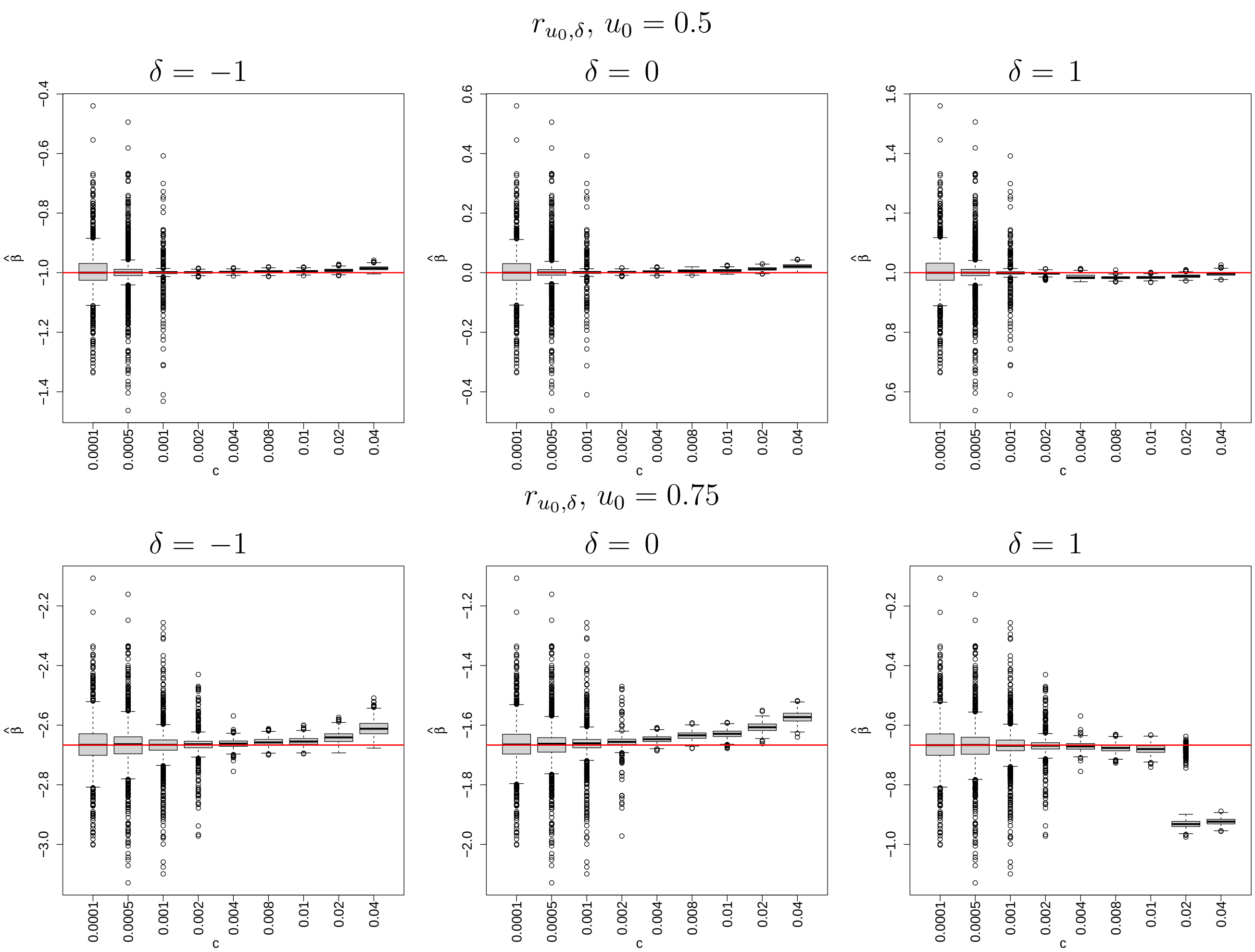

Figure S.5: Boxplots of the estimators $\widehat{\beta}_{\widehat{u}}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.01$. The horizontal red line corresponds to the value β_0.

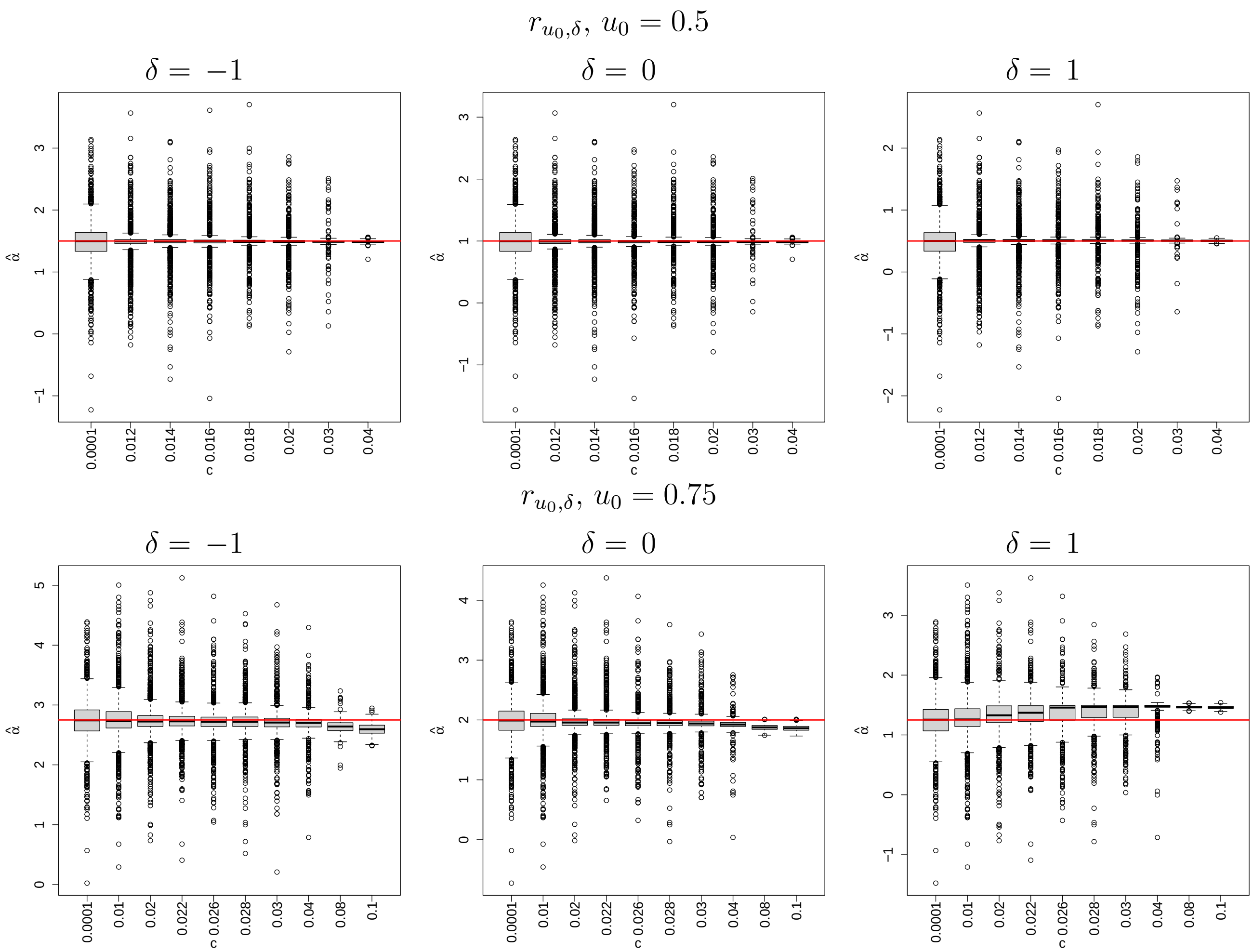

Figure S.6: Boxplots of the estimators $\widehat{\alpha}_{\widehat{u}}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.05$. The horizontal red line corresponds to the value α_0.

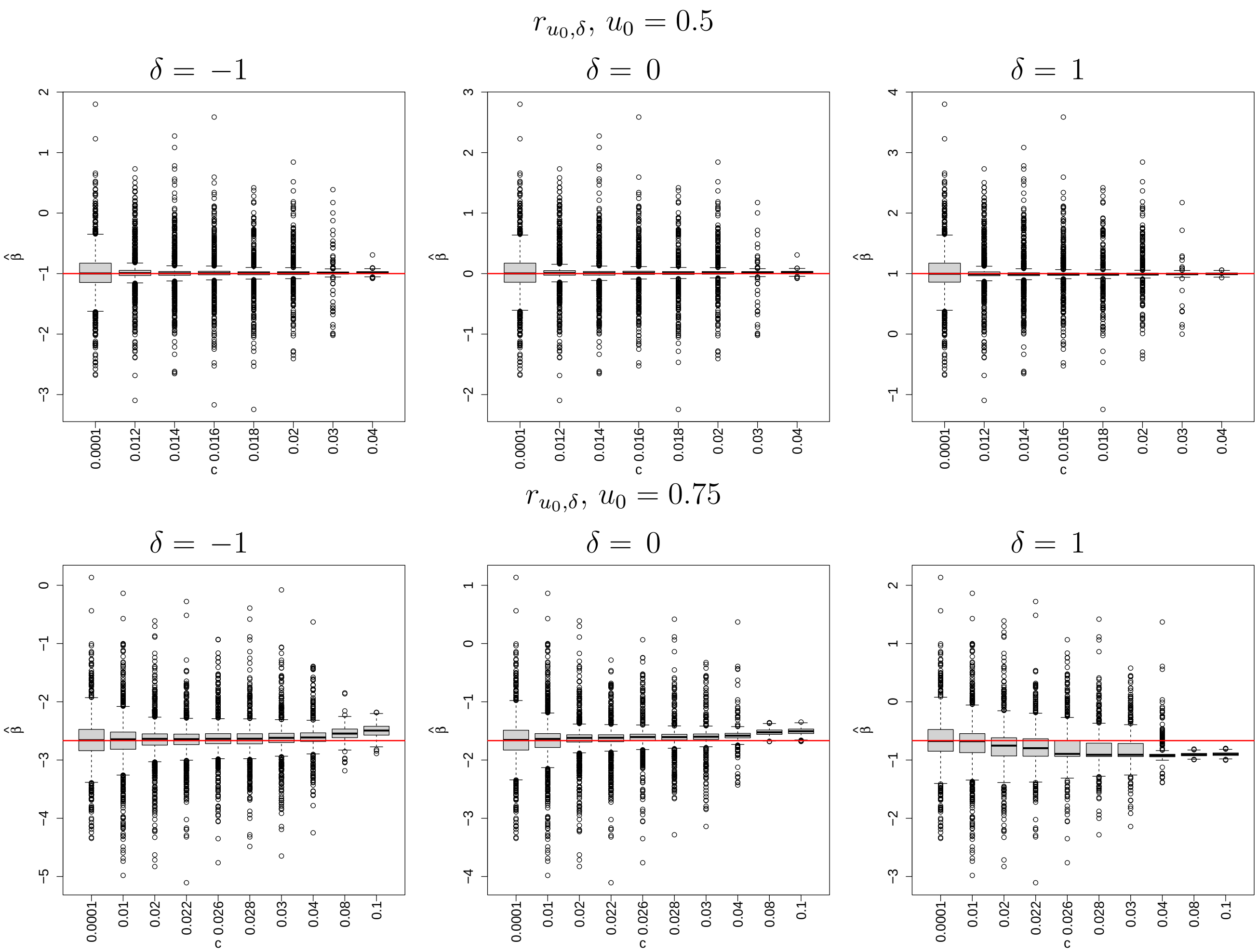

Figure S.7: Boxplots of the estimators $\widehat{\beta}_{\widehat{u}}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.05$. The horizontal red line corresponds to the value β_0.

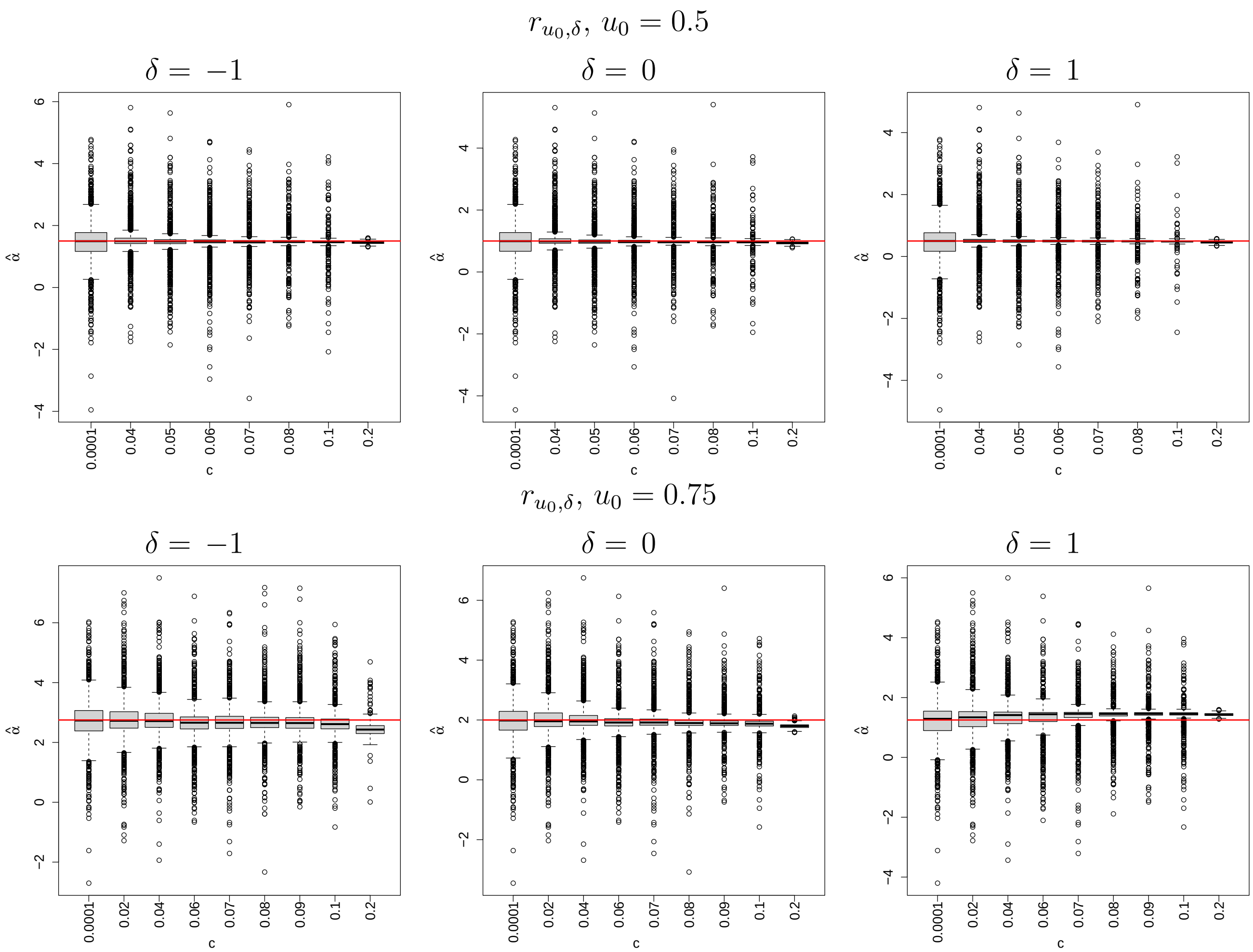

Figure S.8: Boxplots of the estimators $\widehat{\alpha}_{\widehat{u}}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.10$. The horizontal red line corresponds to the value α_0.

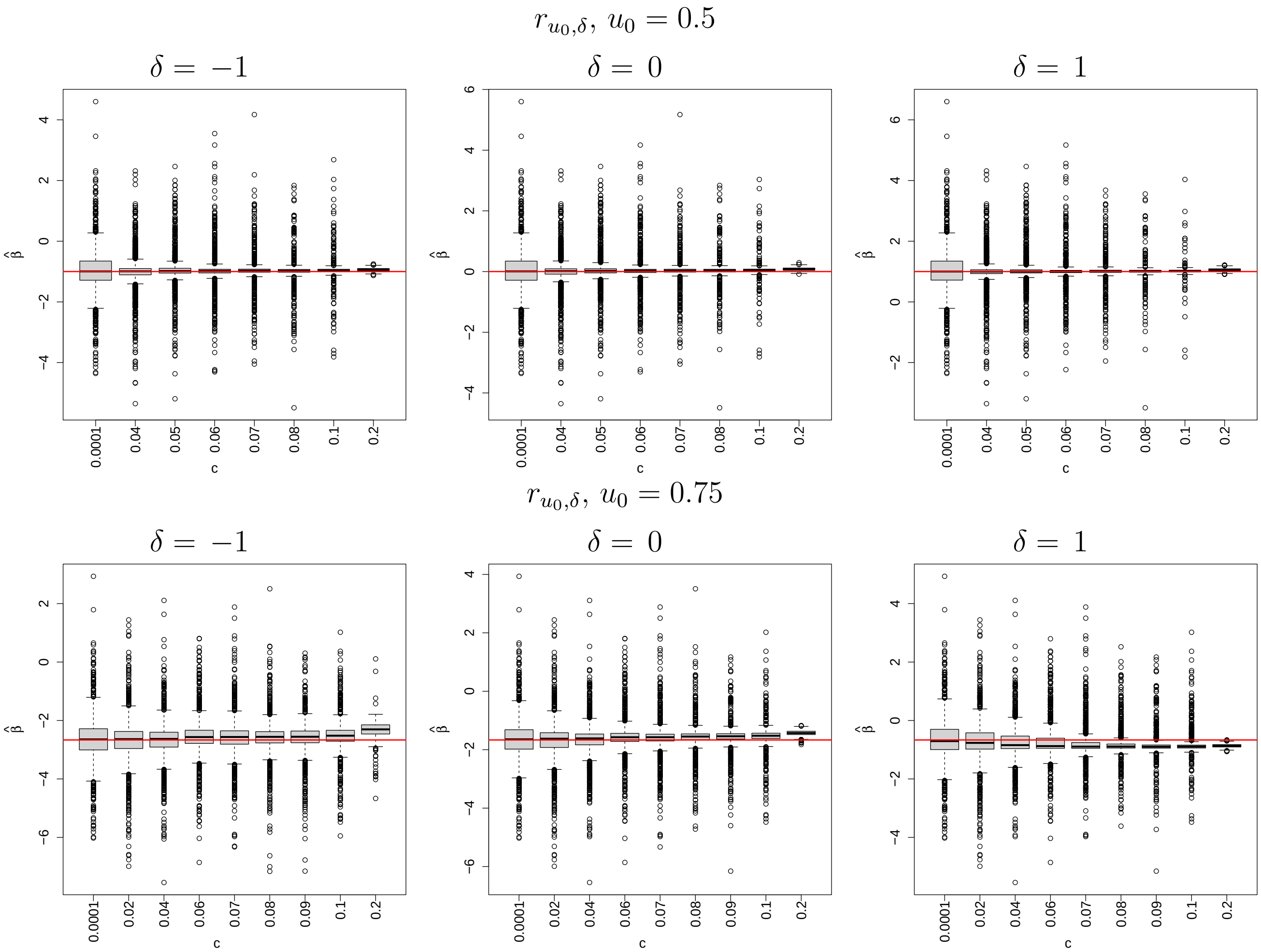

Figure S.9: Boxplots of the estimators $\widehat{\beta}_{\widehat{u}}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.10$. The horizontal red line corresponds to the value β_0.

S.1.2 Numerical results when $f(u) = \exp(u)$

As in Section S.1.1, we consider the sample size $n = 500$ and we summarize the results using the estimates boxplots for different choices of the penalizing constant c. Figures S.10 to S.12 display the boxplots for $u_0 = 0.5$ and 0.75, $\delta = -1, 0, 1$ and $\sigma = 0.01, 0.05$ and 0.1 with varying penalty constants c. The true parameter is indicated with the horizontal solid red line. Besides, Figures S.13 and S.14 present the boxplots of the estimators $\widehat{\alpha}_{\widehat{u}}$ and $\widehat{\beta}_{\widehat{u}}$, respectively, when $\sigma = 0.01$. The boxplots of the estimates of α_0 when $\sigma = 0.05$ and 0.10 are displayed in Figures S.15 and S.17, respectively, while those of $\widehat{\beta}_{\widehat{u}}$ are given in Figures S.16 and S.18. In each Figure, the horizontal solid red line corresponds to the true parameters α_0 and β_0, respectively.

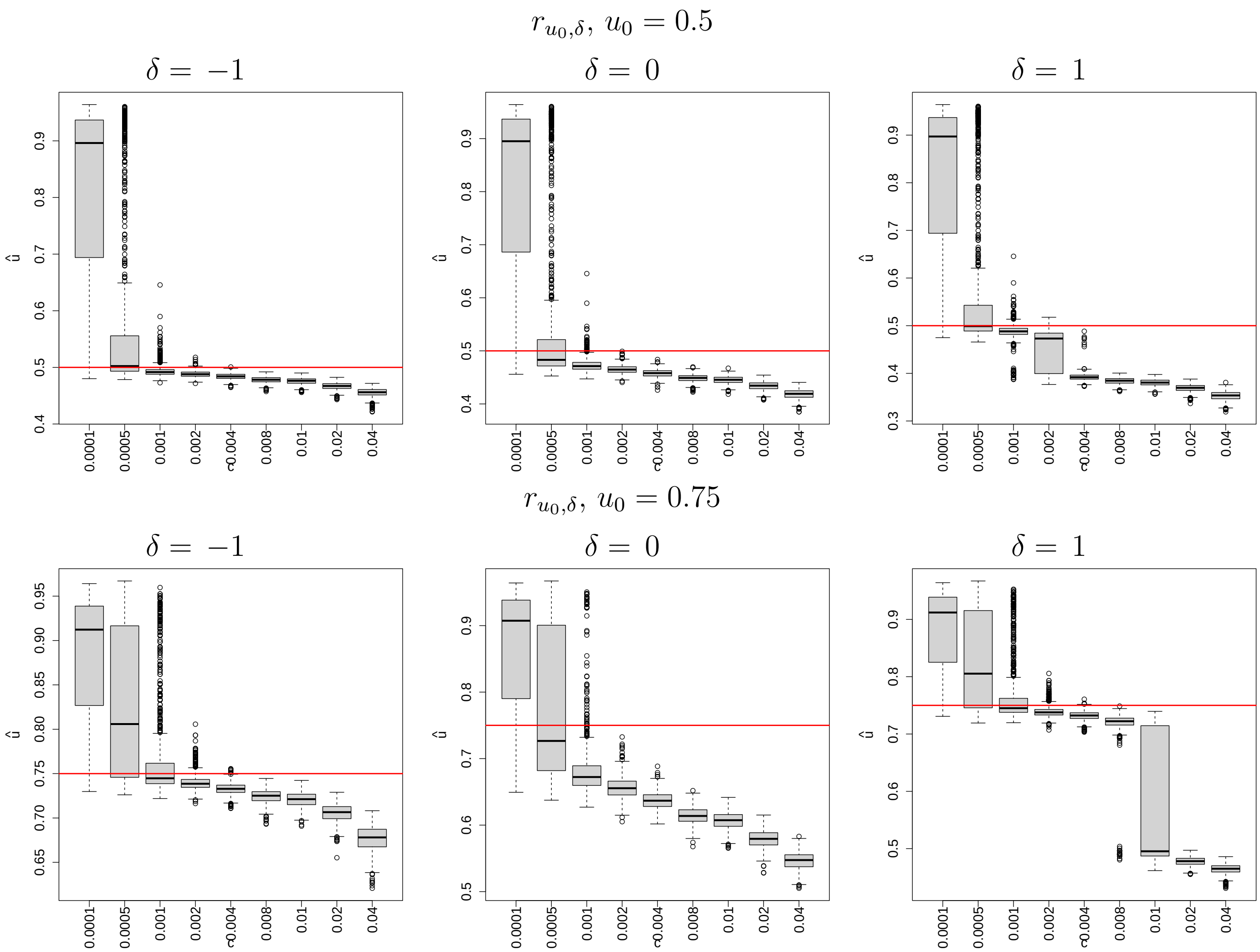

Figure S.10: Boxplots of the estimators $\widehat{u}$ for different choices of the penalizing constant c, when $f(u) = \exp(u)$, $n = 500$ and $\sigma = 0.01$. The horizontal red line corresponds to the true value u_0.

As Figures S.10 to S.12 reveal, the constant leading to boxplots centered at the true value is smaller than the one obtained when $f(u) = u$, see Figures S.1 to S.3. This behaviour may be explained by the fact that the exponential penalization function is larger than $f(u) = u$, so smaller values of c seem more appropriate for estimation purposes.

Regarding the estimates of the regression parameters α_0 and β_0, the same comments given in Section S.1.1 remain valid, that is, for small values of c, the boxplots of $\widehat{\alpha}_{\widehat{u}}$ and $\widehat{\beta}_{\widehat{u}}$ are centered around the true values, while for large values of the threshold when $u_0 = 0.75$, the boxplots of the estimates are mainly completely above or below the true value, indicating their poor performance.

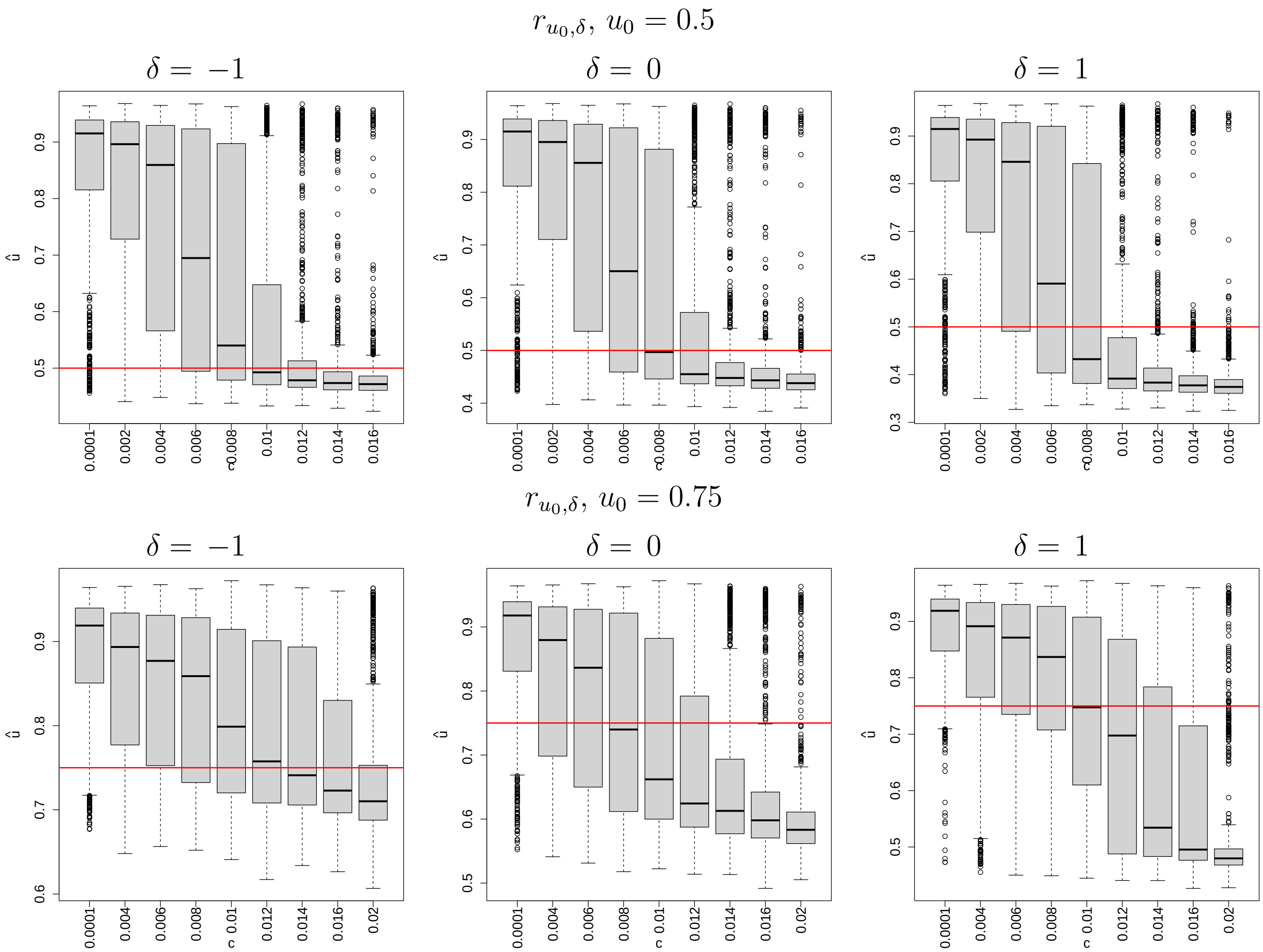

Figure S.11: Boxplots of the estimators $\widehat{u}$ for different choices of the penalizing constant c, when $f(u) = \exp(u)$, $n = 500$ and $\sigma = 0.05$. The horizontal red line corresponds to the true value u_0.

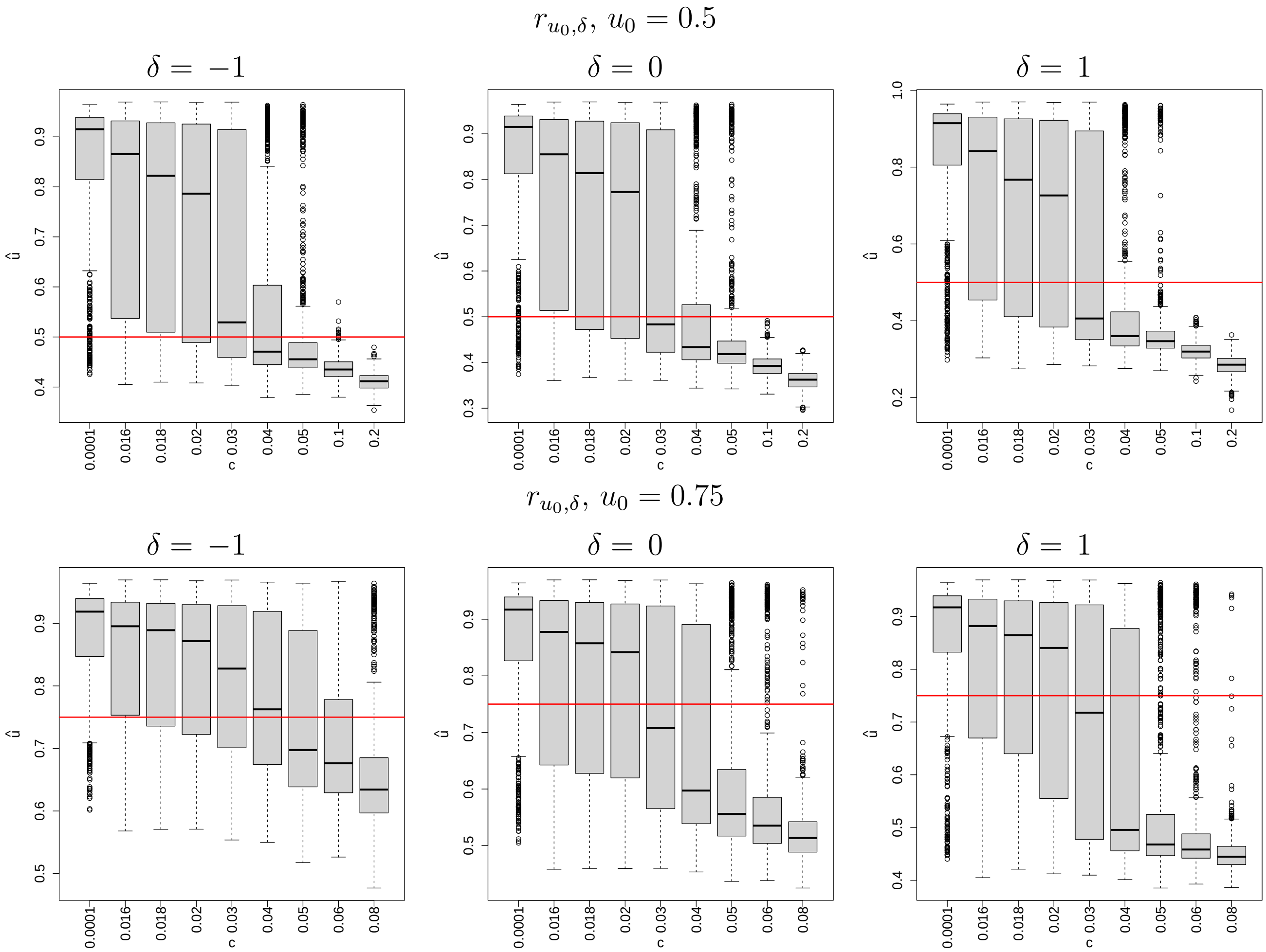

Figure S.12: Boxplots of the estimators $\widehat{u}$ for different choices of the penalizing constant c, when $f(u) = \exp(u)$, $n = 500$ and $\sigma = 0.10$. The horizontal red line corresponds to the true value u_0.

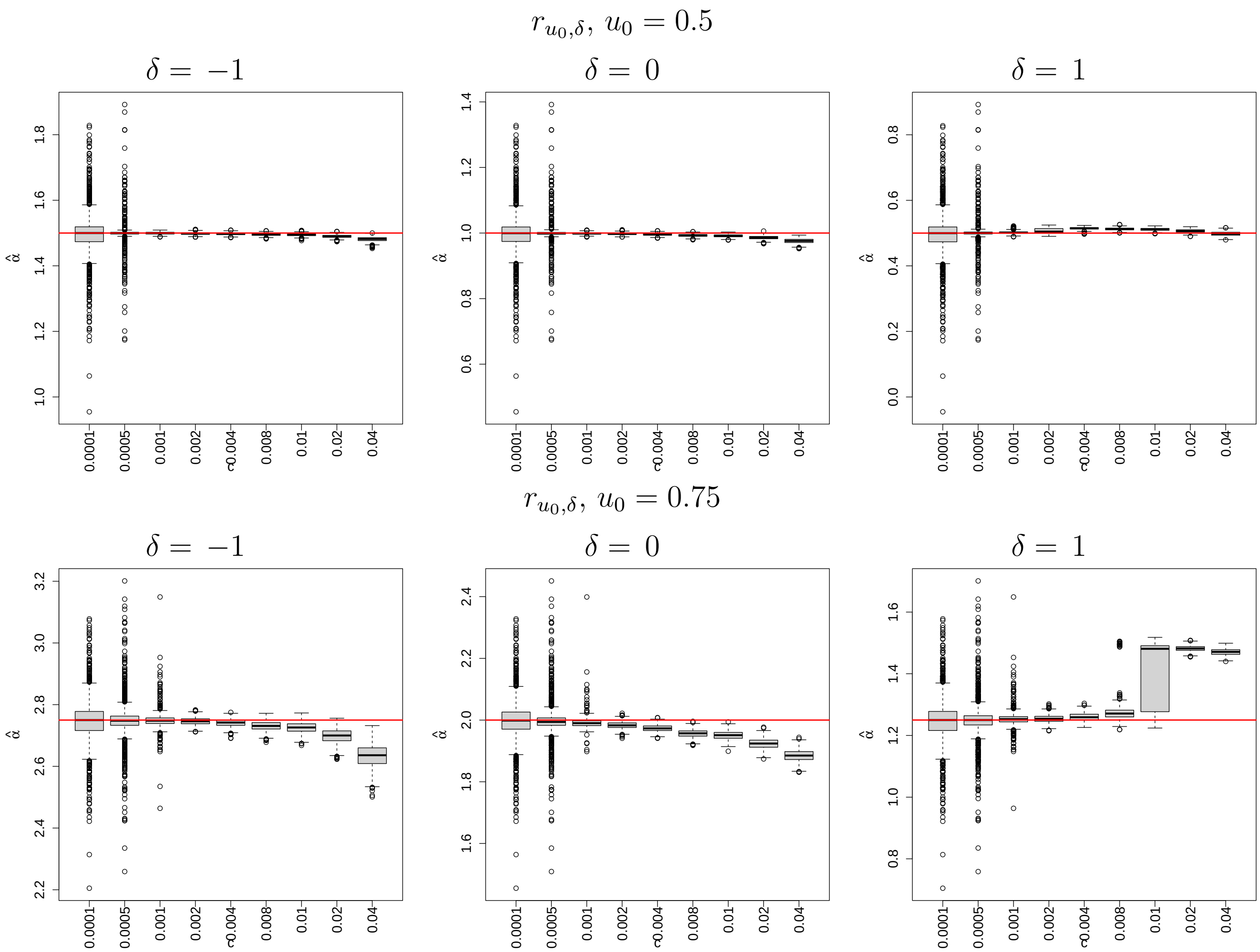

Figure S.13: Boxplots of the estimators $\widehat{\alpha}_{\widehat{u}}$ for different choices of the penalizing constant c, when $f(u) = \exp(u)$, $n = 500$ and $\sigma = 0.01$. The horizontal red line corresponds to the value α_0.

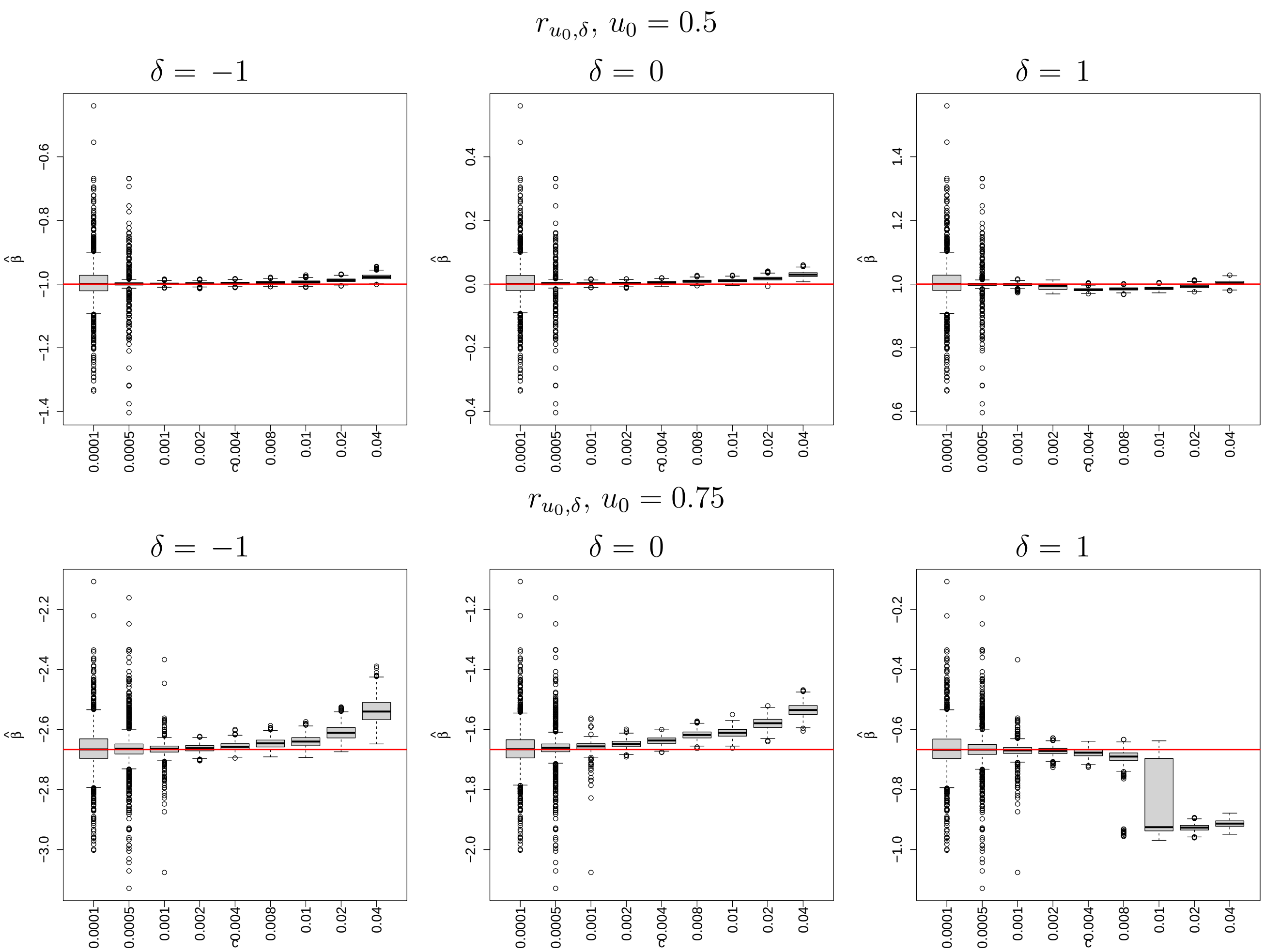

Figure S.14: Boxplots of the estimators $\widehat{\beta}_{\widehat{u}}$ for different choices of the penalizing constant c, when $f(u) = \exp(u)$, $n = 500$ and $\sigma = 0.01$. The horizontal red line corresponds to the value β_0.

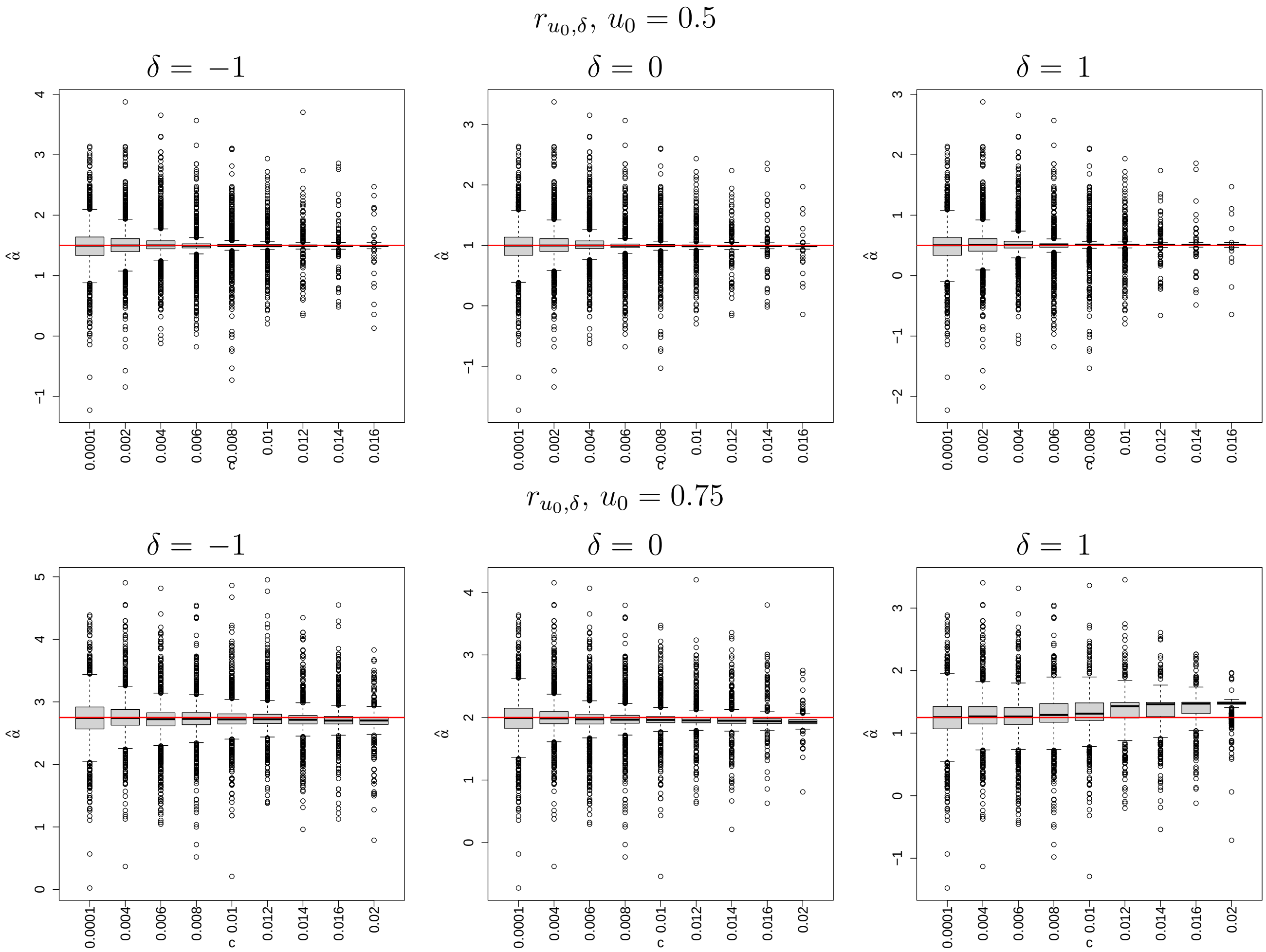

Figure S.15: Boxplots of the estimators $\widehat{\alpha}_{\widehat{u}}$ for different choices of the penalizing constant c, when $f(u) = \exp(u)$, $n = 500$ and $\sigma = 0.05$. The horizontal red line corresponds to the value α_0.

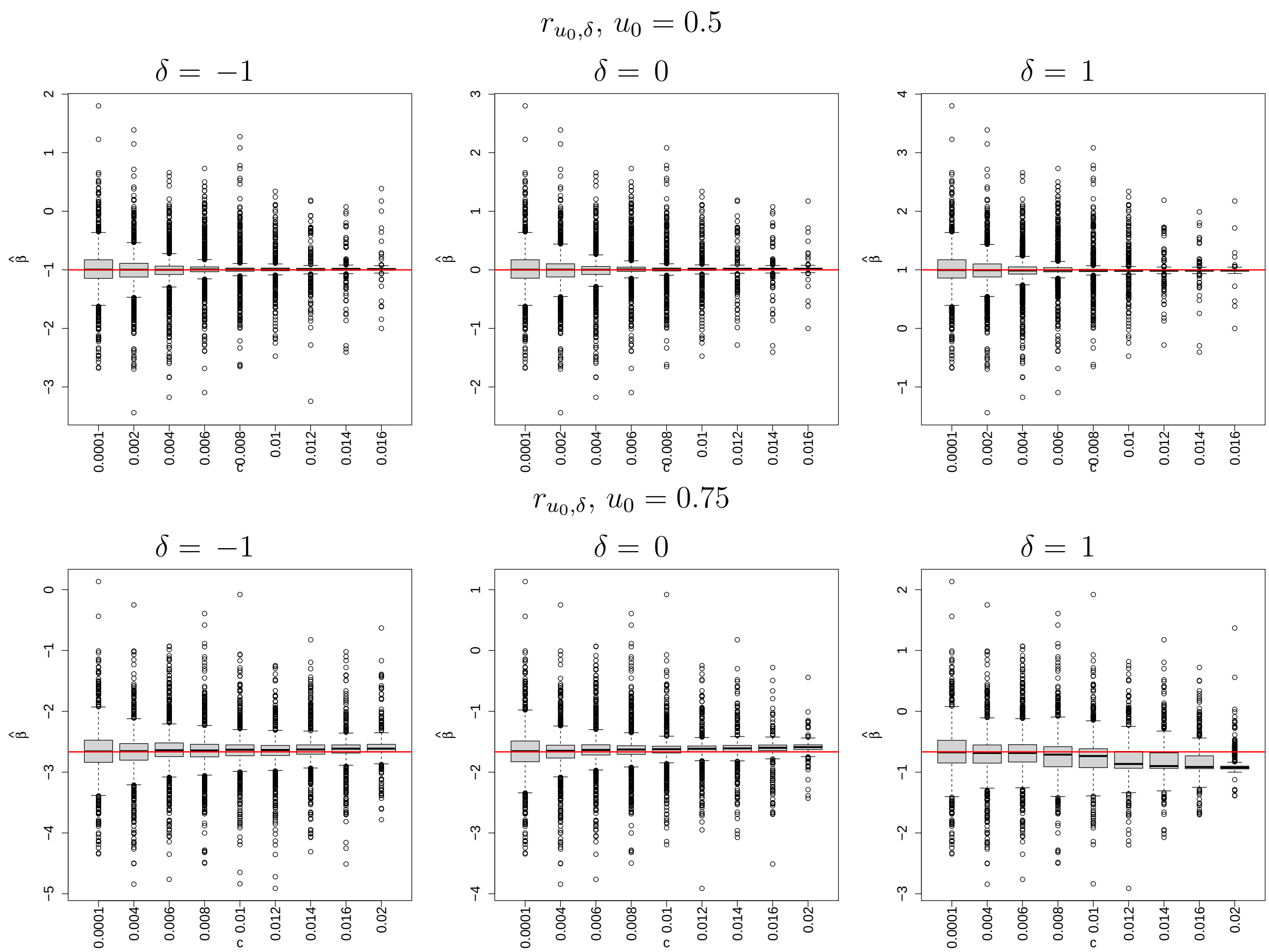

Figure S.16: Boxplots of the estimators $\widehat{\beta_{\widehat{u}}}$ for different choices of the penalizing constant c, when $f(u) = \exp(u)$, $n = 500$ and $\sigma = 0.05$. The horizontal red line corresponds to the value β_0.

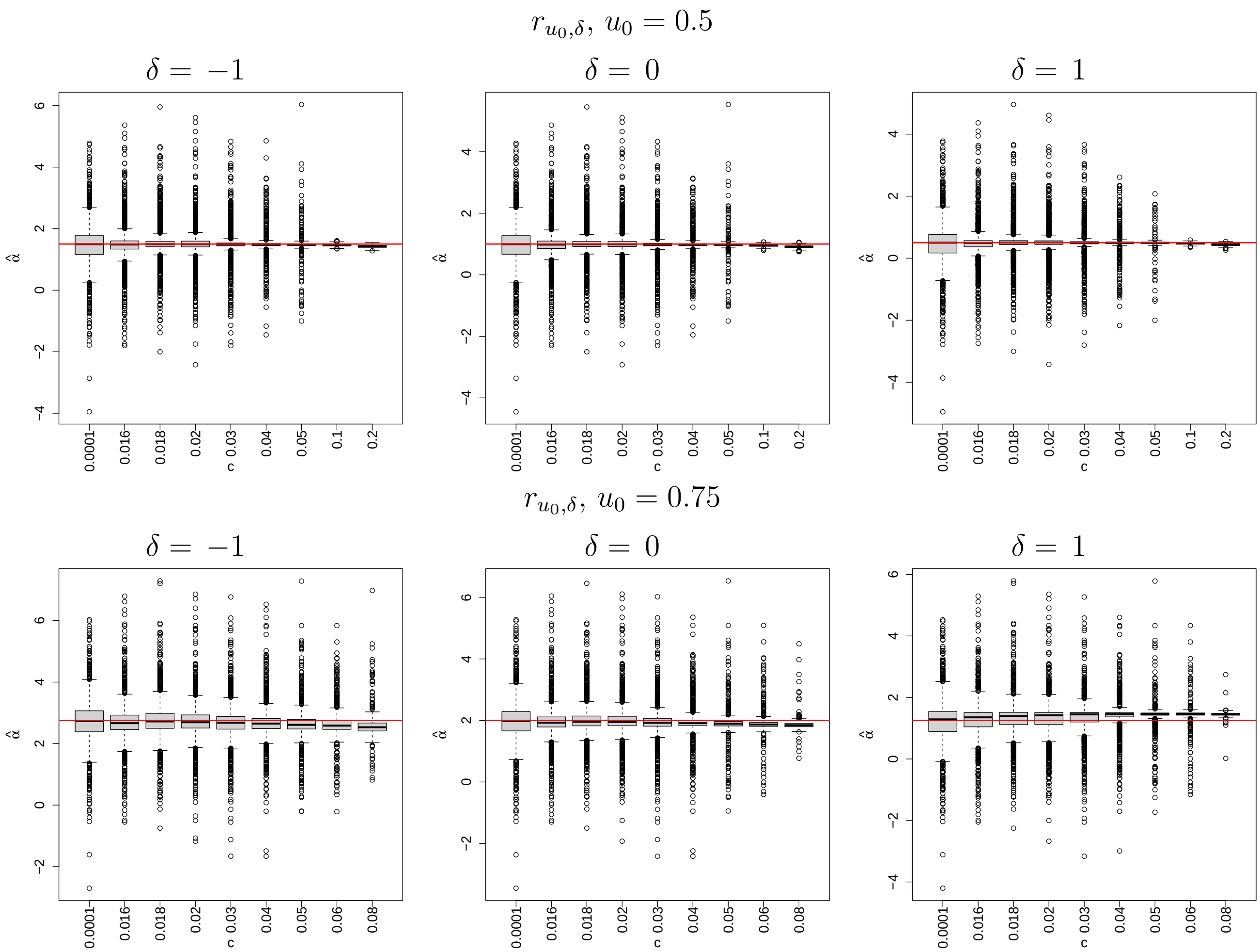

Figure S.17: Boxplots of the estimators $\widehat{\alpha}_{\widehat{u}}$ for different choices of the penalizing constant c, when $f(u) = \exp(u)$, $n = 500$ and $\sigma = 0.10$. The horizontal red line corresponds to the value α_0.

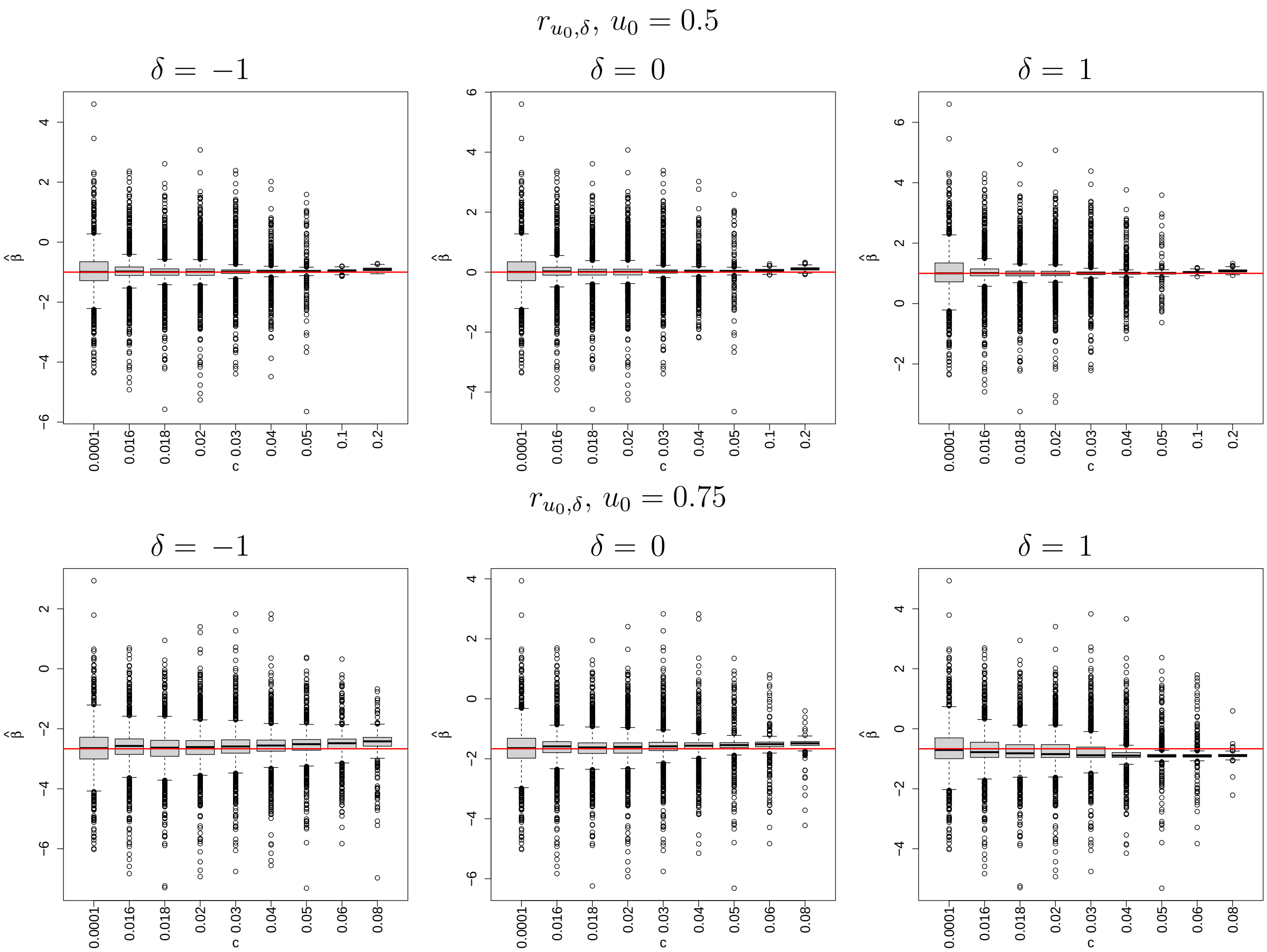

Figure S.18: Boxplots of the estimators $\widehat{\beta}_{\widehat{u}}$ for different choices of the penalizing constant c, when $f(u) = \exp(u)$, $n = 500$ and $\sigma = 0.10$. The horizontal red line corresponds to the value β_0.

S.2 Numerical performance of the method implemented in Zeileis et al. (2002)

To complement the analysis presented in the main document, we include here the results of a simulation study conducted to compare the performance of our proposed threshold detection procedure with that of the method described in Zeileis et al. (2003), which is implemented in the `strucchange` package in `R`, see Zeileis et al. (2002). For fair comparisons and taking into account that our regression function has only one threshold point, we applied the method of Zeileis et al. (2003) fixing the number of change points, given by the parameter `breaks` in the function `breakpoints`, equal to 1. It is worth noticing that the approach in Zeileis et al. (2003) is based on detecting structural breaks in piecewise linear regression models, while our method is more flexible since it only assumes that the model becomes linear only beyond an unknown threshold. We considered two settings: in the first one, as mentioned in the main document, in the piecewise linear regression model considered in Zeileis et al. (2003), the covariates were taken as $\mathbf{z}_i = (1, x_i)^\top$. In the second case, we chose $n = 500$, $\delta = 0.5$ and as covariates $\mathbf{z}_i = (1, x_i, x_i^2, x_i^3)^\top$ in the function `breakpoints`. The choice of δ ensures that only one change point is present. Moreover, we expect that this last scenario will be more beneficial for the approach in Zeileis et al. (2003), since we impose the correct degree of the component below u_0.

S.2.1 Performance when $\mathbf{z}_i = (1, x_i)^\top$

The aim of this section is to evaluate the performance of the method described in Zeileis et al. (2003) when estimating the threshold location u_0, and the regression parameters α_0 and β_0, under the scenarios described in Section S.1 and to compare the obtained results with those reported in Section 4 of the main document and S.1. To provide a graphical summary, easy to interpret, for each scenario, we display the parallel boxplots of the estimates $\widehat{u}$, $\widehat{\alpha}$, and $\widehat{\beta}$ when varying the sample size n. These plots graphically summarizes the estimates variability and bias across simulation runs and allow to visualize their performance as the sample size increases.

Figures S.19 to S.21 display the boxplots of $\widehat{u}$, for $\sigma = 0.01, 0.05$ and 0.10, respectively. The same scale is used in all panels of the same row to facilitate comparisons. The horizontal red lines correspond to u_0. Similarly, Figures S.22, S.24 and S.26 and Figures S.23, S.25 and S.27 present the boxplots of the estimates of α_0 and β_0, respectively. It should be noted that in some panels the red line corresponding to the true value is not present, since it is not within the boxplots range. To facilitate a visual comparison with Figures S.1 to S.9, the boxplots corresponding to $n = 500$ are given in cyan. These plots highlight the differences between our estimates and those proposed in Zeileis et al. (2003), under controlled deviations from the piecewise-linear model assumption. As expected, for all scenarios, a large bias is observed when estimating the threshold using the method in Zeileis et al. (2003). Regarding the estimation of the regression parameters α_0 and β_0, different behaviours are observed according to the values of δ. For $\delta = 0.75$, the estimates considered in the above mentioned paper present a poor performance for all sample sizes. In contrast, when $\delta = 0.5$, these estimates improve their bias performance as the errors standard deviation increases, see Figures S.26 and S.27, for instance. A comparison with the boxplots of our proposal with $n = 500$ reveal that our procedure outperforms the one considered in Zeileis et al. (2003) both regarding the estimation of the change point and of the regression parameters. This behavior is clearly explained by the fact that the underlying regression model is not piecewise linear in x as assumed in Zeileis et al. (2003).

Summarizing, in situations where the shape of the regression function is not known to be linear beyond an unknown breakpoint, the proposal in this paper shows its advantage when estimating the threshold and the regression parameters.

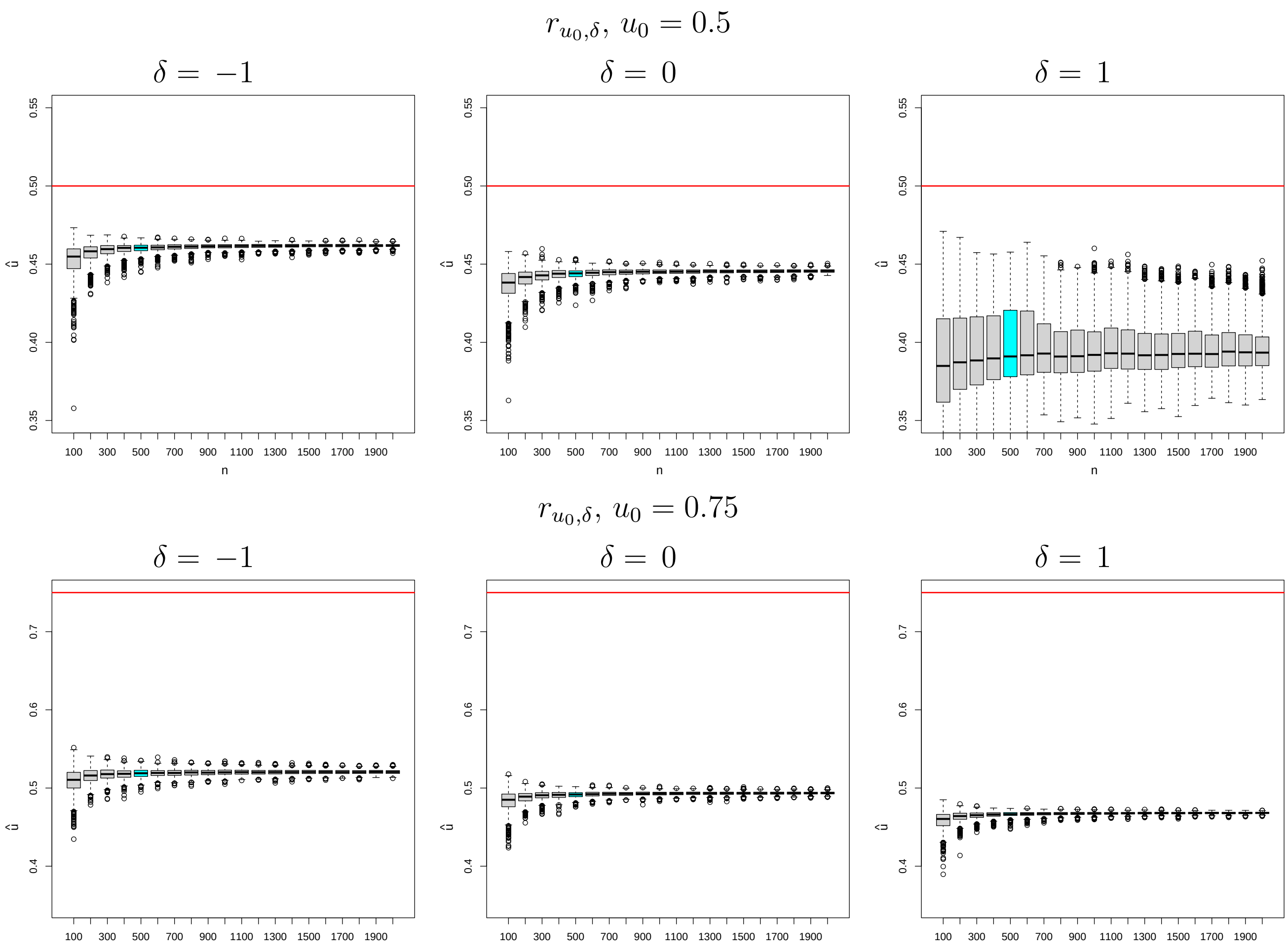

Figure S.19: Boxplots of the estimators $\widehat{u}$ computed using the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.01$. The horizontal red line corresponds to the true value u_0.

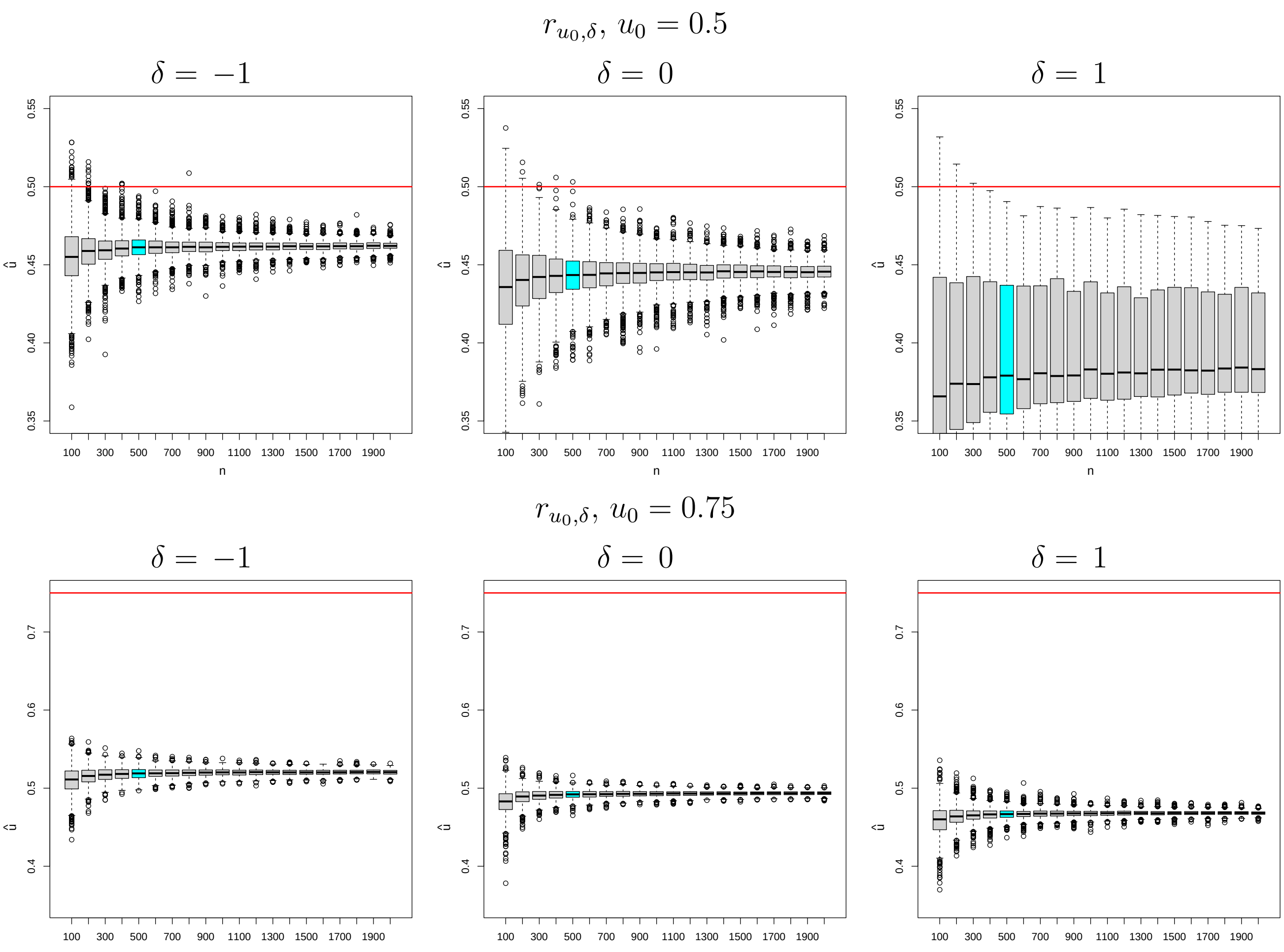

Figure S.20: Boxplots of the estimators $\widehat{u}$ computed using the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.05$. The horizontal red line corresponds to the true value u_0.

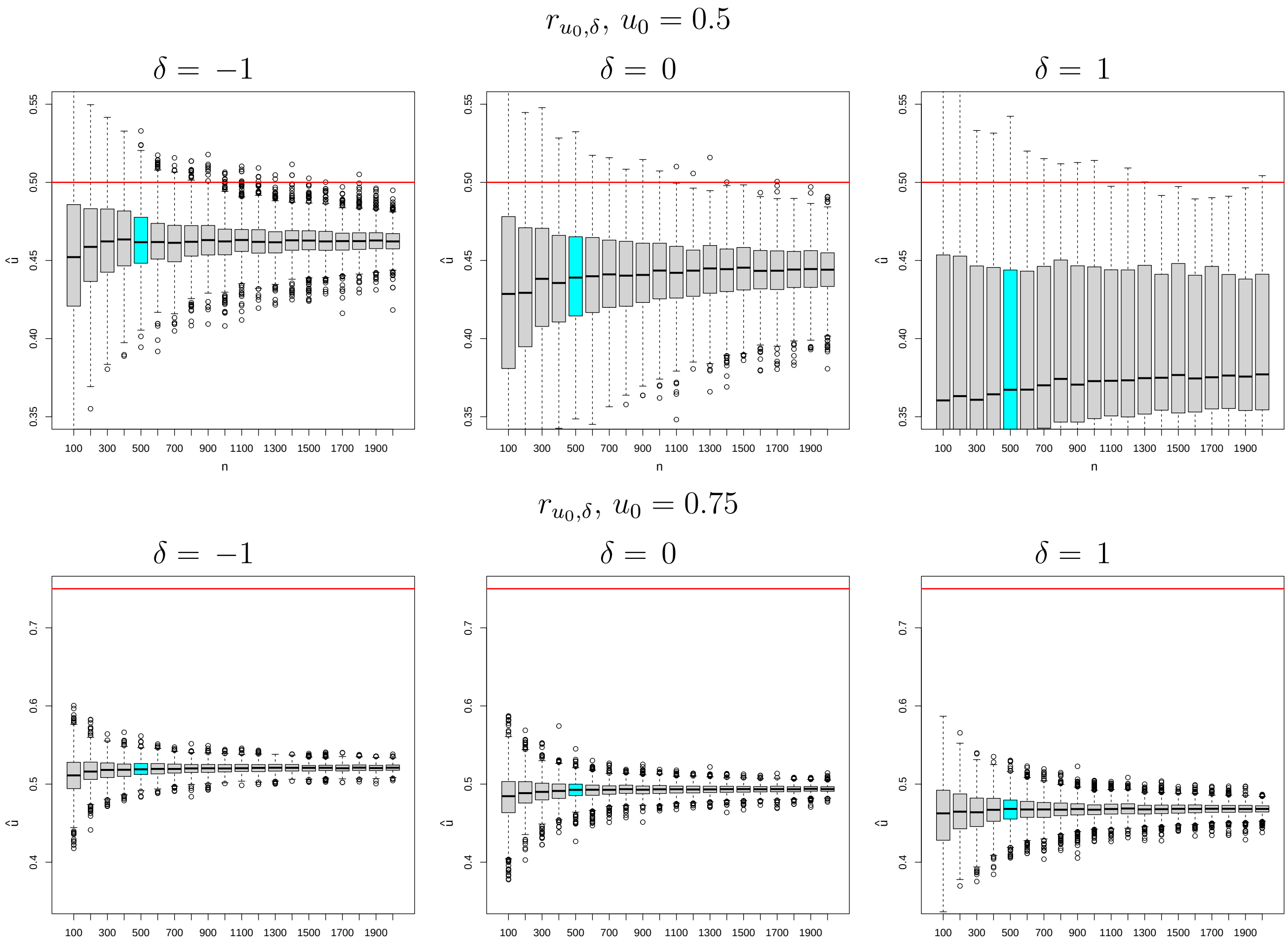

Figure S.21: Boxplots of the estimators $\widehat{u}$ computed using the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.10$. The horizontal red line corresponds to the true value u_0.

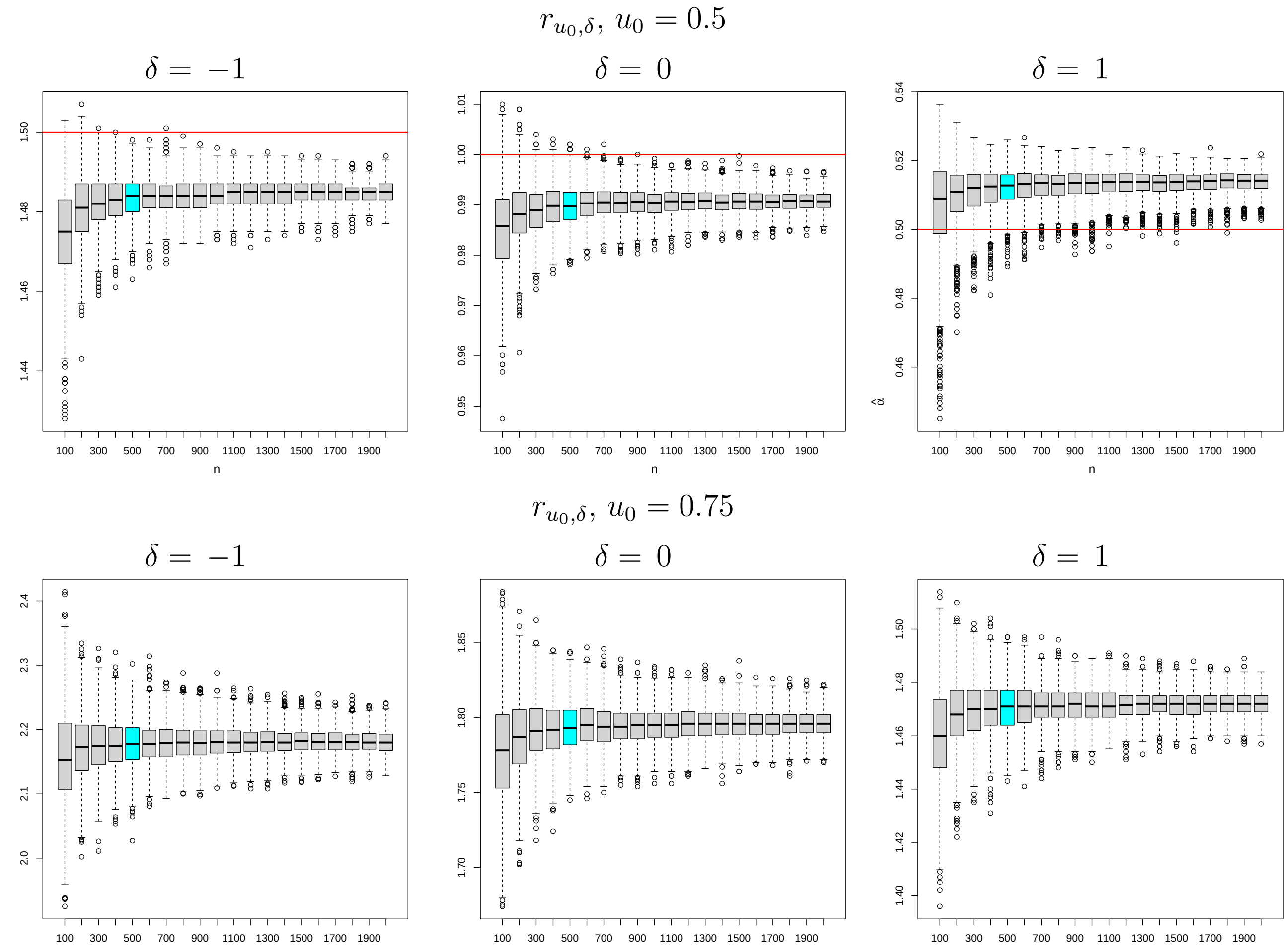

Figure S.22: Boxplots of the estimators $\widehat{\alpha}$ obtained with the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.01$. The horizontal red line, when present, corresponds to the value α_0.

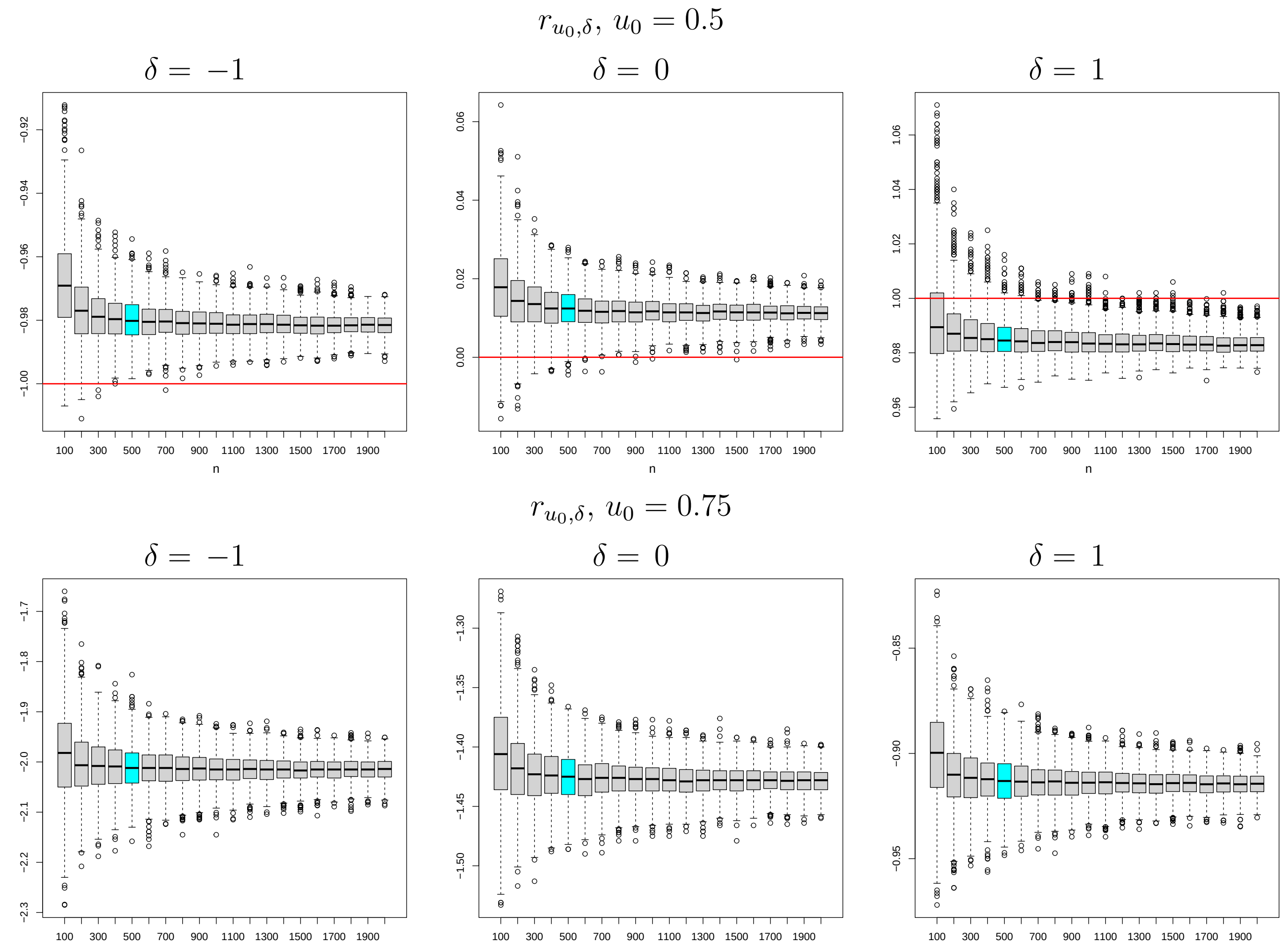

Figure S.23: Boxplots of the estimators $\widehat{\beta}$ obtained with the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.01$. The horizontal red line, when present, corresponds to the value β_0.

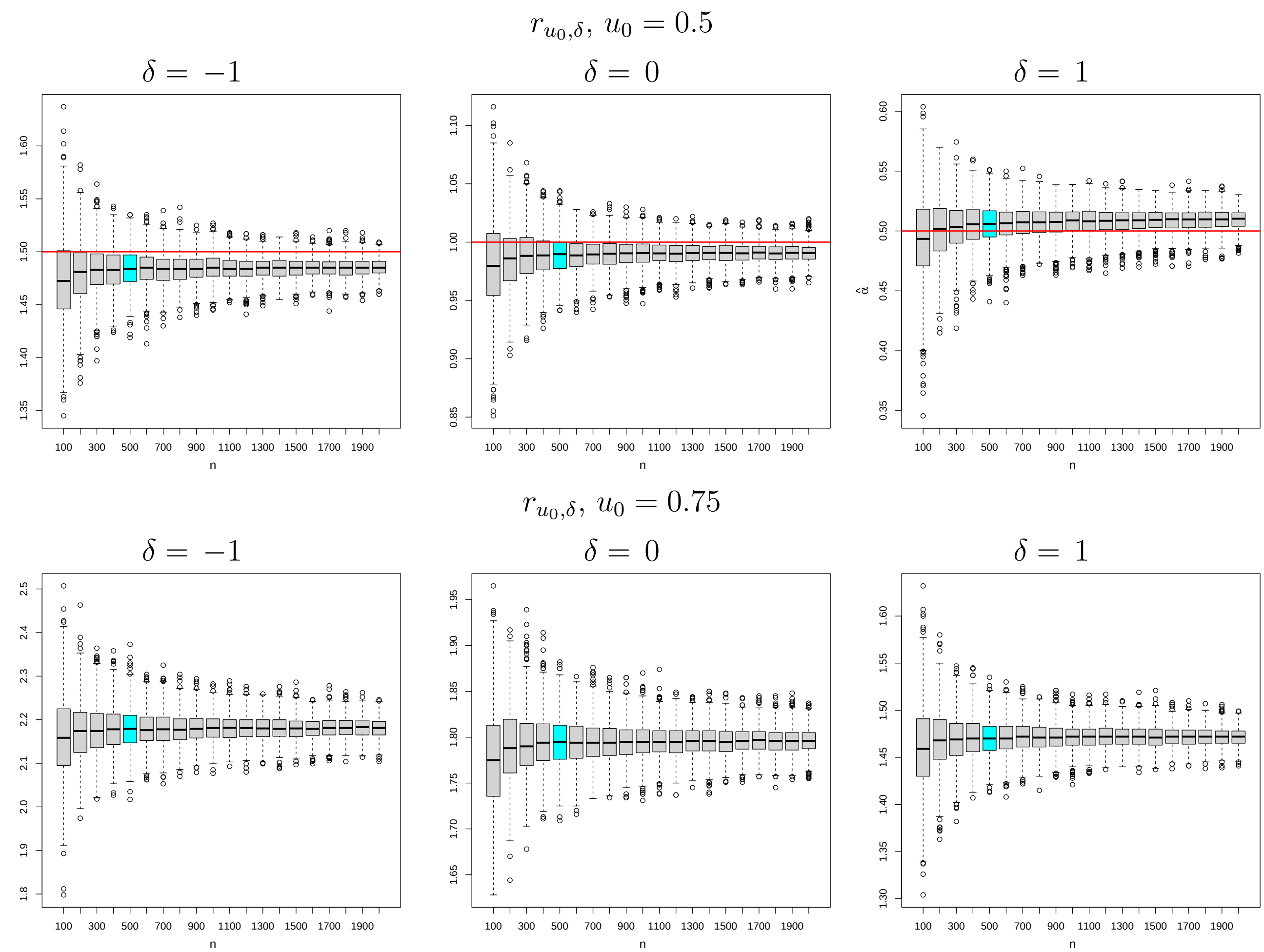

Figure S.24: Boxplots of the estimators $\widehat{\alpha}$ obtained with the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.05$. The horizontal red line, when present, corresponds to the value α_0.

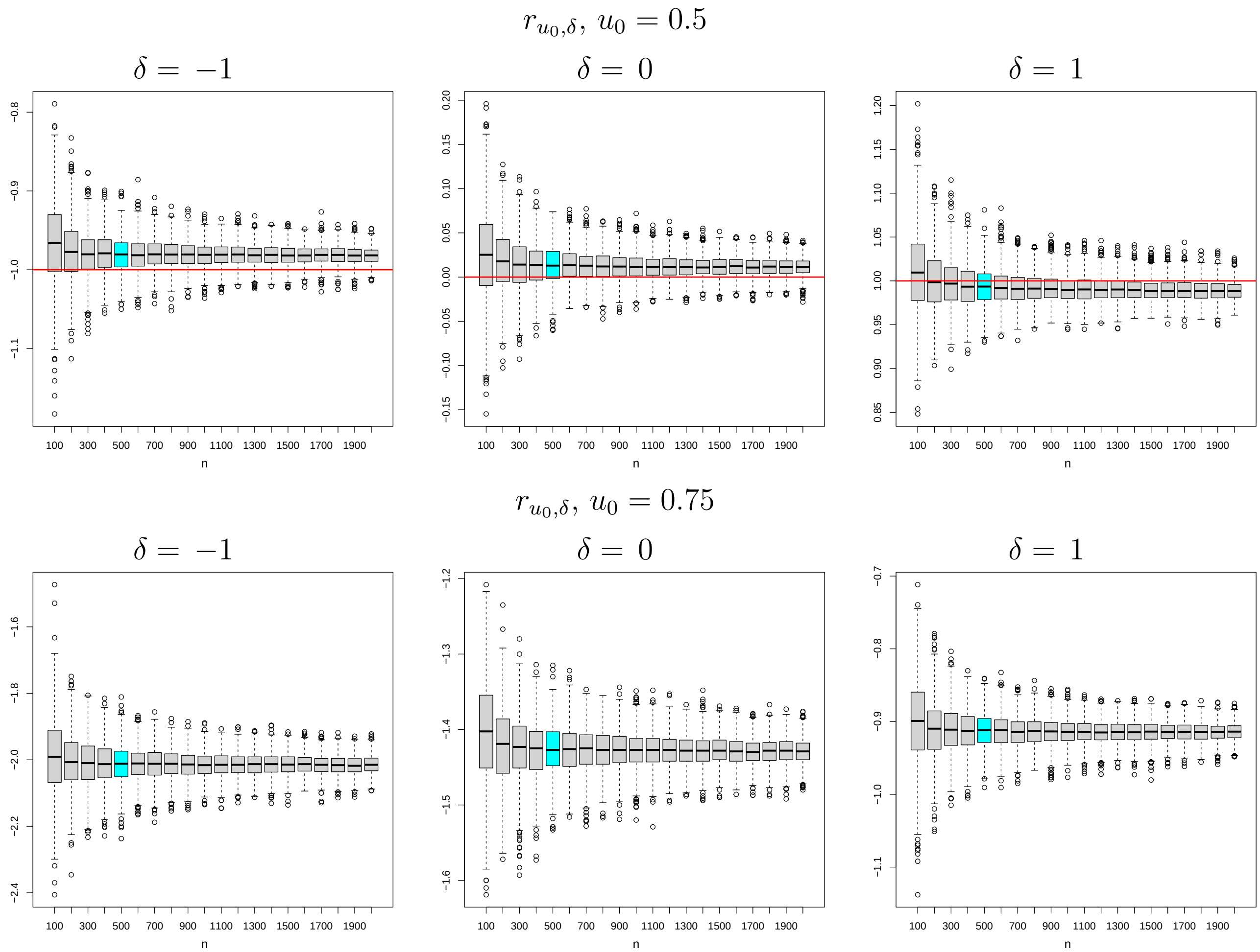

Figure S.25: Boxplots of the estimators $\widehat{\beta}$ obtained with the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.05$. The horizontal red line, when present, corresponds to the value β_0.

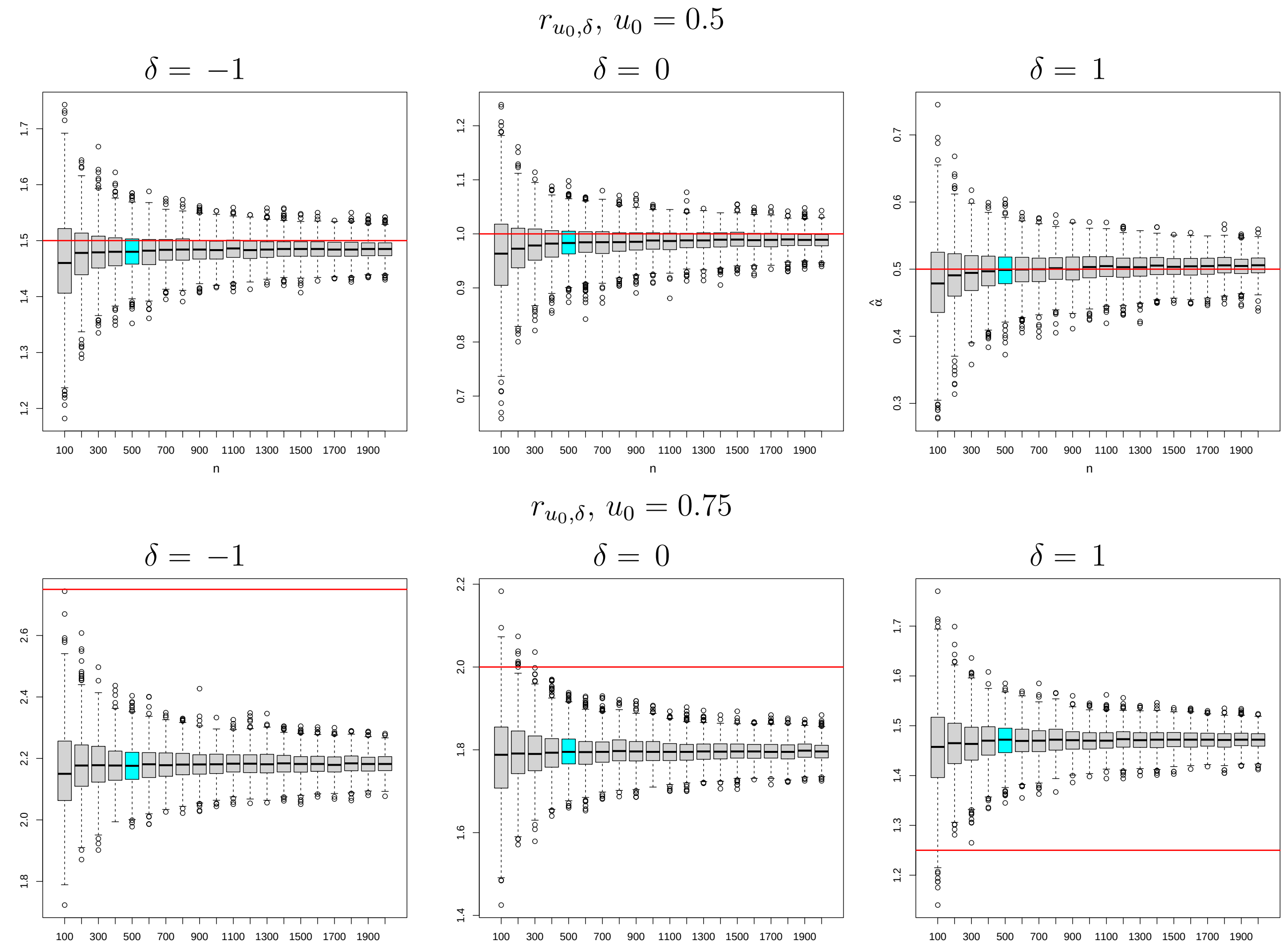

Figure S.26: Boxplots of the estimators $\widehat{\alpha}$ computed using the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.10$. The horizontal red line corresponds to the value α_0.

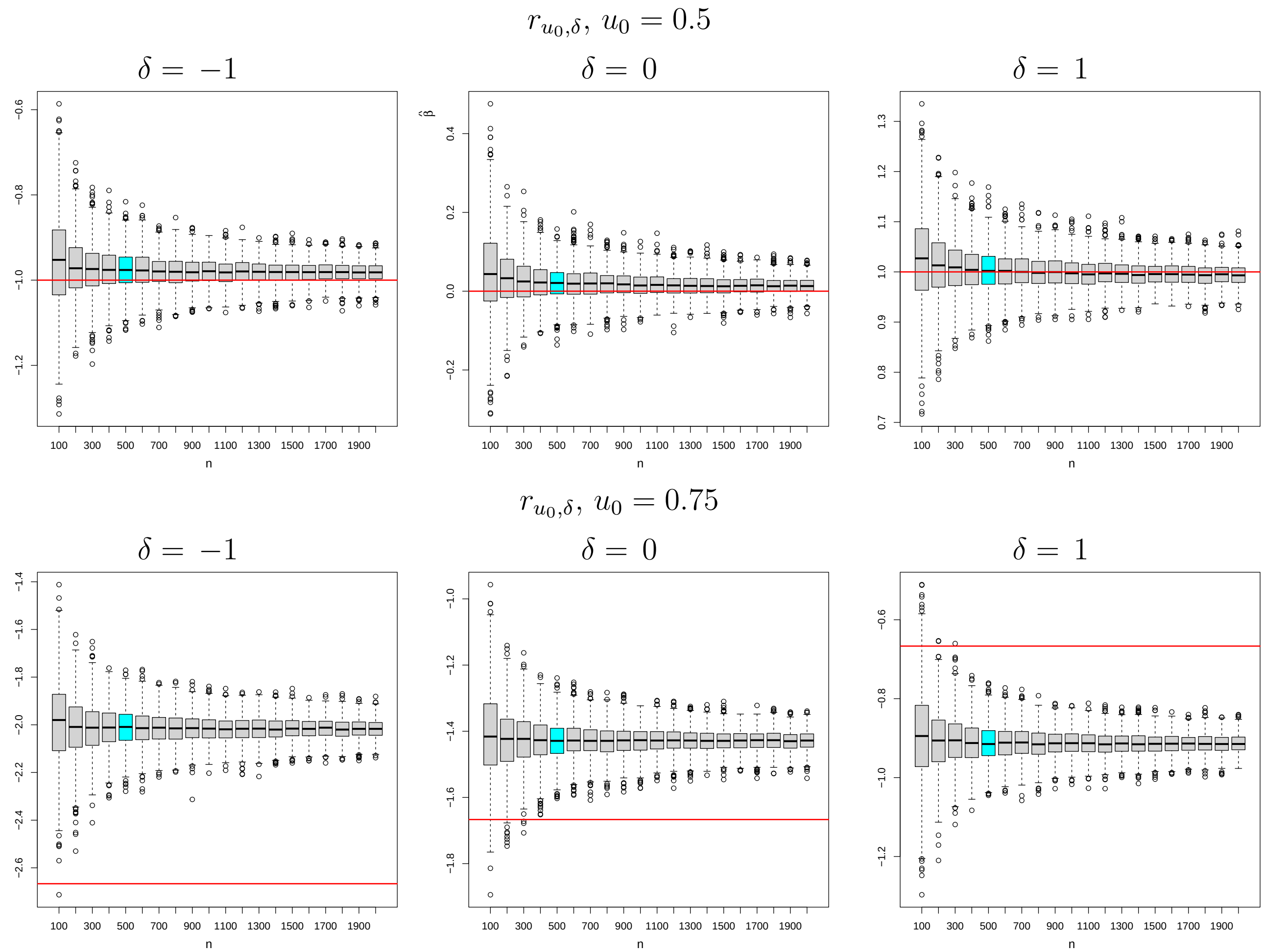

Figure S.27: Boxplots of the estimators $\widehat{\beta}$ computed using the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.10$. The horizontal red line corresponds to the value β_0.

S.2.2 Performance when $\mathbf{z}_i = (1, x_i, x_i^2, x_i^3)^\top$

In this section, we report the results obtained when a piecewise polynomial with degree 3 is considered in the function `breakpoints`. We only present the results when $n = 500$ and $u_0 = 0.5$ and in all figures we display parallel boxplots as σ varies. For notational purposes, and since the regression function above the threshold equals $r(x) = \alpha_0 + \beta_0 x$, the coefficients above the estimated change point $\widehat{u}$ are denoted as $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\beta}_2$ and $\widehat{\beta}_3$, where the two last ones correspond to the estimates associated to x^2 and x^3, whose true value is 0 in the considered model.

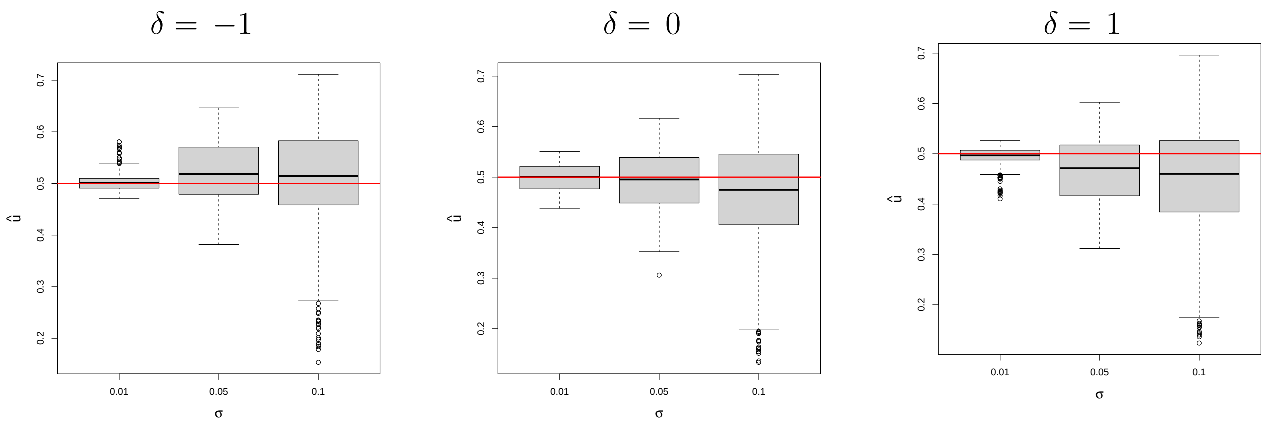

Figure S.28: Boxplots of the estimators $\widehat{u}$ obtained with the method implemented by Zeileis et al. (2002) when a third degree piecewise polynomial is considered. The horizontal red line corresponds to the value u_0.

Figure S.28 presents the boxplots of the threshold estimators, while Figure S.29 displays the boxplots of the estimators $\widehat{\alpha}$ and $\widehat{\beta}$. Unlike the case when a simple piecewise regression model is considered, that is, when $\mathbf{z}_i = (1, x_i)^\top$, the estimators perform quite well, since the boxplots are centered. However, it is worth mentioning that the estimators of α_0 and β_0 show a larger variability than those obtained with our proposal, see Figures S.4 to S.9 and Figures S.13 to S.18. For that reason, in cyan we display the boxplots within a reduced range to facilitate the comparison with the later figures.

Figure S.30 displays the boxplots of the estimators $\widehat{\beta}_2$ and $\widehat{\beta}_3$. Again, due to their large variability, the boxplots in cyan correspond to a reduced range. Note that, when $\delta = -1$ and $\sigma = 0.01$, most estimates of the second and third degree polynomial are below the true value, suggesting to the practitioner that the model above the threshold may be a polynomial of degree $p = 3$ instead of a simple regression model. The same behaviour arises when $\delta = 1$, but in this case the estimates are mostly above 0. When $\sigma = 0.05$, for $\delta = -1$ and 1 the boxplots are not centered but the true value belongs to the central box, instead when $\sigma = 0.10$, the boxplots are centered but the estimates present a large variability. When $\delta = 0$, for the three values of σ the boxplots cross the line corresponding to the true value 0. However, except when $\sigma = 0.01$ the estimators variability is large.

Summarizing, even when the procedure implemented by Zeileis et al. (2002) correctly detects the true threshold, the estimators of the intercept and slope show a larger dispersion than those corresponding to our proposal. Furthermore, the obtained coefficients for x^2 and x^3 are in some situations very biased and may lead to the wrong conclusion, suggesting a model with more coefficients than those really present. A key point in the application of the estimating method described in Zeileis et al. (2002) is that the model should be assumed as a piecewise linear model, in the case, of the present simulation as a piecewise polynomial with known degree. In contrast, our procedure does not assume any model for the regression function below the threshold, making it more flexible to situations where, for instance, the regression function $r(x)$ is nonlinear below u_0.

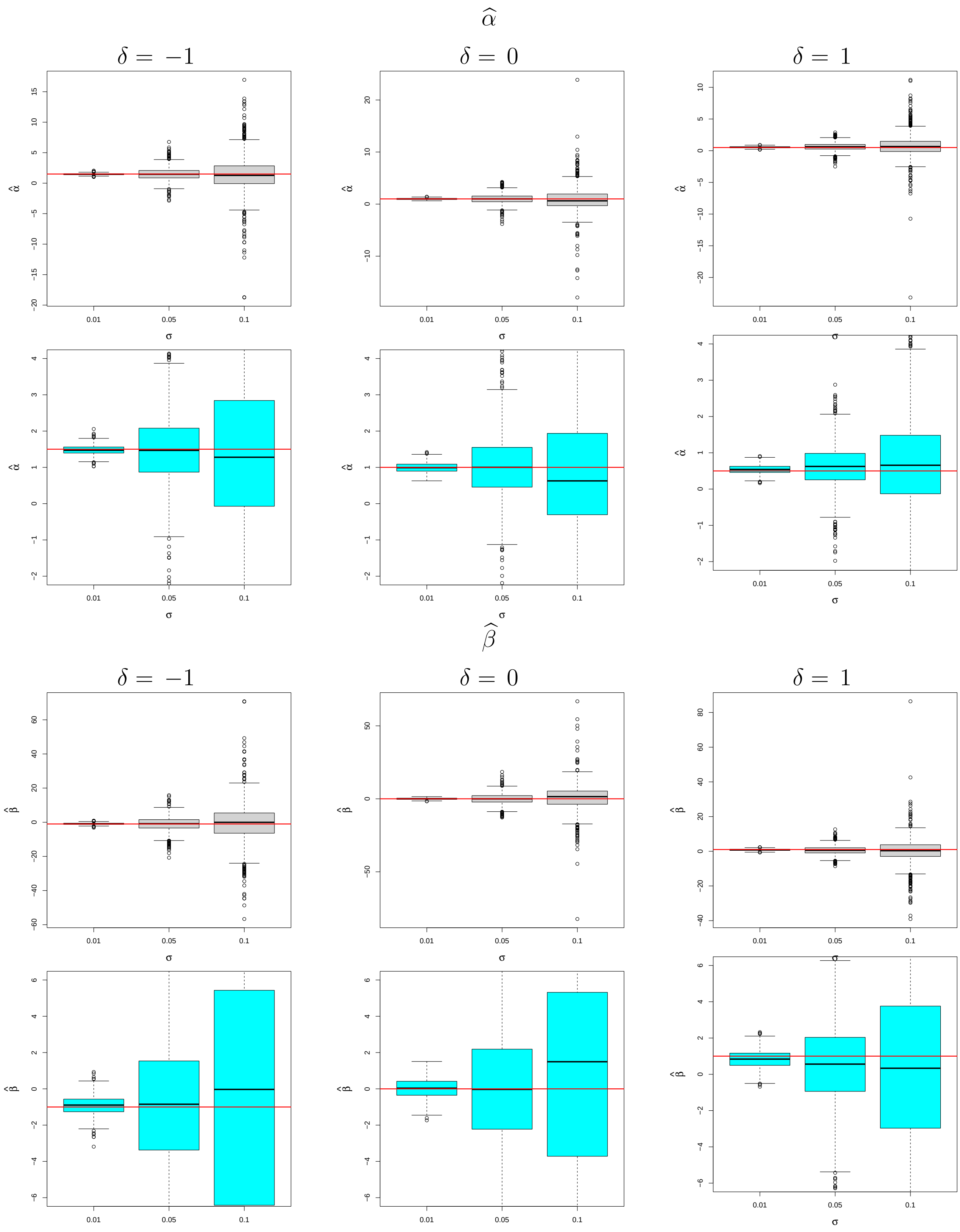

Figure S.29: Boxplots of the estimators $\widehat{\alpha}$ and $\widehat{\beta}$ obtained with the method implemented by Zeileis et al. (2002) when a third degree piecewise polynomial is considered. The horizontal red line corresponds to the value of the parameters α_0 and β_0. The panels with the boxplots in cyan correspond to a reduced estimators range.

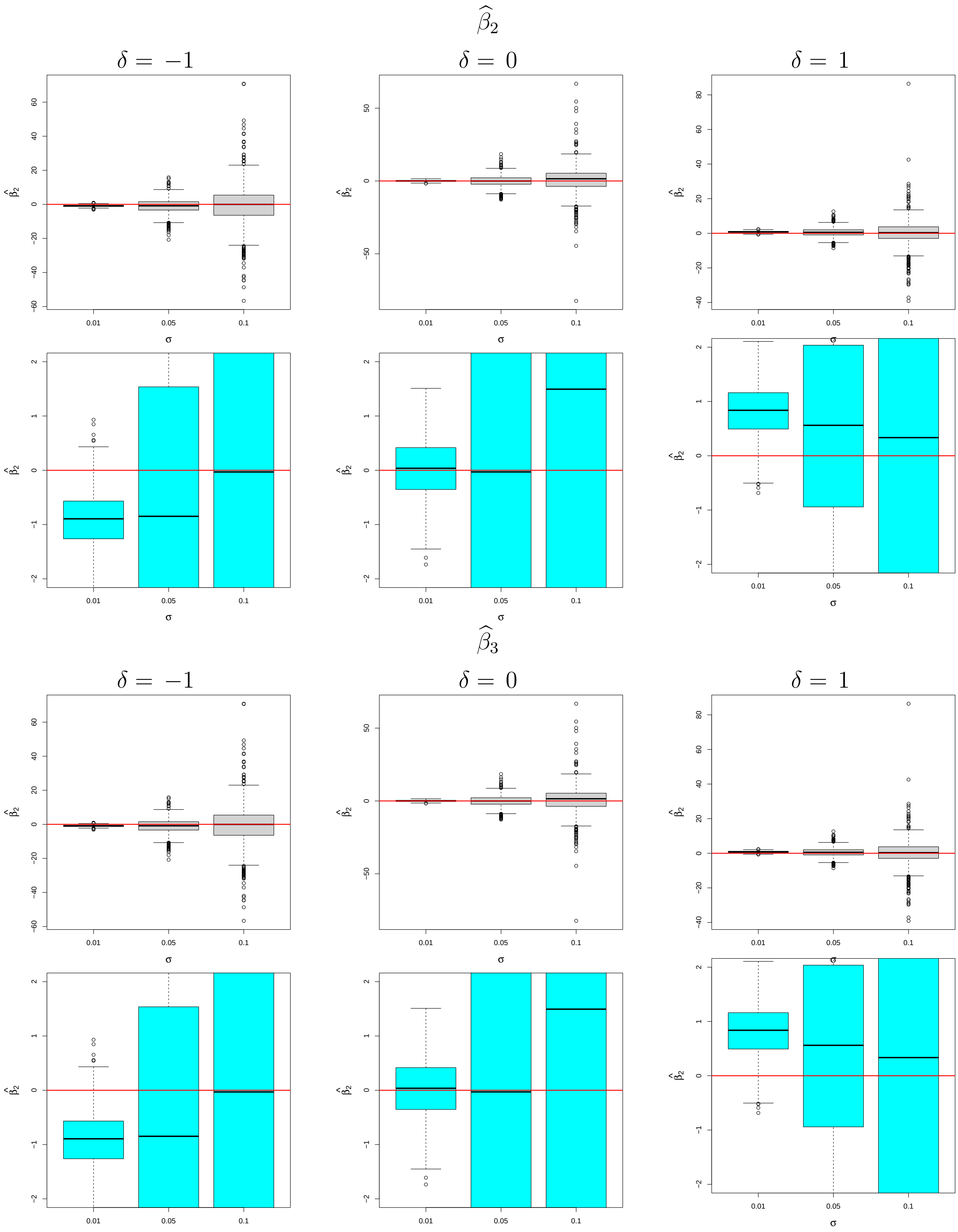

Figure S.30: Boxplots of the estimators $\widehat{\beta}_2$ and $\widehat{\beta}_3$ obtained with the method implemented by Zeileis et al. (2002) when a third degree piecewise polynomial is considered. The horizontal red line corresponds to the true value 0. The panels with the boxplots in cyan correspond to a reduced estimators range.

S.3 Numerical results for Laplace distributed errors

In this section, we consider a framework where the errors have tails heavier than the normal distribution. From now on, we denote as $\mathcal{DE}(\mu, \varsigma^2)$, where $\mu \in \mathbb{R}$ and $\varsigma > 0$, the Laplace distribution with density

$$f(t) = \frac{1}{2\varsigma} \exp\left\{-\frac{|t-\mu|}{\varsigma}\right\}.$$

As above, the samples (X_i, Y_i), $i = 1, \dots, n$, are generated with the same distribution as (X, Y), where $X \sim \mathcal{U}(0, 1)$, $Y = r_{u_0,\delta}(X) + \varepsilon$ and ε is independent of X. Taking into account that if $Z \sim \mathcal{DE}(\mu, \varsigma^2)$, $\mathbb{E}(Z) = \mu$ and $\mathbb{V}ar(Z) = 2\,\varsigma^2$ and to ensure that the errors under this setting have the same variance as in the normally distributed one, we consider the situation where $\varepsilon \sim \mathcal{DE}(0, \sigma^2/2)$.

As in Section S.1.1, to study the performance of the proposed procedure, we consider the sample size $n = 500$ and we summarize the results using the estimates boxplots for different choices of the penalizing constant c. Figures S.31 to S.33 display the boxplots of the threshold estimators for $u_0 = 0.5$ and 0.75, $\delta = -1, 0, 1$ and $\sigma = 0.01, 0.05$ and 0.1 with varying penalty constants c. The true parameter is indicated with the horizontal solid red line. Besides, Figures S.34 and S.35 present the boxplots of the estimators $\widehat{\alpha}_{\widehat{u}}$ and $\widehat{\beta}_{\widehat{u}}$, respectively, when $\sigma = 0.01$. The boxplots of the estimates of α_0 when $\sigma = 0.05$ and 0.10 are displayed in Figures S.36 and S.38, respectively, while those of $\widehat{\beta}_{\widehat{u}}$ are given in Figures S.37 and S.39. In each Figure, the horizontal solid red line corresponds to the true parameters α_0 and β_0, respectively. The boxplots show that the estimators perform as when the errors are normally distributed. It is worth mentioning that the value of the constant c where the median over replications of the estimates $\widehat{u}$ is close to the true value is larger than when considering normal errors.

As in Section S.2, we also study the numerical behaviour of the threshold detection procedure described in Zeileis et al. (2002). We begin by considering the same framework as in Section S.2.1, that is, when the covariates were taken as $\mathbf{z}_i = (1, x_i)^\top$, even when for Gaussian errors our procedure already outperformed the one considered in Zeileis et al. (2002). Figures S.40 to S.42 display the boxplots of $\widehat{u}$, for $\sigma = 0.01, 0.05$ and 0.10, respectively, when varying the sample size n. The same scale is used in all panels of the same row to facilitate comparisons. The horizontal red lines correspond to u_0. Similarly, Figures S.43, S.45 and S.47 and Figures S.44, S.46 and S.48 present the boxplots of the estimates of α_0 and β_0, respectively. It should be noted that in some panels the red line corresponding to the true value is not present, since it is not within the boxplots range. As in Section S.2.1, the boxplots corresponding to $n = 500$ are given in cyan. As when the errors have a normal distribution, for all scenarios, a large bias is observed when estimating the threshold using the method in Zeileis et al. (2003). Again, different behaviours are observed when estimating α_0 and β_0 according to the values of δ, but the conclusions given for normal errors remain valid. More precisely, for $\delta = 0.75$, the estimates considered in Zeileis et al. (2003) present a poor performance for all sample sizes, while for $\delta = 0.5$, these estimates improve their bias performance as the errors standard deviation increases. It is worth mentioning that, when comparing these estimators with our proposal for $n = 500$, the boxplots show the advantages of our procedure both regarding the estimation of the change point and of the regression parameters. As mentioned above this behavior may be explained by the fact

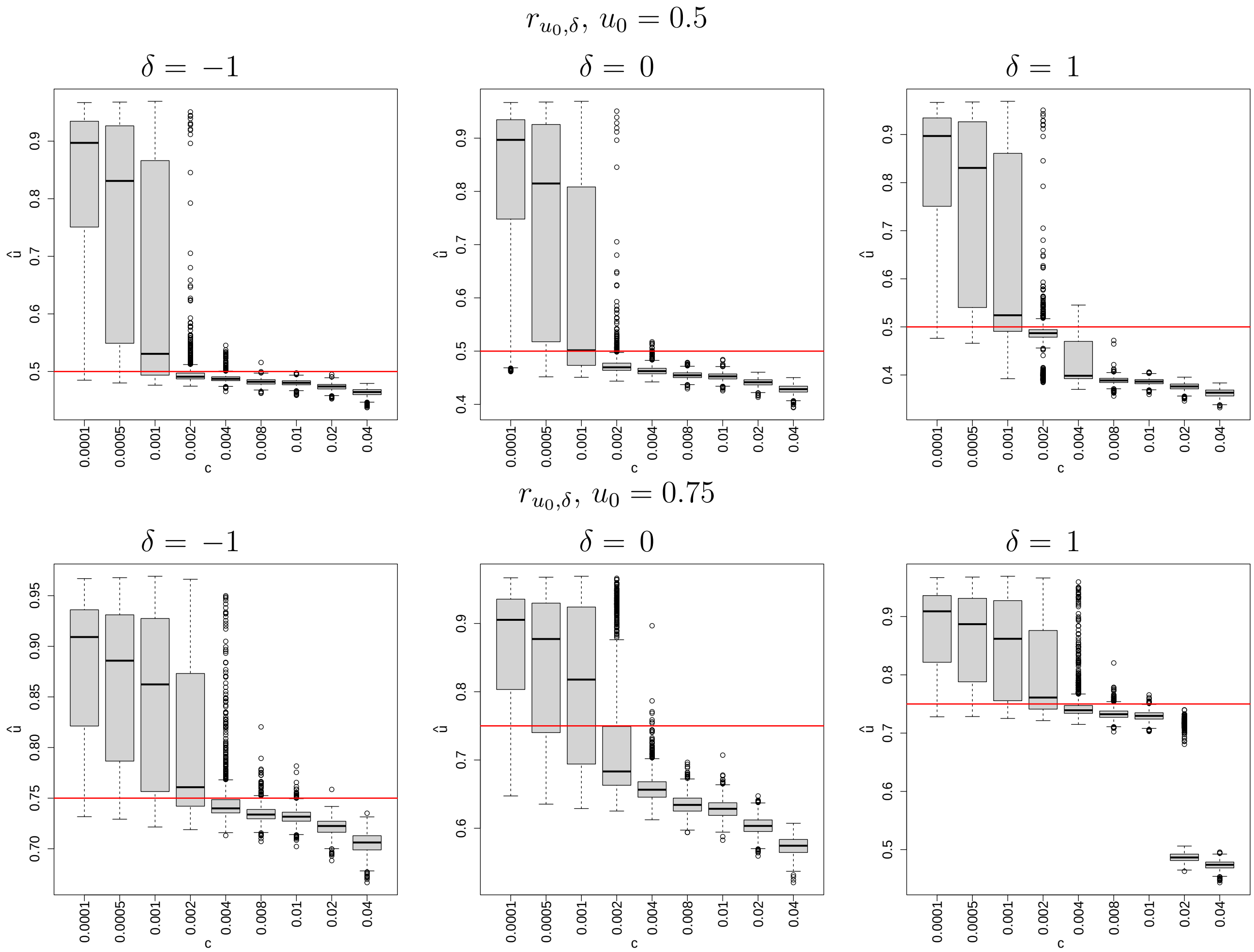

Figure S.31: Boxplots of the estimators $\widehat{u}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.01$, when the errors have a Laplace distribution. The horizontal red line corresponds to the true value u_0.

that the underlying regression model is not piecewise linear in x.

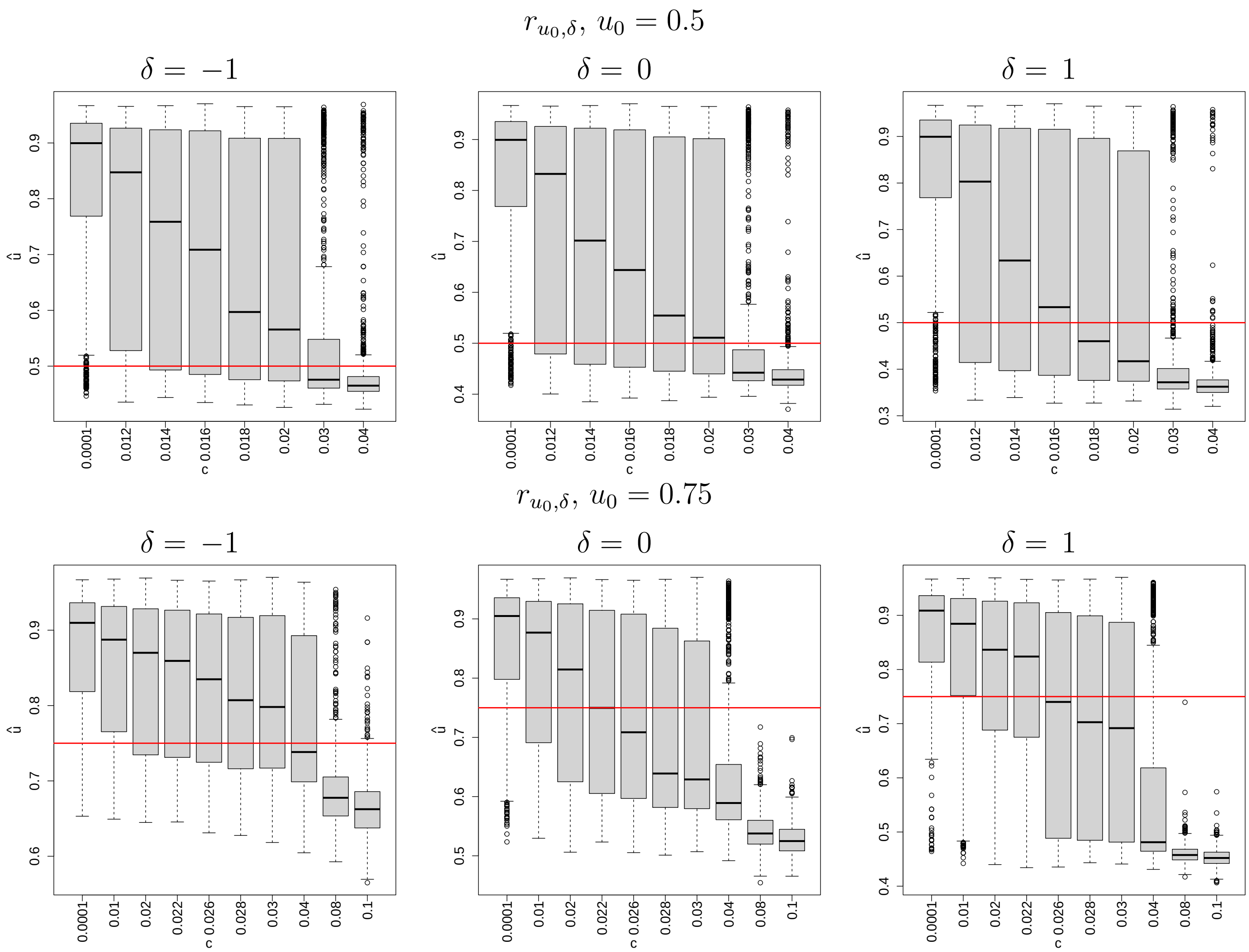

Figure S.32: Boxplots of the estimators $\widehat{u}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.05$, when the errors have a Laplace distribution. The horizontal red line corresponds to the true value u_0.

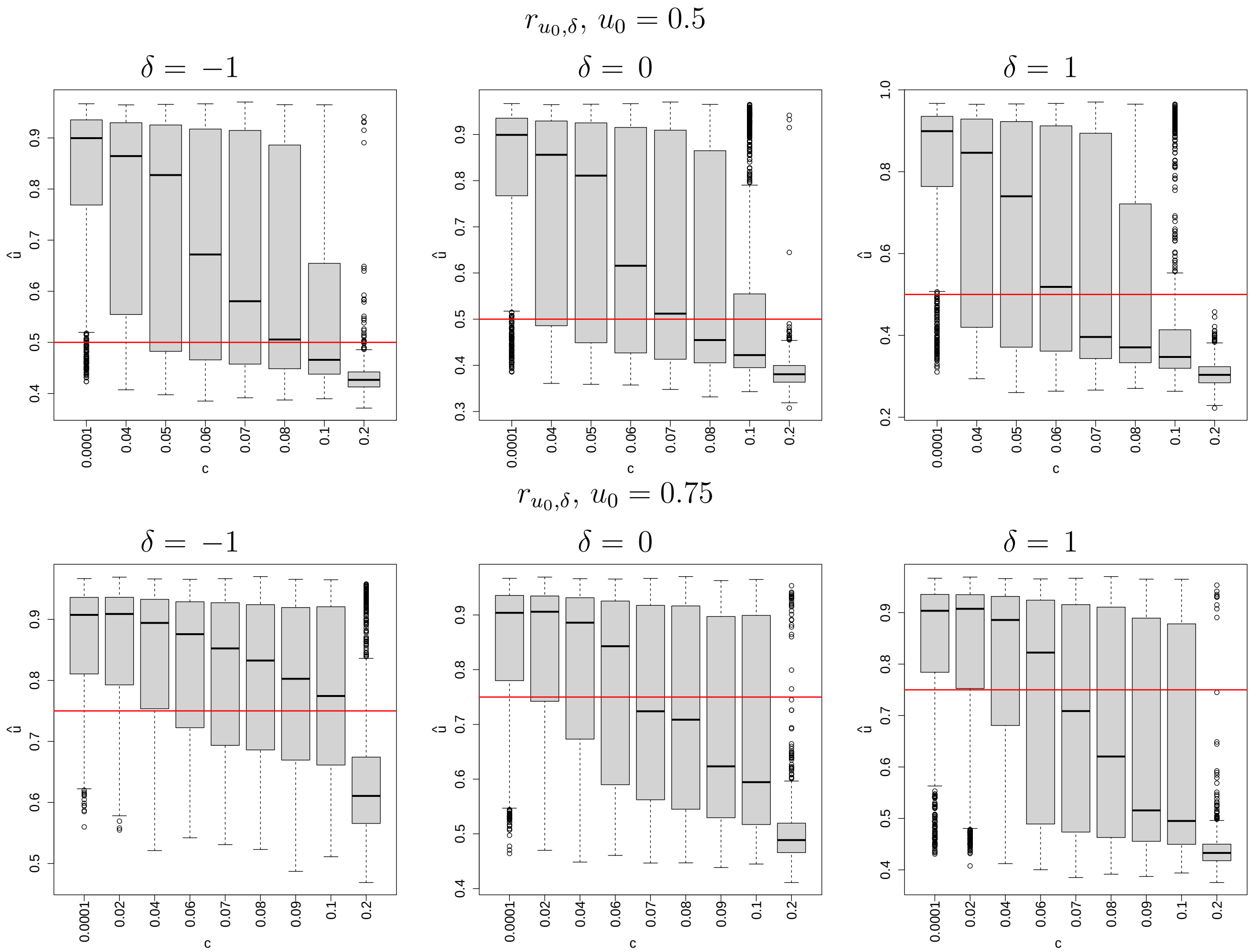

Figure S.33: Boxplots of the estimators $\widehat{u}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.10$, when the errors have a Laplace distribution. The horizontal red line corresponds to the true value u_0.

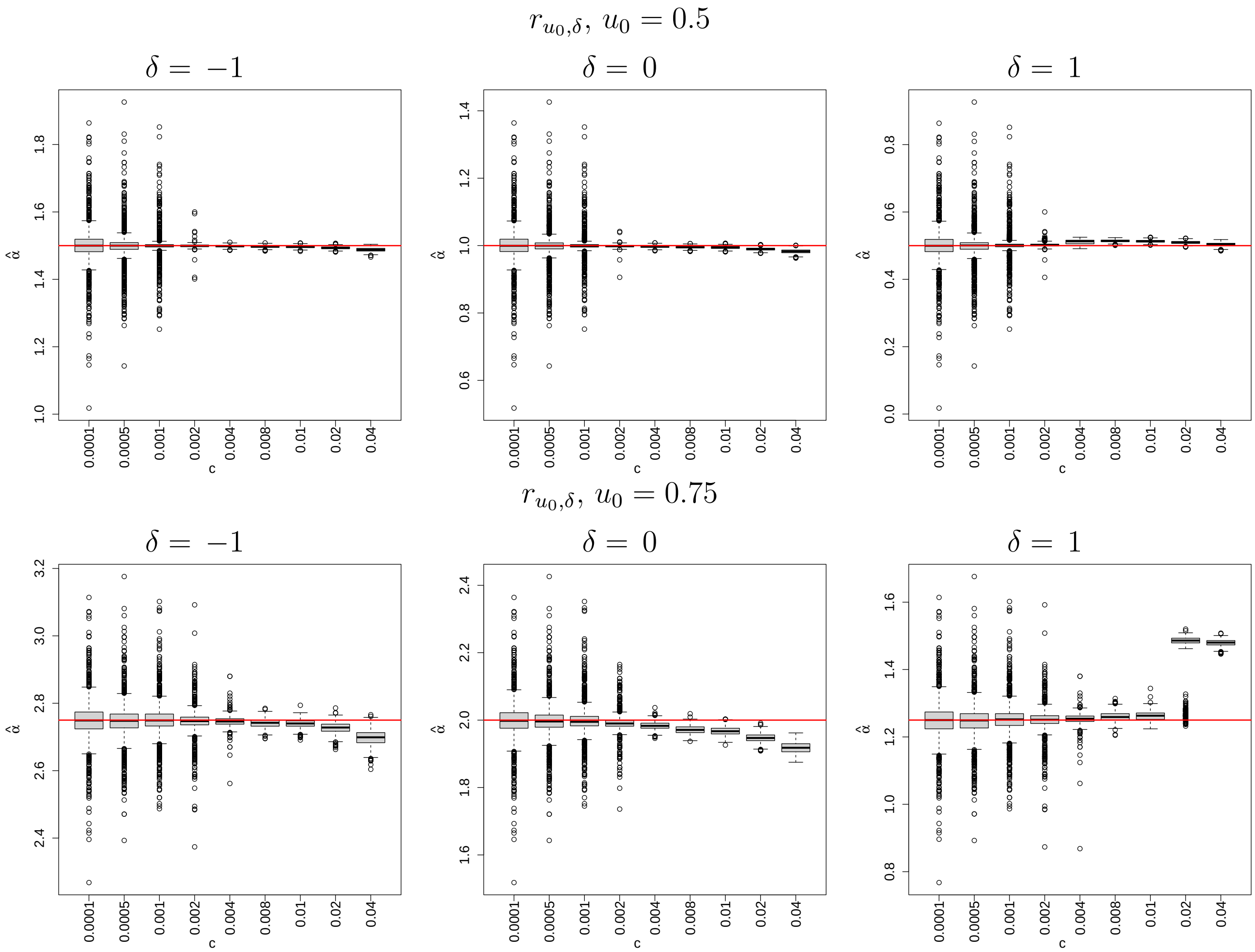

Figure S.34: Boxplots of the estimators $\widehat{\alpha}_{\widehat{u}}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.01$, when the errors have a Laplace distribution. The horizontal red line corresponds to the value α_0.

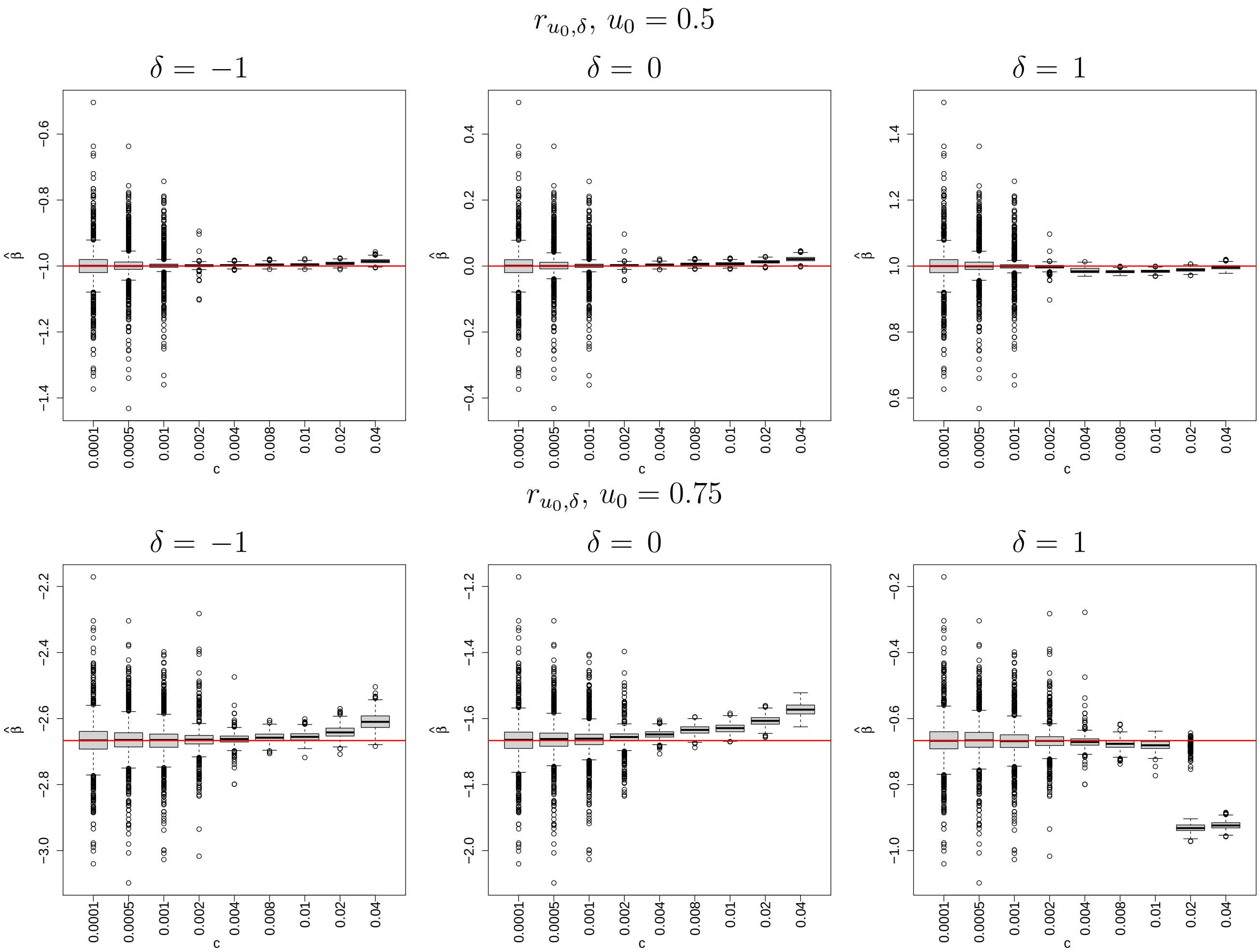

Figure S.35: Boxplots of the estimators $\widehat{\beta}_{\widehat{u}}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.01$, when the errors have a Laplace distribution. The horizontal red line corresponds to the value β_0.

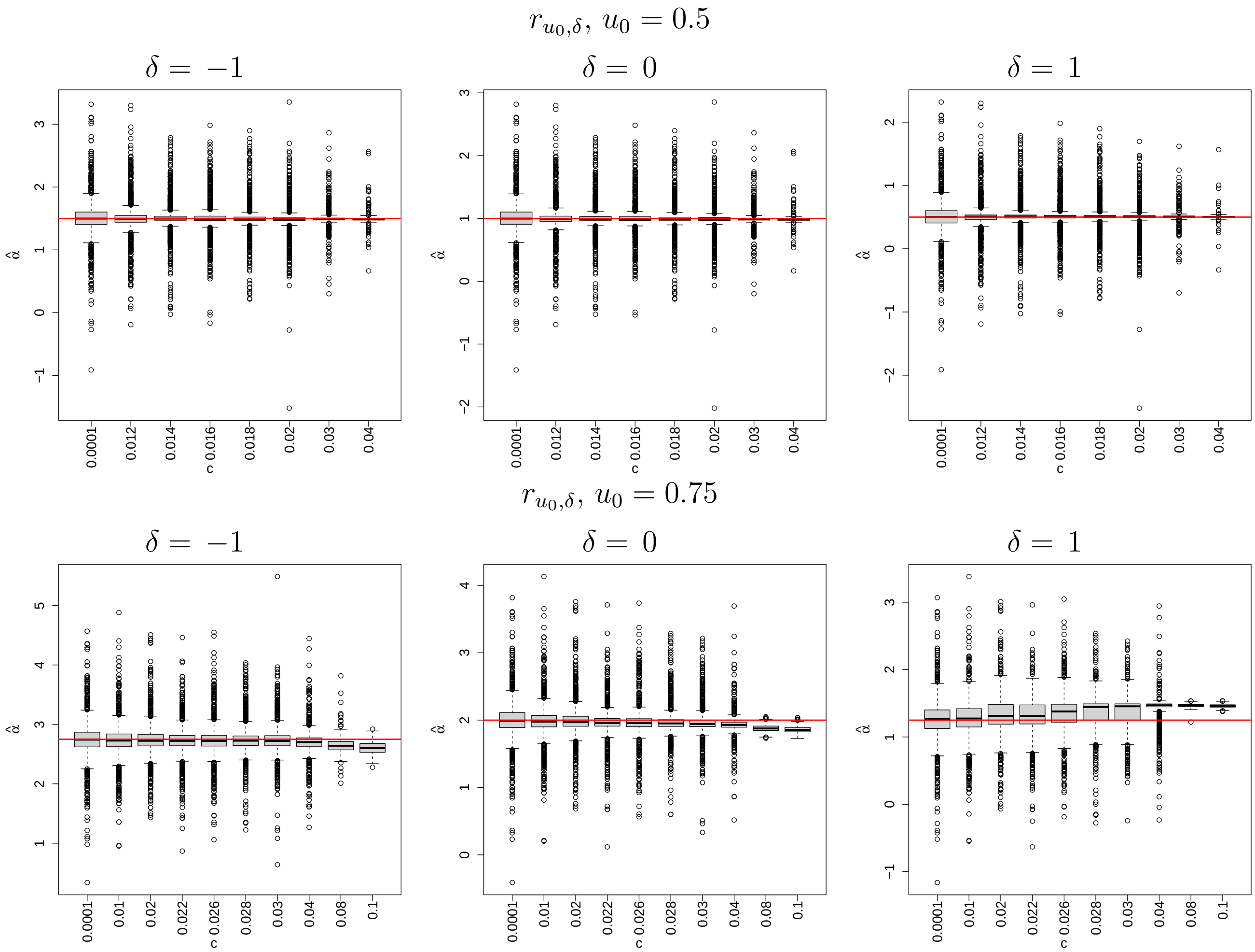

Figure S.36: Boxplots of the estimators $\widehat{\alpha}_{\widehat{u}}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.05$, when the errors have a Laplace distribution. The horizontal red line corresponds to the value α_0.

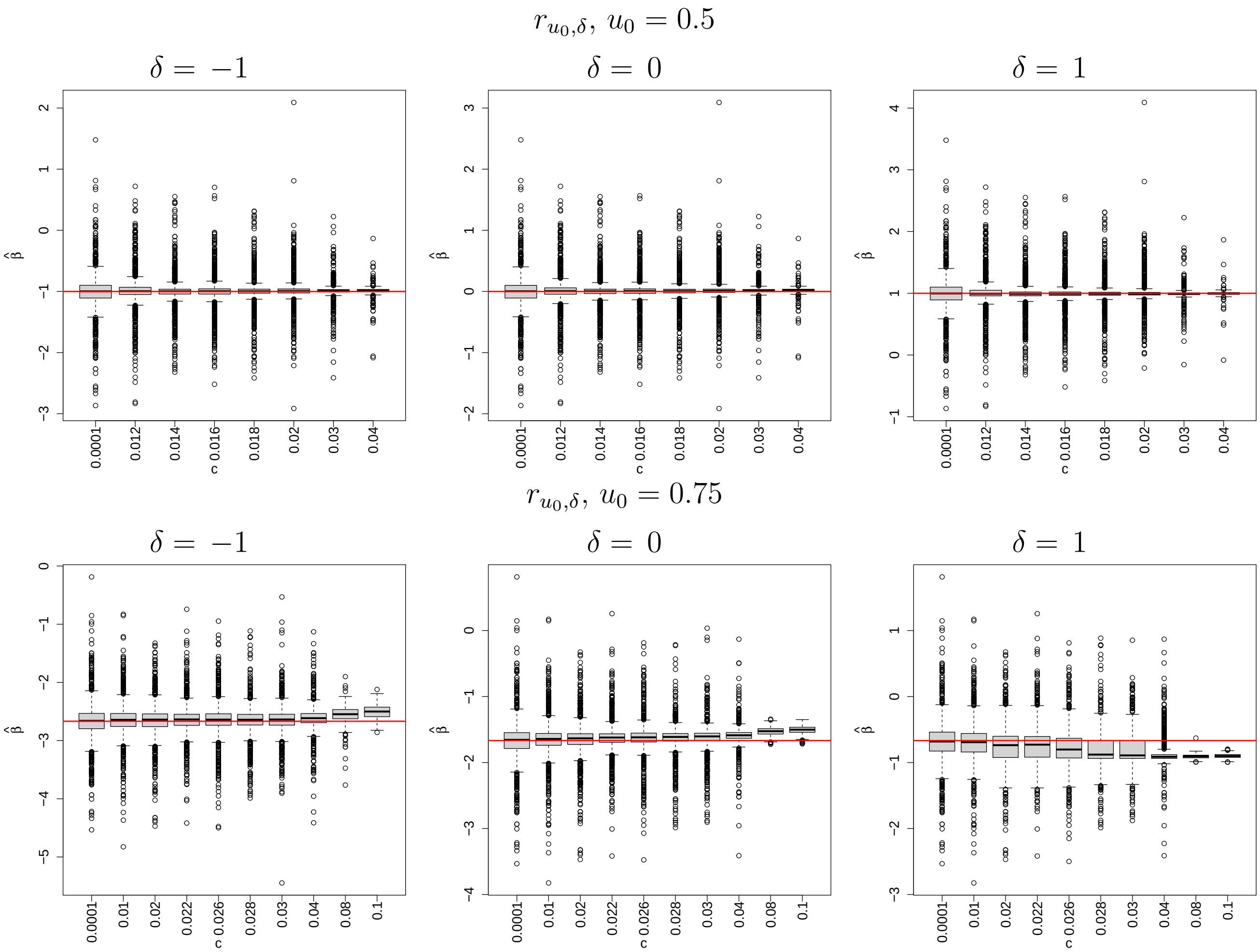

Figure S.37: Boxplots of the estimators $\widehat{\beta}_{\widehat{u}}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.05$, when the errors have a Laplace distribution. The horizontal red line corresponds to the value β_0.

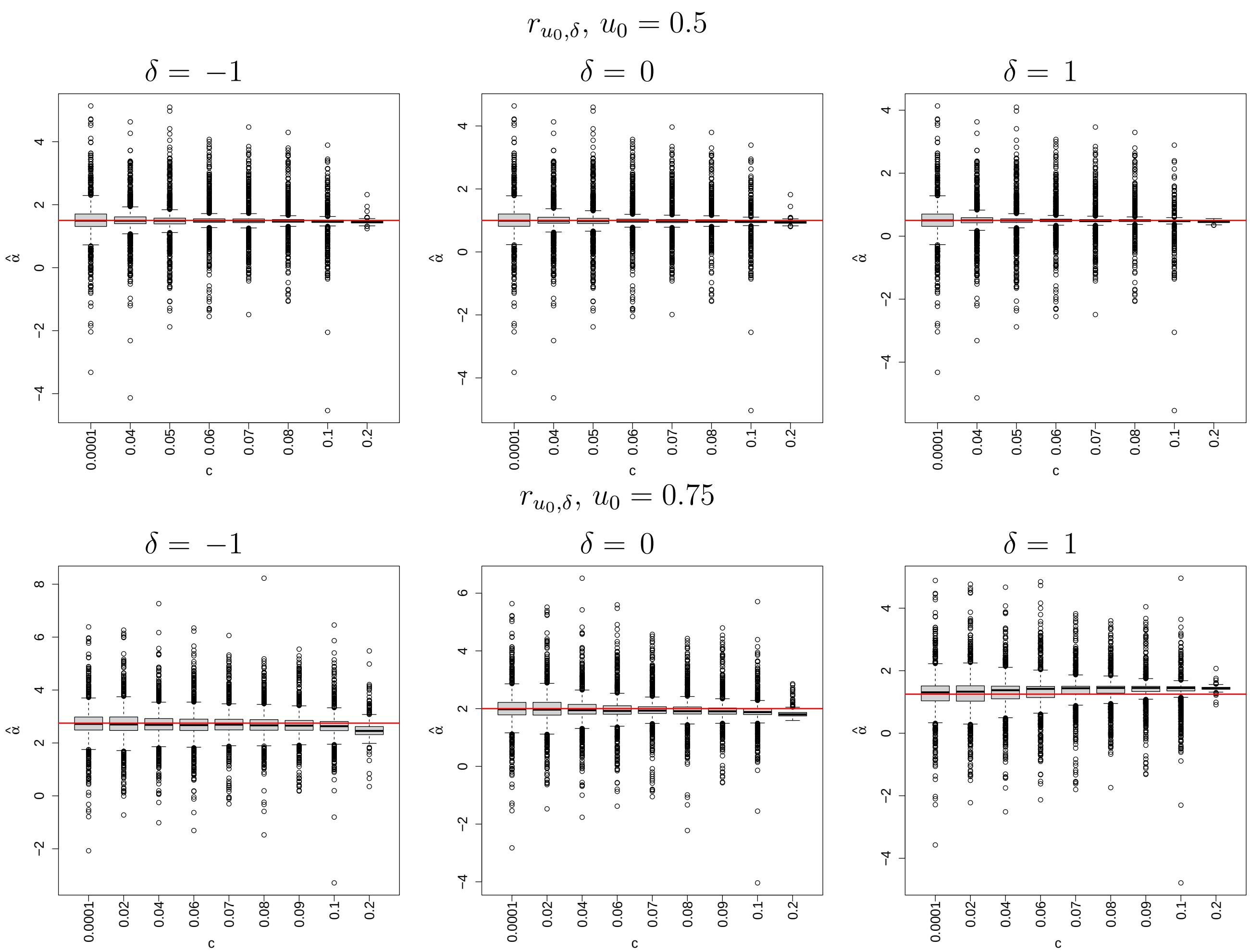

Figure S.38: Boxplots of the estimators $\widehat{\alpha}_{\widehat{u}}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.10$, when the errors have a Laplace distribution. The horizontal red line corresponds to the value α_0.

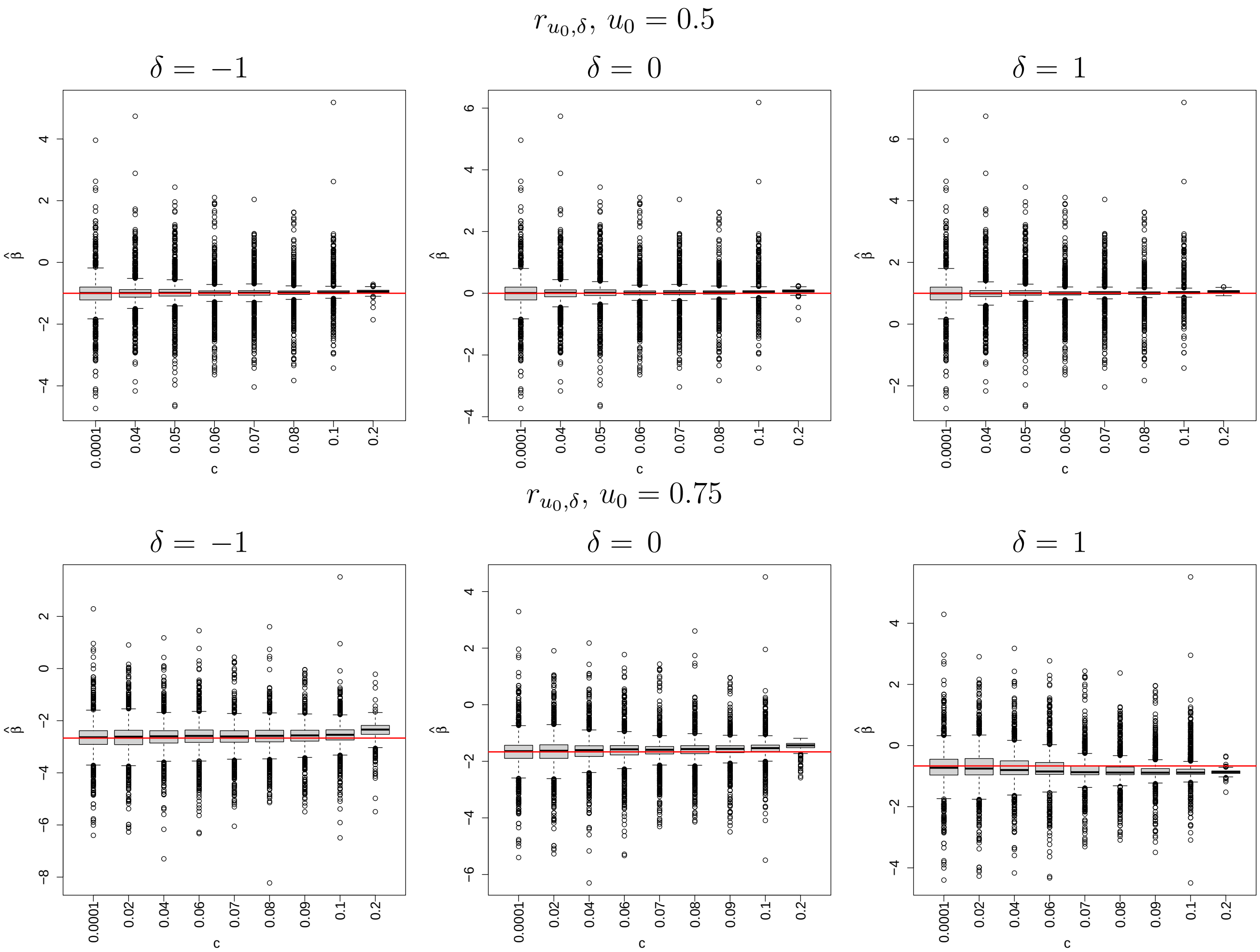

Figure S.39: Boxplots of the estimators $\widehat{\beta}_{\widehat{u}}$ for different choices of the penalizing constant c, when $n = 500$ and $\sigma = 0.10$, when the errors have a Laplace distribution. The horizontal red line corresponds to the value β_0.

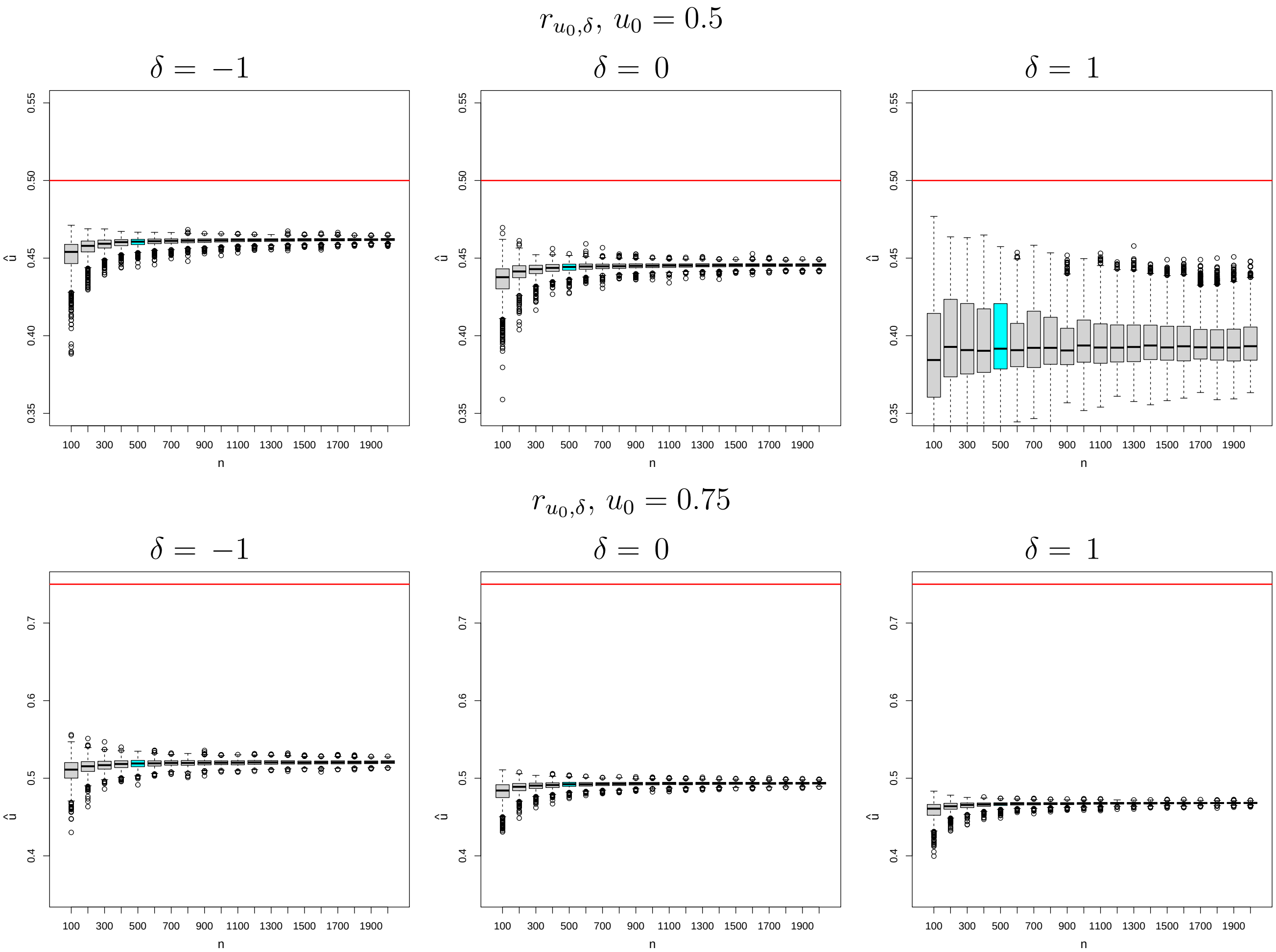

Figure S.40: Boxplots of the estimators $\widehat{u}$ computed using the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.01$, when $\varepsilon \sim \mathcal{DE}(0, \sigma^2/2)$. The horizontal red line corresponds to the true value u_0.

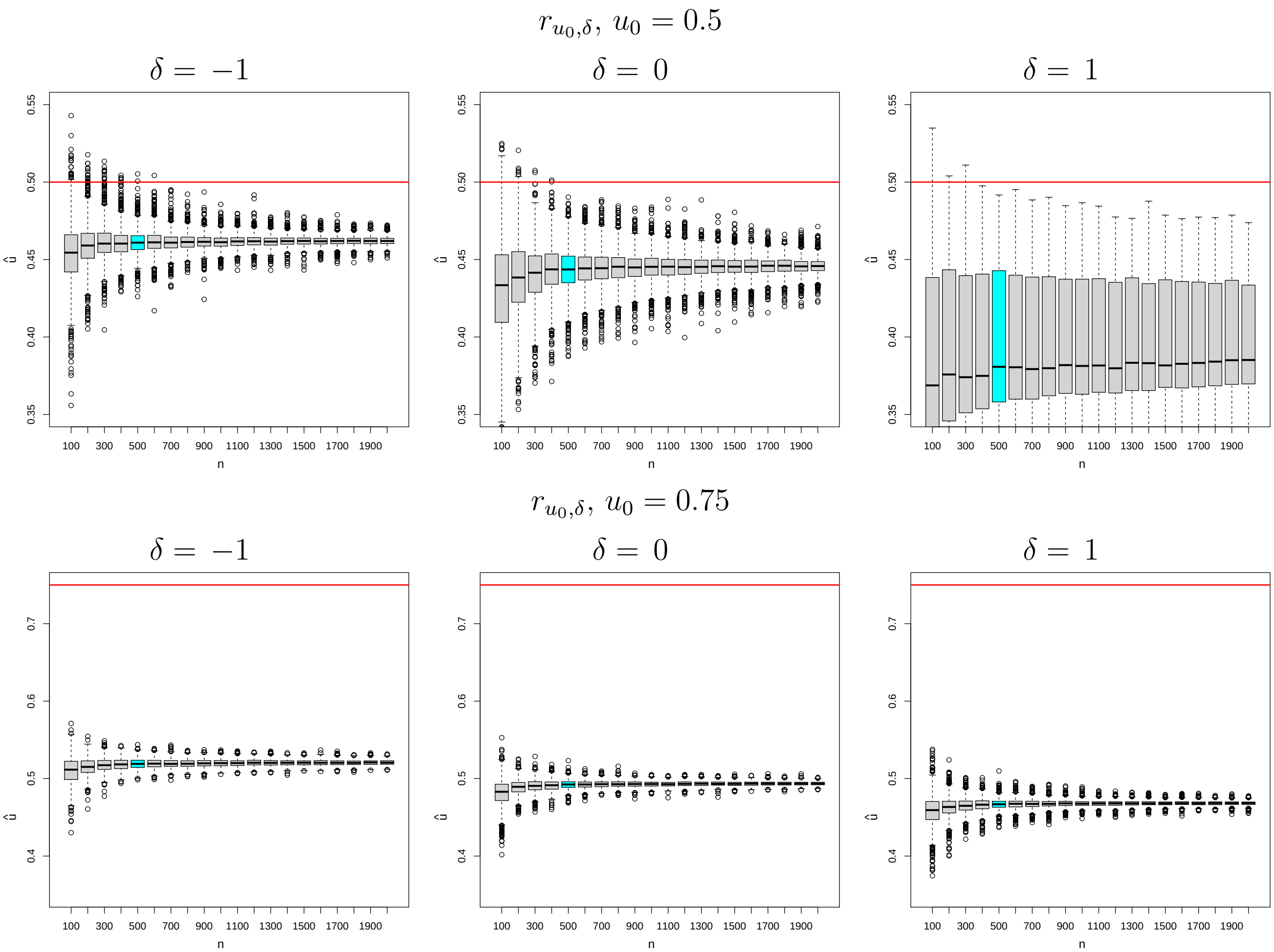

Figure S.41: Boxplots of the estimators $\widehat{u}$ computed using the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.05$, when $\varepsilon \sim \mathcal{DE}(0, \sigma^2/2)$. The horizontal red line corresponds to the true value u_0.

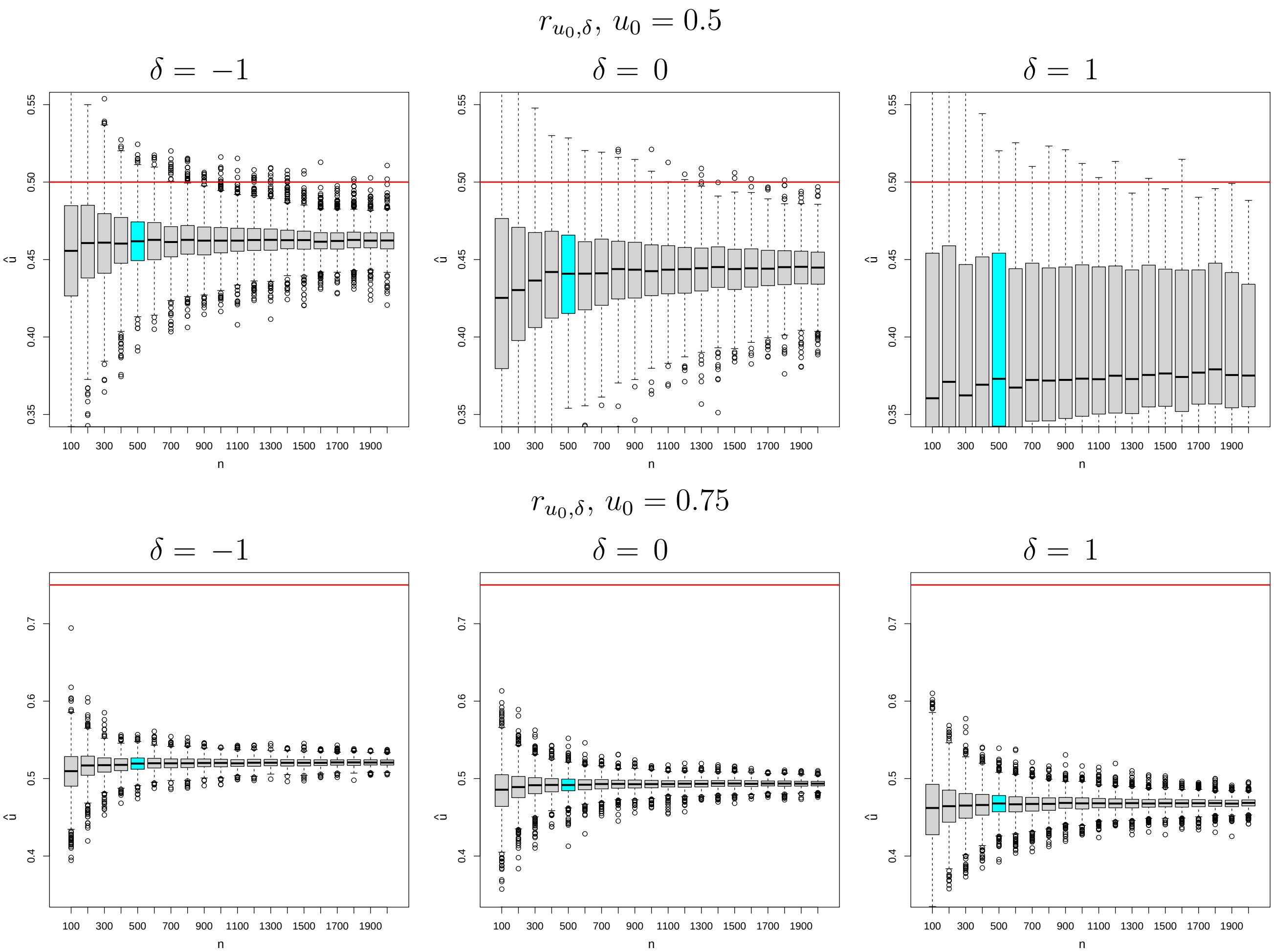

Figure S.42: Boxplots of the estimators $\widehat{u}$ computed using the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.10$, when $\varepsilon \sim \mathcal{DE}(0, \sigma^2/2)$. The horizontal red line corresponds to the true value u_0.

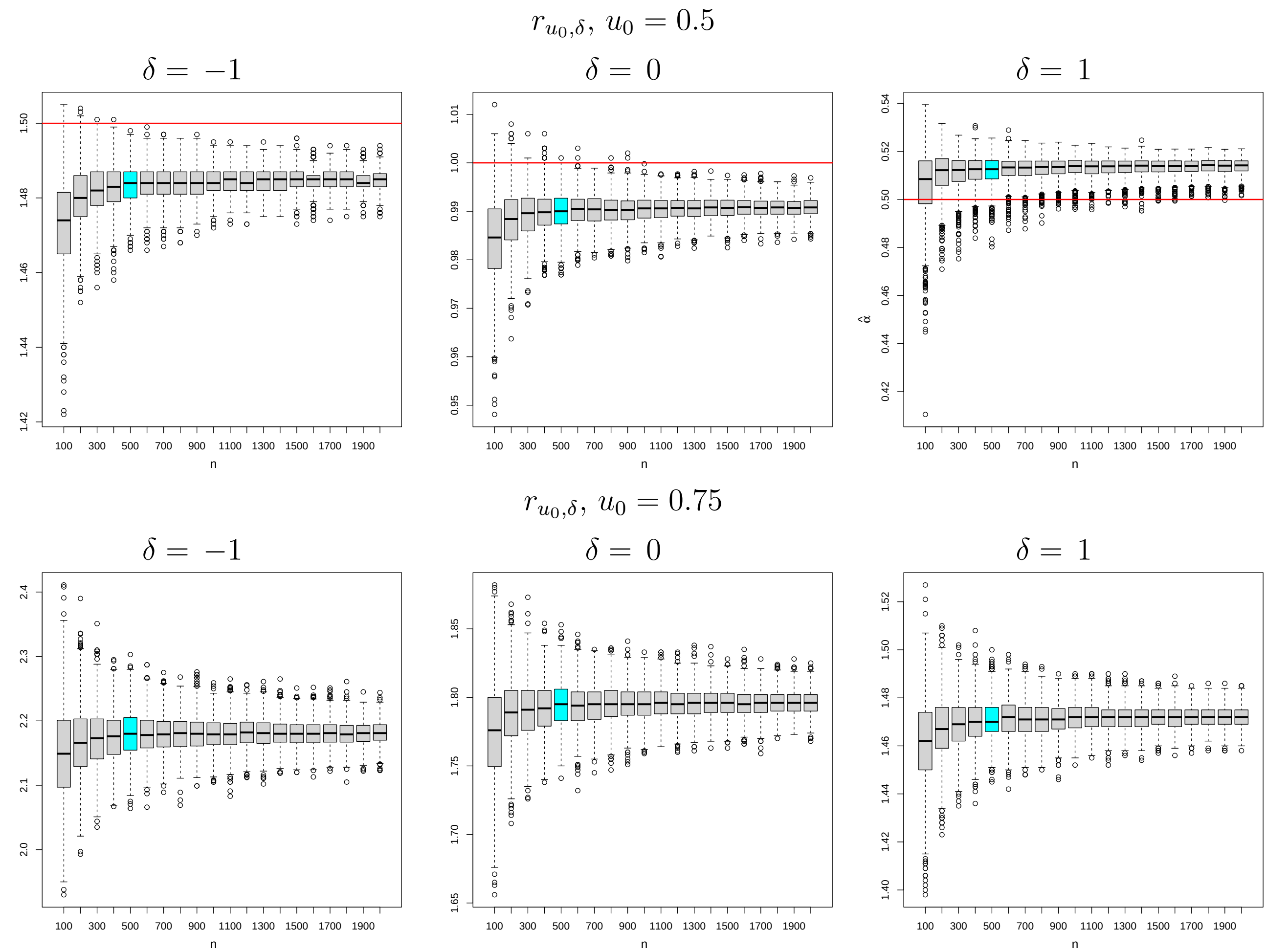

Figure S.43: Boxplots of the estimators $\widehat{\alpha}$ obtained with the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.01$, when $\varepsilon \sim \mathcal{DE}(0, \sigma^2/2)$. The horizontal red line, when present, corresponds to the value α_0.

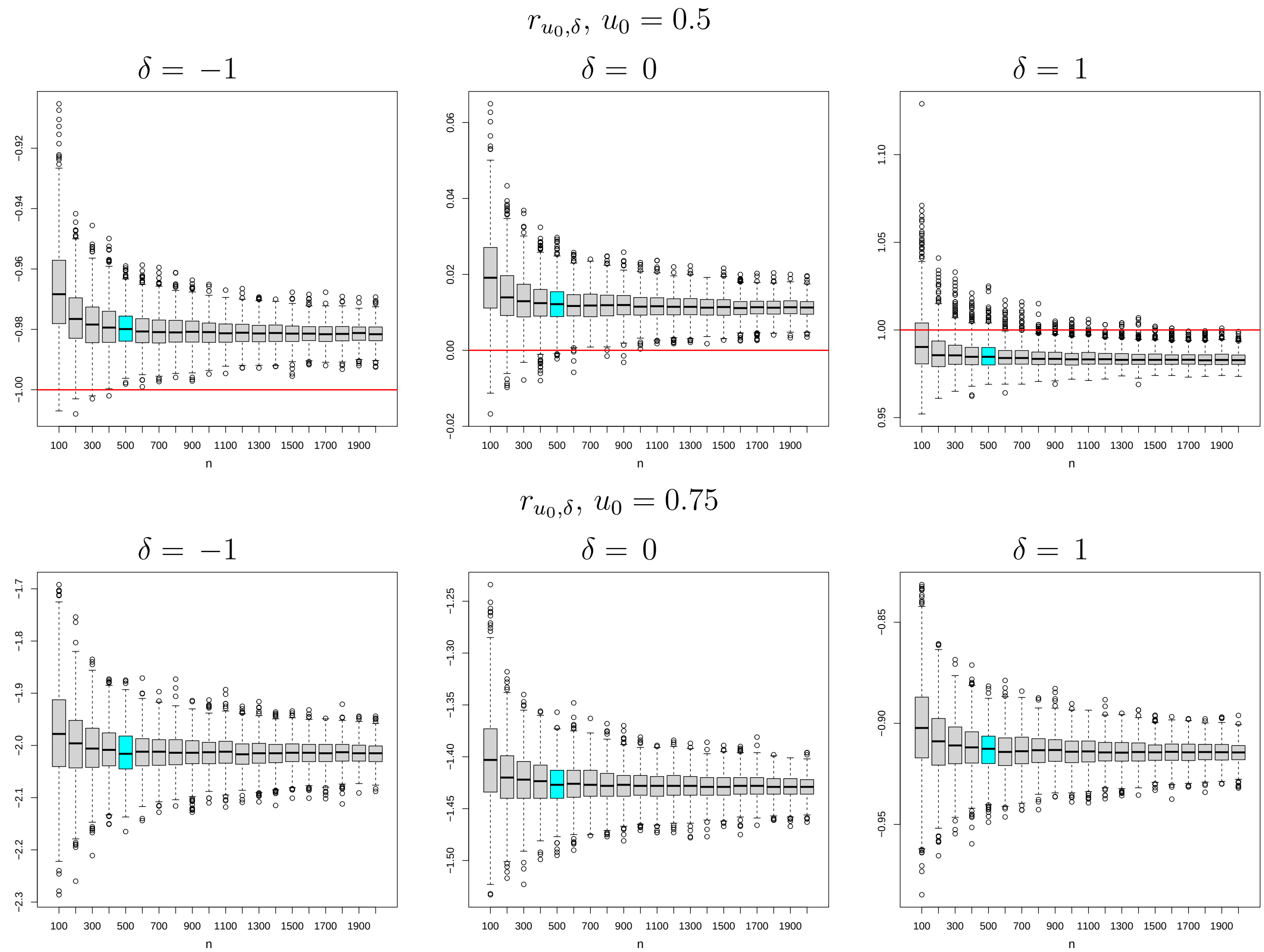

Figure S.44: Boxplots of the estimators $\widehat{\beta}$ obtained with the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.01$, when $\varepsilon \sim \mathcal{DE}(0, \sigma^2/2)$. The horizontal red line, when present, corresponds to the value β_0.

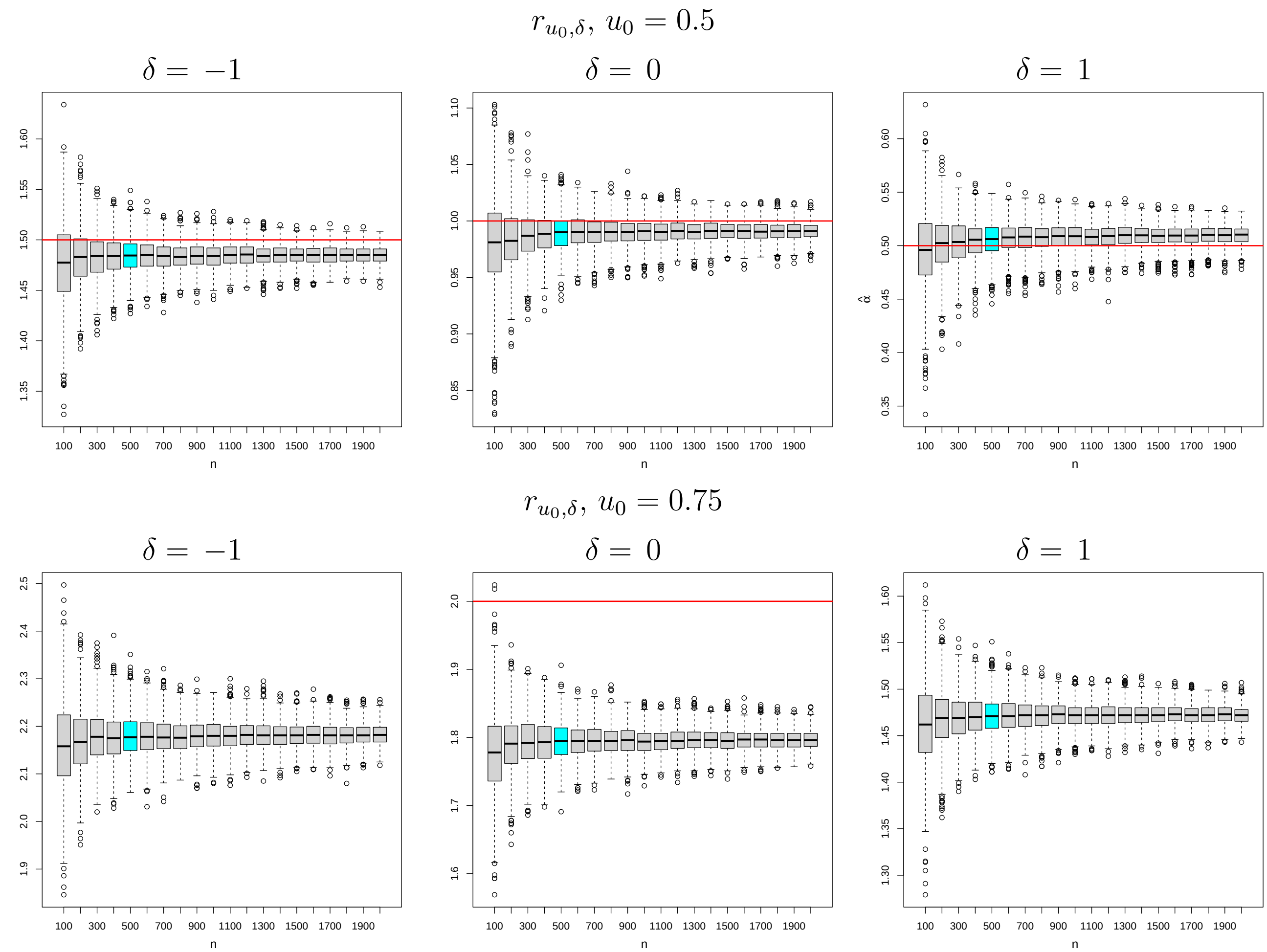

Figure S.45: Boxplots of the estimators $\widehat{\alpha}$ obtained with the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.05$, when $\varepsilon \sim \mathcal{DE}(0, \sigma^2/2)$. The horizontal red line, when present, corresponds to the value α_0.

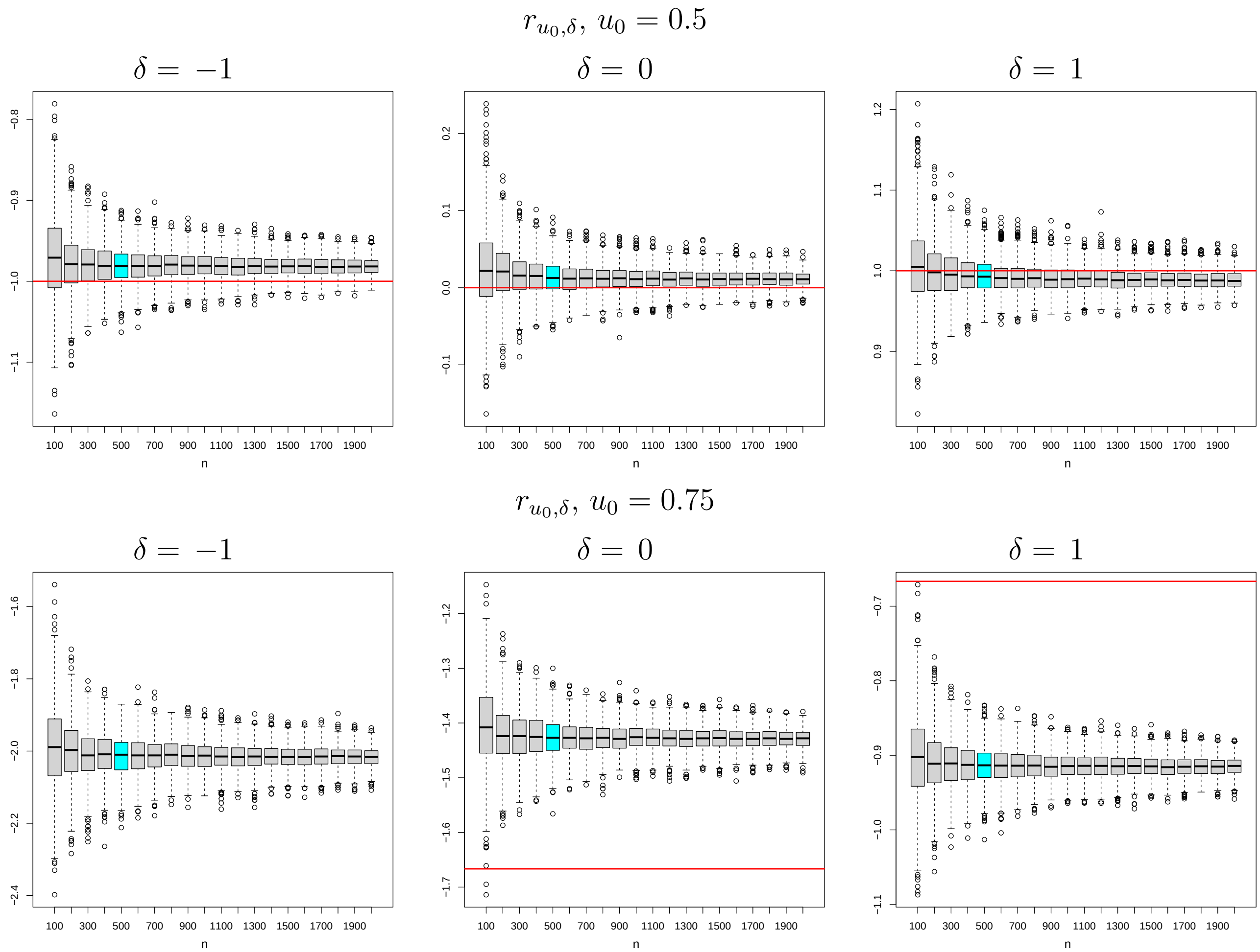

Figure S.46: Boxplots of the estimators $\widehat{\beta}$ obtained with the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.05$, when $\varepsilon \sim \mathcal{DE}(0, \sigma^2/2)$. The horizontal red line, when present, corresponds to the value β_0.

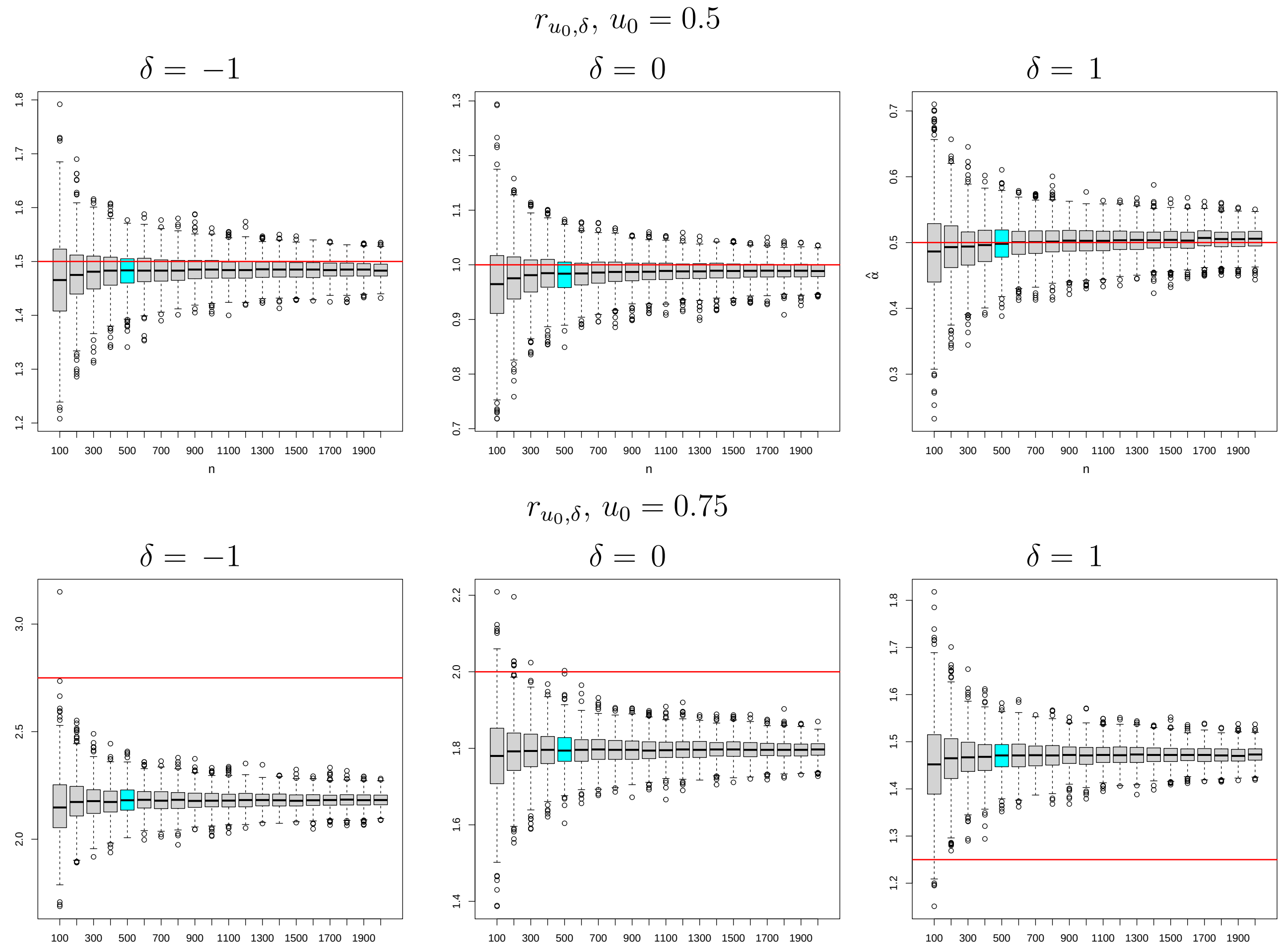

Figure S.47: Boxplots of the estimators $\widehat{\alpha}$ computed using the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.10$, when $\varepsilon \sim \mathcal{DE}(0, \sigma^2/2)$. The horizontal red line corresponds to the value α_0.

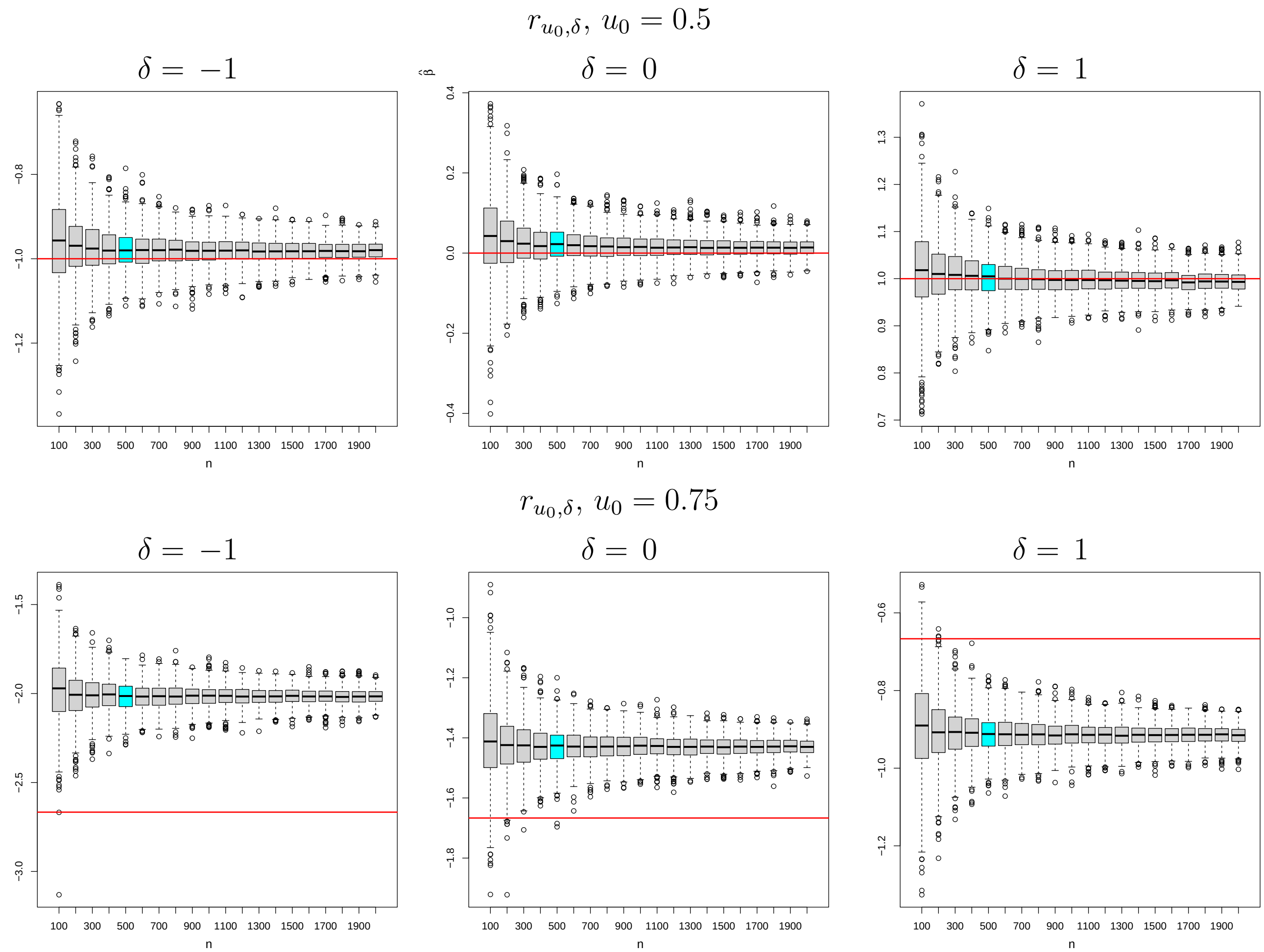

Figure S.48: Boxplots of the estimators $\widehat{\beta}$ computed using the method implemented by Zeileis et al. (2002) for different choices of sample size and $\sigma = 0.10$, when $\varepsilon \sim \mathcal{DE}(0, \sigma^2/2)$. The horizontal red line corresponds to the value β_0.

We now summarize the results obtained when a piecewise polynomial with degree 3 is considered in the function `breakpoints`. As in Section S.2.2, we choose $n = 500$ and $u_0 = 0.5$. In all figures, we display parallel boxplots as σ varies. As mentioned in Section S.2.2, since the true regression function above the threshold equals $r(x) = \alpha_0 + \beta_0 x$ and a third degree piecewise polynomial is fitted, the coefficients corresponding to the fitted polynomial above the estimated change point $\widehat{u}$ are denoted as $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\beta}_2$ and $\widehat{\beta}_3$, where the two last ones correspond to the estimates associated to x^2 and x^3, whose true value is 0 in the considered model.

Figure S.49 presents the boxplots of the threshold estimators, while Figure **??** displays the boxplots of the estimators $\widehat{\alpha}$ and $\widehat{\beta}$, the plots in cyan correspond to the boxplots within a reduced range to facilitate the comparison with those corresponding to our proposal. As in Section S.2.2, the boxplots of $\widehat{u}$ are almost centered when $\sigma = 0.01$. However, as revealed in Table S.1, when $\sigma = 0.05$ and $\delta = -1$ as well as when $\sigma = 0.10$ for any value of δ, the distances between the true value and the median are slightly larger than when the errors are normally distributed. Regarding the estimators of α_0 and β_0, their behaviour is similar to the oner described in Section S.2.2.

	Normal Errors			Laplace Errors		
σ	δ			δ		
	-1	0	1	-1	0	1
0.01	0.0008	0.0000	-0.0032	0.0012	-0.0028	-0.0024
0.05	0.0184	-0.0046	-0.0288	0.0209	-0.0031	-0.0261
0.10	0.0147	-0.0251	-0.0399	0.0235	-0.0305	-0.0413

Table S.1: Values of $\underset{1\le s\le Nrep}{\text{median}}(\widehat{u}_s) - u_0$, where $\widehat{u}_s$ stands for the threshold estimate obtained at the $s-$th replication, when $u_0 = 0.5$, $n = 500$ and a third degree piecewise polynomial is considered.

Figure **??** displays the boxplots of the estimators $\widehat{\beta}_2$ and $\widehat{\beta}_3$. Again, due to their large variability, the boxplots in cyan correspond to a reduced range. The conclusions given in in Section S.2.2 specially regarding the behaviour for the different values of δ and σ are still valid and, in particular, when comparing its the performance with that of our proposal. As revealed when the errors are normal, the obtained coefficients for x^2 and x^3 are in some situations very biased and may suggest a model with more coefficients than those really present.

We also want to recall that, as previously mentioned, our procedure does not assume a given parametric model for the regression function below the threshold, making it more flexible than the proposal in Zeileis et al. (2002).

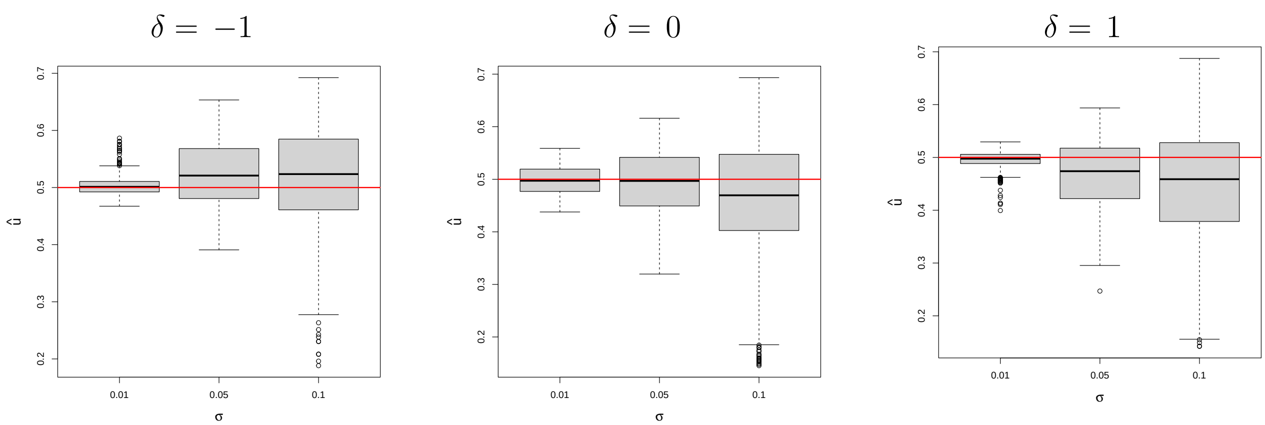

Figure S.49: Boxplots of the estimators $\widehat{u}$ obtained with the method implemented by Zeileis et al. (2002) when a third degree piecewise polynomial is considered and $\varepsilon \sim \mathcal{DE}(0, \sigma^2/2)$. The horizontal red line corresponds to the value u_0.

References

Zeileis, A., Leisch, F., Hornik, K., and Kleiber, C. (2002). strucchange: An R package for testing for structural change in linear regression models. *Journal of Statistical Software*, 7(2):1–38.

Zeileis, A., Leisch, F., Hornik, K., and Kleiber, C. (2003). Computational and visual methods for detecting structural changes in linear regression models. *Computational Statistics & Data Analysis*, 44(1-2):109–123.