Robust Box-Cox transformations for simple regression

A. Marazzi and V. J. Yohai

Abstract. The use of the Box-Cox family of transformations is a popular approach to make data behave according to a linear regression model. The regression coefficients, as well as the parameter λ defining the transformation, are generally estimated by maximum likelihood, assuming homoscedastic normal errors. These estimates are nonrobust; in addition, consistency to the true parameters holds only if the assumptions of normality and homoscedasticity are satisfied. We present here a new class of estimates, for the case of simple regression, which are robust and consistent even if the assumptions of normality and homoscedasticity do not hold.

1. Introduction

The Box-Cox family of transformations has become a widely used tool to make data behave according to a linear regression model. Sakia (1992) has given an excellent review of the work relating to this transformation. The response variable, transformed according to the Box-Cox procedure, is usually assumed to be linearly related to its covariates and the errors normally distributed with constant variance. The regression coefficients, as well as the parameter λ defining the transformation, are generally estimated by maximum likelihood (ML). Unfortunately, near normality and homoscedasticity are hard to attain simultaneously with a single transformation. In addition, the ML-estimate is not consistent under non-normal or heteroscedastic errors and it is not robust.

Carroll and Ruppert (1988) proposed bounded influence estimates based on the normal homoscedastic model to limit the influence of a moderate number of outliers. Within their approach, heteroscedasticity was modeled with the help of a weighting system. Moreover, various semiparametric and nonparametric approaches to relax the parametric structure of the response distribution have been

Received by the editors May 2, 2003.
1991 *Mathematics Subject Classification.* Primary 62J05; Secondary 62G35.
Key words and phrases. Box-Cox transformations, robust estimation, heteroscedasticity.
This work was completed with the support of Grant 2053-066895.01 from the Swiss National Science Foundation and Grant PICT–99 03–06277 from the Agencia Nacional de Promoción de la Ciencia y la Tecnología, Argentina.

studied. References concerning these proposals can be found in Foster et al. (2001), where a semiparametric procedure based on a minimum distance criterion between a semiparametric and a nonparametric estimate of the error distribution is developed. In general, these approaches do not provide effective protection against heavy contamination and heteroscedasticity.

In this paper we present a new class of estimates for the case of simple regression which are robust and consistent even if the assumptions of normality and homoscedasticity do not hold. The new estimates are based on minimization (with respect to λ) of a measure of autocorrelation among the residuals with respect to a robust estimate of the regression coefficients (for given λ). In order to compute the autocorrelation, the residuals have to be ordered according to the values of the regressor. We use the residual autocorrelation as a measure of functional relationship between the residuals and the regressor. This measure is reminiscent of a proposal by Maravall (1983) to detect nonlinearity in a time series.

2. The model

We consider a bivariate random sample $(x_1, y_1), \ldots, (x_n, y_n)$ and assume that the response and the explanatory variables are linked by the linear relationship

$$y_i^{(\lambda_0)} = \alpha_0 + \beta_0 x_i + h(x_i) u_i, \tag{2.1}$$

where α_0, β_0 and λ_0 are real parameters and $y_i > 0$. The function $h(\cdot)$ is unknown and

$$y^{(\lambda)} = \begin{cases} (y^\lambda - 1)/\lambda & \text{if} \quad \lambda \neq 0, \\ \ln(\lambda) & \text{if} \quad \lambda = 0, \end{cases} \tag{2.2}$$

denotes the usual Box-Cox transformation. We assume that the errors u_i are i.i.d. according to a common cdf F, that u_i is independent of x_i, and $E(u_i) = 0$.

3. The robust autocorrelation estimate

Suppose that (α_n, β_n) are robust estimates of intercept and slope for the simple linear regression model, e.g., MM-estimates (Yohai, 1987). Let $(\alpha_n(\lambda), \beta_n(\lambda))$ be the result of applying these estimates to the responses $y_i^{(\lambda)}$ for a given λ and $s_n(\lambda)$ a robust measure of the residual scale, e.g., the one used to compute the MM-estimate (usually, the initial high breakdown point S-estimate of scale; Rousseeuw and Leroy, 1987). If the residuals are computed using the true parameter λ_0, their conditional mean is close to zero for all values of x. On the other hand, when the residuals are computed using a $\lambda \neq \lambda_0$, there is a functional relationship between the residual conditional mean and x. A suitable value of λ should therefore minimize a measure of dependency between residuals and x. One such measure is the *robust residual autocorrelation* $\rho_n(\lambda)$. Its definition depends on whether there are tied x-values or not.

3.1. The case without tied x-values

Firstly, we suppose that all the values $x_1, \dots, x_n$ are distinct. In this case $\rho_n(\lambda)$ is given by the following steps:

1. Sort x_i, $i = 1, \dots, n$ in ascending order and let $i_1, \dots, i_n$ be the corresponding permuted indices. Thus, $x_{i_1} < x_{i_2} < \dots < x_{i_n}$.

2. For a given λ, let $r_j(\lambda) = y_{i_j}^{(\lambda)} - \alpha_n(\lambda) - x_{i_j}\beta_n(\lambda)$ $(j = 1, \dots, n)$ and

$$v_j(\lambda) = \psi\left(\frac{r_j(\lambda)}{s_n(\lambda)}\right), \quad j = 1, \dots, n.$$

3. Set

$$\rho_n(\lambda) = \frac{1}{n}\sum_{j=1}^{n-1} v_j(\lambda)v_{j+1}(\lambda), \tag{3.1}$$

where $\psi(\cdot)$ is any monotone, odd, and bounded function, e.g., the well known Huber's function.

3.2. The case with tied x-values

We now assume that there are only k distinct x-values and that n_1 observations are equal to the smallest, n_2 to the next smallest, ..., n_k to the largest. Thus, $n_1 + n_2 + \dots + n_k = n$. In this case, we modify the terms in (3.1) as follows:

1. Replace the first $n_1 - 1$ terms by

$$\frac{(n_1 - 1)}{\binom{n_1}{2}} \sum_{i=1}^{n_1-1} \sum_{j=i+1}^{n_1} v_i(\lambda)v_j(\lambda).$$

2. Replace the term $v_{n_1}(\lambda)v_{n_1+1}(\lambda)$ by

$$\frac{1}{n_1 n_2} \sum_{i=n_1+1}^{n_1+n_2} \sum_{j=1}^{n_1} v_i(\lambda)v_j(\lambda).$$

3. Replace the following $n_2 - 1$ terms by

$$\frac{(n_2 - 1)}{\binom{n_2}{2}} \sum_{i=n_1+1}^{n_1+n_2-1} \sum_{j=i+1}^{n_1+n_2} v_i(\lambda)v_j(\lambda).$$

and so on. Note that this replacement procedure is equivalent to arbitrarily permuting the tied observations and computing their correlations by averaging over the permutations.

The robust autocorrelation is positive when there is dependency between residuals and the regressor x; otherwise it is close to zero. However, the equation $\rho_n(\lambda) = 0$ may have more than one or no solutions. Therefore, we define the *robust autocorrelation estimate* of λ_0 as the global minimum of $\rho_n(\lambda)$. Another proposal for robust autocorrelation can be found in Ma and Genton, 2001.

Remark 3.1. We tried to extend the robust autocorrelation estimate to the case of several covariates. For that purpose the residuals were sorted according to the fitted values (in place of the x-values). Unfortunately, simulation results showed that the performance of this proposal is not satisfactory. The reason of this poor performance may be due to the fact that the fitted value corresponding to a value $\lambda \neq \lambda_0$ may be almost uncorrelated with the fitted value associated to $\lambda = \lambda_0$.

4. Consistency

For simplicity, we consider the model given by (2.1) and $y^{(\lambda)} = y^\lambda$. If $\lambda \neq 0$, this model is equivalent to the one given by (2.1) and (2.2). We prove consistency of the robust autocorrelation estimate supposing that $\lambda \in \Lambda = [l_1, l_2]$ with $l_1 > 0$ and $l_2 < \infty$. We make the following assumptions.

(A1) In the model (2.1), $(u_1, x_1), (u_2, x_2), \ldots, (u_n, x_n)$ are i.i.d. and u_i and x_i are independent. The distribution F of u_i is continuous, symmetric, and $E(u_i) = 0$. The distribution G of x_i is continuous.

(A2) h is continuous.

(A3) ψ is continuous, odd, monotone, and bounded.

(A4) (Uniform consistency of the estimates α_n, β_n, and s_n). For all $\lambda \in \Lambda$, there exist real numbers $\varepsilon_1(\lambda) > 0$, $s_0 > 0$ and continuous functions $\alpha(\lambda)$, $\beta(\lambda)$, $s(\lambda)$ such that $\alpha(\lambda_0) = \alpha_0$, $\beta(\lambda_0) = \beta_0$, $s(\lambda) > s_0 > 0$, and

$$\begin{aligned}
p\lim_{n\to\infty} \sup_{|\lambda^*-\lambda|<\varepsilon_1(\lambda)} |\alpha_n(\lambda^*) - \alpha(\lambda^*)| &= 0, \\
p\lim_{n\to\infty} \sup_{|\lambda^*-\lambda|<\varepsilon_1(\lambda)} |\beta_n(\lambda^*) - \beta(\lambda^*)| &= 0, \\
p\lim_{n\to\infty} \sup_{|\lambda^*-\lambda|<\varepsilon_1(\lambda)} |s_n(\lambda^*) - s(\lambda^*)| &= 0,
\end{aligned}$$

where $p\lim$ denotes convergence in probability.

(A5) (Robust identifiability condition). Let

$$r(\lambda, \alpha, \beta, u, x) = y^\lambda - \alpha - \beta x = (\alpha_0 + \beta_0 x + h(x)u)^{\lambda/\lambda_0} - \alpha - \beta x,$$

$$d(\lambda, \alpha, \beta, s, x) = E(\psi(r(\lambda, \alpha, \beta, u, x)/s)|x),$$

where x and y denote generic values of the covariate and the response. Then, for any $\lambda \neq \lambda_0$, α, β and $s > 0$, there exists an interval $I(\lambda, \alpha, \beta, s)$ such that $P(x \in I(\lambda, \alpha, \beta, s)) > 0$ and for all $x \in I(\lambda, \alpha, \beta, s)$

$$d(\lambda, a(\lambda), \beta(\lambda), s, x) \neq 0.$$

We consider the estimates:

$$\hat{\lambda}_n = \arg\min_{\lambda\in\Lambda} \rho_n(\lambda), \qquad \hat{\alpha}_n = \alpha_n(\hat{\lambda}_n), \quad \hat{\beta}_n = \beta_n(\hat{\lambda}_n).$$

Theorem 4.1. *Under (A1)-(A5), $\hat{\lambda}_n \to \lambda_0$, $\hat{\alpha}_n \to \alpha_0$, and $\hat{\beta}_n \to \beta_0$ in probability.*

The proof is given in the appendix.

Remark 4.2. Suppose that (A5) does not hold. Then, there exist values $\lambda_1 \neq \lambda_0$, α_1, β_1 and s such that, with probability one, $y^{\lambda_1} = \alpha_1 + \beta_1 x + u$, where u satisfies the condition $E(\psi_s(u)|x) = 0$ with $\psi_s(u) = \psi(u/s)$. But, according to (A1), the model (2.1) satisfies the same condition and, therefore, it is not identifiable. When $h(x) = 1$ and $\beta_0 = 0$, this condition does not hold; in this case λ_0 in model (2.1) is clearly not identifiable.

Remark 4.3. At present, we did not check that the MM-estimates and the scale estimates we are suggesting for α_n, β_n, and s_n satisfy the uniform consistency condition (A4). However, this assumption seems very plausible. The methods of Berrendero and Zamar (2003) could be used to prove this property.

5. Empirical results

Bivariate observations (x_i, y_i) were generated according to the model

$$y_i^{\lambda_0} = \alpha_0 + \beta_0 x_i + h(x_i)u_i, \quad i = 1, \dots, n,$$

where the u_i were i.i.d., such that $u_i = be_i$, $E(e_i) = 0$, the distribution of e_i was uniform $U(-0.5, 0.5)$, normal $N(0, 1)$, Student t_6, t_3, contaminated normal $0.9N(0, 1) + 0.1N(0, 25)$, or exponential, and the factor b was chosen so that MAD(u_i)=1/3. In addition, $\alpha_0 = 10$, $\beta_0 = 2$, $\lambda_0 = 0.5$, $n = 100$, $x_i = 0.2 \cdot i$ $(i = 1, \dots, n)$. Two options for $h(x)$ were used: $h(x) = 3$ (homoscedastic case) and $h(x) = x/2$ (heteroscedastic case). Each experiment was based on 1000 samples. The parameters α_0, β_0 and the function h were chosen in order to have positive responses, clear identifiability, and perceptible homo/heteroscedasticity. The choice of λ_0 was not relevant, because of the equivariance of $\hat{\lambda}_n$ $(\hat{\lambda}_n(y_i^a, \dots, y_n^a) = \hat{\lambda}_n(y_1, \dots, y_n)/a)$. Various parameter sets provided similar results.

The results are reported in Table 1 and Table 2 and include the average bias of the simulated estimates (ave=1000×mean($\hat{\lambda}_n - \lambda_0$)) the root of the mean squared error ($\sqrt{\text{mse}}$=1000 × (mean$(\hat{\lambda}_n - \lambda_0)^2)^{0.5}$) and the median absolute deviation with respect to λ_0 (mde=1000×median($\hat{\lambda}_n - \lambda_0$)/0.6745). RAC denotes the robust autocorrelation estimator defined in Section 3, where $\alpha_n(\lambda)$, $\beta_n(\lambda)$, are MM-estimators (with breakdown point 0.5 and efficiency 0.95 for normal errors, as defined in Yohai, 1987) and $s_n(\lambda)$ the robust measure of the residual scale provided by an initial S-estimate (Rousseeuw and Leroy, 1987, p. 135, with tuning constant $c = 1.547$). AC denotes the autocorrelation estimator when $\alpha_n(\lambda)$, $\beta_n(\lambda)$, and $s_n(\lambda)$ are least squares estimates. F is the estimator defined in Foster et al. (2001) and RF is a modification that uses the MM-estimates $\alpha_n(\lambda)$, $\beta_n(\lambda)$ mentioned above in place of least squares estimates. BI denotes the bounded influence estimator as defined in Carroll and Ruppert (1988, p.186, with tuning constant $a = 1.5$ and based on 10 iterations). ML denotes the maximum likelihood estimator.

Table 1. Simulation results: homoscedastic case

		RAC	AC	F	RF	BI	ML
Unif	bias	0	0	0	0	-2	0
	$\sqrt{\text{mse}}$	22	20	22	21	23	18
	mde	22	21	22	21	23	18
Gauss	bias	-1	-1	0	0	-3	0
	$\sqrt{\text{mse}}$	28	26	27	26	26	23
	mde	28	26	28	27	24	23
t6	bias	0	0	1	1	-4	0
	$\sqrt{\text{mse}}$	29	30	29	28	26	29
	mde	30	31	30	29	27	29
t3	bias	-2	-1	-1	-1	-6	1
	$\sqrt{\text{mse}}$	31	46	30	29	27	54
	mde	31	36	28	28	27	45
CntG	bias	0	-2	-1	-1	-4	0
	$\sqrt{\text{mse}}$	30	48	31	30	27	54
	mde	29	46	31	30	26	56
Exp	lme	-2	-1	-1	-1	2	-26
	$\sqrt{\text{mse}}$	28	38	24	22	26	44
	mde	27	36	23	22	27	44

Table 2. Simulation results: heteroscedastic case

		RAC	AC	F	RF	BI	ML
Unif	bias	1	2	-34	-34	-120	-136
	$\sqrt{\text{mse}}$	45	40	62	54	124	139
	mde	44	40	64	53	179	205
Gauss	bias	2	3	-33	-32	-142	-192
	$\sqrt{\text{mse}}$	46	50	62	54	146	195
	mde	46	49	61	52	207	285
t6	bias	5	6	-30	-30	-149	-222
	$\sqrt{\text{mse}}$	49	59	63	55	153	226
	mde	50	58	61	54	221	332
t3	bias	1	6	-36	-34	-158	-234
	$\sqrt{\text{mse}}$	47	97	66	56	162	252
	mde	44	71	68	56	233	365
CntG	bias	3	8	-36	-33	-146	-216
	$\sqrt{\text{mse}}$	48	122	68	58	150	242
	mde	45	93	67	58	216	339
Exp	lme	22	6	-56	-54	-171	-313
	$\sqrt{\text{mse}}$	70	80	87	74	179	319
	mde	71	72	94	82	255	467

In the cases of homoscedastic short tailed error distributions, the performances of the various estimators were similar. For long tailed distributions, BI, RAC and RF provided effective and similar protections (RF being slightly superior for exponential errors). In the heteroscedastic symmetric cases, only the autocorrelation estimates RAC and AC were not biased and the robustness of RAC was very satisfying. Unfortunately, all estimates (with the exception of AC) were strongly biased in the asymmetric heteroscedastic case.

Remark 5.1. We used FORTRAN programs loaded into S-plus to evaluate the initial estimates $\alpha_n(\lambda)$, $\beta_n(\lambda)$, and $s_n(\lambda)$ on a grid of 101 equally spaced points between 0.05 and 1.25; linear interpolation was used for other values of λ. On a 733 MHz Pentium III, the computing time (seconds) to obtain the initial estimates on the grid was 0.22 ($n = 20$), 1.60 ($n = 50$), 10.77 ($n = 100$), 78.14 ($n = 200$). The computing time for minimizing $\rho(\lambda)$ was 0.05 ($n = 20$), 0.08 ($n = 50$), 0.15 ($n = 100$), 0.27 ($n = 200$).

6. Appendix

For any $\lambda \in \Lambda$, u, u^*, x, and $\varepsilon > 0$ we define

$$t(\lambda, u, u^*, x, \varepsilon) = \inf_{D(\lambda,\varepsilon,x)} \psi(r(\lambda', \alpha', \beta', u, x)/s')\psi(r(\lambda', \alpha', \beta', u^*, x')/s'),$$

where $D(\lambda, \varepsilon, x)$ is the set of all λ', α', β', s', x' that satisfy the following conditions:

$$|\lambda' - \lambda| \le \varepsilon, |\alpha' - \alpha(\lambda')| \le \varepsilon, |\beta' - \beta(\lambda')| \le \varepsilon, |s' - s(\lambda')| \le \varepsilon, |x' - x| \le \varepsilon.$$

Lemma 6.1. *The following properties hold:*
(i) $d(\lambda, \alpha, \beta, s, x)$ *is continuous;*
(ii) $d(\lambda_0, \alpha_0, \beta_0, s, x) = E(\psi(h(x)u/s)|x) = 0$.

Proof. (i) follows from the Dominated Convergence Theorem and (ii) it is immediate.

Lemma 6.2. *Let, u, u^*, x be independent variables such that u and u^* have distribution F and x distribution G. Then, for a given $\lambda \neq \lambda_0$, there exists $\varepsilon_2(\lambda) \le \varepsilon_1(\lambda)$, such that $E(t(\lambda, u, u^*, x, \varepsilon_2(\lambda))) > 0$.*

Proof. By (A5),

$$P(E(\psi(r(\lambda, \alpha(\lambda), \beta(\lambda), u, x)/s(\lambda))|x) \neq 0) > 0$$

and then

$$E(t(\lambda, u, u^*, x, 0)) = E(E^2(\psi(r(\lambda, \alpha(\lambda), \beta(\lambda), u, x)/s(\lambda))|x)) > 0.$$

The Lemma follows from the continuity of $t(\lambda, u, u^*, x, \varepsilon)$ and the Dominated Convergence Theorem.

Lemma 6.3. *Let $x_1, x_2, \ldots, x_n$ be a random sample of a distribution G, and let $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$ be the ordered sample. Define $d_i = x_{(i+1)} - x_{(i)}$, $i = 1, \ldots, n-1$ and $m_n(\varepsilon) = \#\{i : d_i > \varepsilon\}/n$. Then, for any $\varepsilon > 0$, $m_n(\varepsilon) \to 0$ a.s. as $n \to \infty$.*

Proof. Take $\delta > 0$ and let M such that $P(|x| > M) < \delta$. Then

$$\begin{aligned} m_n(\varepsilon) &\leq \frac{\#\{i : d_i > \varepsilon, |x_i| > M\}}{n} + \frac{\#\{i : d_i > \varepsilon, |x_i| \leq M\}}{n} \\ &\leq \frac{\#\{i : |x_i| > M\}}{n} + \frac{1}{n}\frac{2M}{\varepsilon}. \end{aligned}$$

Therefore,

$$\limsup_{n\to\infty} m_n(\varepsilon) \leq \lim_{n\to\infty} \frac{\#\{i : |x_i| > M\}}{n} < \delta.$$

Since this inequality holds for any $\delta > 0$, the Lemma follows.

Proof of Theorem 4.2

To prove that $p\lim_{n\to\infty} \hat{\lambda}_n = \lambda_0$, it will be enough to show that:
(a) If $\delta > 0$ and $L = \{|\lambda - \lambda_0| > \delta\}$, then there exists $z > 0$, such that

$$p\lim \inf_{n\to\infty} \inf_{\lambda\in L} \rho_n(\lambda) > z;$$

(b) $p\lim \rho_n(\lambda_0) = 0$.
We start proving (a). Since we are assuming that the distribution G of x is continuous we can assume that all the x_i's are different. We can write

$$\begin{aligned} &\rho_n(\lambda) \\ = &\frac{1}{n}\sum_{i=1}^{n-1} \psi\left(\frac{r(\lambda, \alpha_n(\lambda), \beta_n(\lambda), u_{(i)}, x_{(i)})}{s_n(\lambda)}\right) \psi\left(\frac{r(\lambda, \alpha_n(\lambda), \beta_n(\lambda), u_{(i+1)}, x_{(i+1)})}{s_n(\lambda)}\right) \end{aligned}$$

where $x_{(i)} = x_{i_i}$, $u_{(i)} = u_{i_i}$ and $i_1, i_2, \ldots, i_n$ are the indices such that $x_{i_1} < x_{i_2} < \ldots < x_{i_n}$. For any $\lambda \neq \lambda_0$, let $\varepsilon_2(\lambda)$ be defined as in Lemma 6.2. According to the Heine-Borel theorem, we can find $\lambda_1, \ldots, \lambda_k$ in L, such that $L \subset \cup_{j=1}^k L_j$, where $L_j = \{\lambda : |\lambda - \lambda_j| < \varepsilon_2(\lambda)\}$. It is therefore enough to show that there exist numbers $z_j > 0$, $j = 1, \ldots, k$, such that

$$p\lim \inf_{n\to\infty} \inf_{\lambda\in Lj} \rho_n(\lambda) > z_j, \quad j = 1, \ldots, k.$$

Let $I = \{i : x_{(i+1)} - x_{(i)} > \varepsilon_2(\lambda)\}$ and $m_n = \#I/n$. In addition, for a given j let

$$\begin{aligned} H_{n1} &= \left\{ \sup_{|\lambda-\lambda_j|\leq\varepsilon_1(\lambda_j)} |\alpha_n(\lambda) - \alpha(\lambda)| < \varepsilon_2(\lambda_j) \right\}, \\ H_{n2} &= \left\{ \sup_{|\lambda-\lambda_j|\leq\varepsilon_1(\lambda_j)} |\beta_n(\lambda) - \beta(\lambda)| < \varepsilon_2(\lambda_j) \right\}, \\ H_{n3} &= \left\{ \sup_{|\lambda-\lambda_j|\leq\varepsilon_1(\lambda_j)} |s_n(\lambda) - s(\lambda)| < \varepsilon_2(\lambda_j) \right\}, \end{aligned}$$

and

$$H_n = \cap_{l=1}^3 H_{nl}^C.$$

Then, by (A4)

$$\lim_{n\to\infty} P(H_n) = 1. \tag{6.1}$$

In H_n, we have:

$$\begin{aligned}
&\inf_{\lambda\in L_j} \rho_n(\lambda)\\
> \;& \frac{1}{n}\sum_{i\notin I} t(\lambda_i, u_{(i)}, u_{(i+1)}, x_{(i)}, \varepsilon_2(\lambda_j)) +\\
& \frac{1}{n}\sum_{i\in I} \psi\left(\frac{r(\lambda,\alpha_n(\lambda),\beta_n(\lambda),u_{(i)},x_{(i)})}{s_n(\lambda)}\right)\psi\left(\frac{r(\lambda,\alpha_n(\lambda),\beta_n(\lambda),u_{(i+1)},x_{(i+1)})}{s_n(\lambda)}\right)\\
\geq \;& \frac{1}{n}\sum_{i\notin I} t(\lambda_j, u_{(i)}, u_{(i+1)}, x_{(i)}, \varepsilon_2(\lambda_j)) - \frac{m_nK^2}{n}\\
\geq \;& \frac{1}{n}\sum_{i=1}^{n-1} t(\lambda_j, u_{(i)}, u_{(i+1),}x_{(i)}, \varepsilon_2(\lambda_j)) - \frac{2m_nK^2}{n}.
\end{aligned}$$

where $K = \sup\psi$. Therefore, by Lemma 6.3 and (6.1),

$$\begin{aligned}
\liminf_{n\to\infty}\inf_{\lambda\in L_j}\rho_n(\lambda) \;&>\; p\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n-1} t(\lambda_j, u_{(i)}, u_{(i+1)}, x_{(i)}, \varepsilon_2(\lambda_j))\\
&=\; p\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n-1} t(\lambda_j, u_i, u_{r_{i+1}}, x_i, \varepsilon_2(\lambda_j))
\end{aligned}\tag{6.2}$$

where $(r_1,\ldots,r_n)$ is the inverse permutation of $(i_1,\ldots,i_n)$. Since this permutation depends only on the x_i's, but not on the u_i's, we have

$$E\left(\frac{1}{n}\sum_{i=1}^{n-1} t(\lambda_j, u_i, u_{r_{i+1}}, x_i, \varepsilon_2(\lambda_i)\right) = E(t(\lambda_j, u, u^*, x, \varepsilon_2(\lambda_j)), \tag{6.3}$$

where u, u^* and x are independent random variables, the first two with distribution F and the third with distribution G. In addition,

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n-1} t(\lambda_j, u_i, u_{r_{i+1}}, x_i, \varepsilon_2(\lambda_j))\right)\\
=\;& \frac{n-1}{n^2}\,[\mathrm{Var}(t(\lambda_j, u, u^*, x, \varepsilon_2(\lambda_j))\\
&+\mathrm{Cov}(t(\lambda_j, u, u^*, x, \varepsilon_2(\lambda_j))t(\lambda_j, u^*, u^{**}, x^*, \varepsilon_2(\lambda_j)))]
\end{aligned}$$

where u, u^*, u^{**}, x, x^* are independent. Then

$$\lim_{n\to\infty}\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n-1} t(\lambda_j, u_i, u_{r_{i+1}}, x_i, \varepsilon_2(\lambda_j))\right) = 0. \tag{6.4}$$

Therefore, by (6.2), (6.3), (6.4), Chebychev's inequality, and Lemma 6.2, we get

$$p \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n-1} t(\lambda_j, u_i, u_{r_{i+1}}, x_i, \varepsilon_2(\lambda_j))) = E(t(\lambda_j, u, u^*, x, \varepsilon_2(\lambda_j)) > 0,$$

which proves (a).

The proof of part (b) is similar to the proof of part (a). The main difference is that we now use Lemma 6.1 (ii) instead of Lemma 6.2. This proves $p \lim_{n\to\infty} \hat{\lambda}_n = \lambda_0$. Using (A4) we get $p \lim_{n\to\infty} \hat{\alpha}_n = \alpha_0$ and $p \lim_{n\to\infty} \hat{\beta}_n = \beta_0$.

References

[1] Berrendero J.R., Zamar R.H. (2003). On the uniform asymptotic normality of robust estimates. Draft.

[2] Carrol R.J., Ruppert D. (1988). *Transformation and weighting in regression.* Chapman and Hall, New York.

[3] Foster A.M., Tial L., Wei L.W. (2001). Estimation for the Box-Cox transformation model without assuming parametric error distribution. *Journal of the American Statistical Association,* 96, 1097-1101.

[4] Ma Y., Genton M.G.(2001). Highly robust estimation of the autocovariance function. *Journal of Time Series Analysis*, 78, 663-684.

[5] Maravall A. (1983). An application of nonlinear time series forecasting. *Journal of Business & Economic Statistics,* 1, 66-74.

[6] Rousseeuw P. J., Leroy A.M. (1987). *Robust regression and outlier detection.* Wiley, New York.

[7] Sakia, R.M. (1992). The Box-Cox transformation technique: a review. *The Statistician,* 41, 169-178.

[8] Yohai V. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics,* 15, 642-656.

Acknowledgment

Many thanks to Danielle Meylan and Alex Randriamiharisoa for their programming support.

Université de Lausanne, Institut de médecine sociale et préventive, Bugnon 17, CH 1005 Lausanne, Switzerland
E-mail address: alfio.marazzi@unil.ch

Universidad de Buenos Aires, Departamento de Matematica, Ciudad Universitaria, Pabellon 1, 1428 Buenos Aires, Argentina
E-mail address: vyohai@dm.com.ar