



HAL
open science

Laplace's method and BIC model selection for least absolute value criterion

Jean-Marc Bardet

► **To cite this version:**

Jean-Marc Bardet. Laplace's method and BIC model selection for least absolute value criterion. Statistics and Probability Letters, 2022, 195, 10.1016/j.spl.2022.109764 . hal-03779838

HAL Id: hal-03779838

<https://hal.science/hal-03779838v1>

Submitted on 18 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Laplace's method and BIC model selection for least absolute value criterion

Jean-Marc Bardet, bardet@univ-paris1.fr

SAMM, Université Paris 1, 90 rue de Tolbiac, 75013 Paris, FRANCE

September 17, 2022

Abstract

In this paper, we provide an answer to the following question: in the particular case of a non-differentiable likelihood, is the formula for the BIC model selection criterion the same? More precisely, we obtain the Laplace method for a sum-of-absolute-values function and we deduce that the usual BIC formula with penalty in $\log(n)$ remains the same in the context of a selection of explanatory variables by least absolute value regression.

Keywords: Laplace's method; BIC criterion; Least absolute values regression; Model selection; Robust statistics

1 Introduction

Let us consider here the classical framework of the observation of a sample $Y := (Y_1, \dots, Y_n)$, a family of random variables, and a finite family \mathcal{M} of possible models m for Y . The approach proposed by Schwarz in [19] leading to the BIC model selection procedure is one of the most fruitful answer to this problem. It is a criterion of model selection used in many fields of statistics, see for example a general presentation in the book [14]. Moreover, it has the particularity compared to the AIC or Cp criteria to be asymptotically consistent under certain conditions. A first proof in linear least squares regression can be found in [15], recent results on general time series were obtained in [3].

The BIC criterion consists on choosing the most probable model after observing Y , *i.e.*

$$\hat{m}_B = \operatorname{argmax}_{m \in \mathcal{M}} \{P(m | Y)\}. \quad (1.1)$$

In the sequel we will assume that all models have the same probability of occurring, which implies that we can write $P(m | Y) = |\mathcal{M}|^{-1} \frac{P(Y|m)}{P(Y)}$, with Bayes formula and using the notation $|A| = \operatorname{Card}(A)$ when $\operatorname{Card}(A) < \infty$.

In a parametric framework where for any models m there exists a parameter $\theta_m \in \Theta_m \subset \mathbb{R}^{|m|}$ identifying m , this induces (see more details in [18])

$$P(Y | m) = \int_{\Theta_m} L_n(Y, \theta_m) d\lambda_{|m|}(\theta_m) \quad \text{for any } m \in \mathcal{M},$$

where $L_n(Y, \theta_m)$ is the likelihood of Y given $\theta \in \Theta_m$ and $\lambda_{|m|}$ is the Lebesgue measure on $\mathbb{R}^{|m|}$. As a consequence maximizing $P(m|Y)$ is equivalent to minimize

$$\widehat{S}_n(m, Y) = -2 \log \left(\int_{\Theta_m} \exp \left(\log (L_n(Y, \theta_m)) \right) d\lambda_{|m|}(\theta_m) \right). \quad (1.2)$$

Assume now that the log-likelihood $\ell_m(\theta)$ of (Y_1, \dots, Y_n) defined from the model m can be expressed. Since $\exp \left(\log (L_n(Y, \theta_m)) \right) = \exp \left(-n \left(-\frac{1}{n} \ell_n(\theta_m) \right) \right)$ and $-\frac{1}{n} \ell_n(\theta_m)$ is almost surely a converging function of θ under classical assumptions, a Laplace's method can be used for obtaining the asymptotic behavior of $\int_{\Theta_m} \exp \left(-n \left(-\frac{1}{n} \ell_n(\theta_m) \right) \right) d\theta_m$ when $\theta_m \in \Theta_m \mapsto \ell_n(\theta_m)$ is sufficiently often differentiable.

More precisely, in [10], p.480, under assumptions such as $\theta_m \in \Theta_m \mapsto \ell_n(\theta_m)$ is six times continuously differentiable, the following expansion obtain from Laplace's method has been established:

$$\begin{aligned} & \int_{\Theta_m} \exp (\ell_n(\theta_m)) d\lambda_{|m|}(\theta_m) \\ &= \left(\frac{2\pi}{n} \right)^{|m|/2} \exp (\ell_n(\widehat{\theta}_m)) \left(\det \left(-\frac{1}{n} \frac{\partial^2}{\partial \theta_m^2} \ell_n(\widehat{\theta}_m) \right) \right)^{-1/2} \left(1 + O(1/n) \right) \quad a.s. \end{aligned} \quad (1.3)$$

where $\widehat{\theta}_m = \operatorname{argmax}_{\theta_m \in \Theta_m} \{ \ell_n(\theta_m) \}$, the usual maximum likelihood estimator, is supposed to be unique.

Using the consistency of $\widehat{\theta}_m$ this induces:

$$\widehat{S}_n(m, Y) = -2 \ell_n(\widehat{\theta}_m) + |m| \log(n) - 2 \pi |m| + \log (\det(I(\widehat{\theta}_m))) + O_P(n^{-1/2}), \quad (1.4)$$

where $I(\widehat{\theta}_m) = -\lim_{n \rightarrow \infty} \frac{1}{n} \frac{\partial^2}{\partial \theta_m^2} \ell_n(\widehat{\theta}_m)$, which is typically the Fisher information matrix in case of independent identically distributed random variables (see also [13] for more details).

Considering the first two terms of this development with respect to which the others are negligible, Schwarz obtained the BIC criterion for the m model which is defined by:

$$BIC(m) = -2 \ell_n(\widehat{\theta}_m) + |m| \log(n). \quad (1.5)$$

This expression of the BIC criterion is now mostly used as is, using a log-likelihood, typically Gaussian, which is not necessarily that of the sample considered. But even in this case, the asymptotic consistency property of the criterion can be established (see for example [15] in least squares regression, or [3] for affine causal time series).

The purpose of this paper is to provide the same kind of results, *i.e.* Laplace's method and BIC expansion, but in a particular case of non differentiable likelihood: the case of independent random variables distributed following a Laplace distribution. This case is very interesting in practice, because it is well known that just as the Gaussian case leads to least squares estimators, the Laplacian case leads to least absolute values estimators. Indeed, if Y_i follows a Laplace's distribution with $E[Y_i] = m_i(\alpha_m) \in \mathbb{R}$ and $E|Y_i| = \lambda_m \in (0, \infty)$ for any $i \in \mathbb{Z}$, then:

$$L_n(Y, \theta_m) = \left(\frac{\lambda_m}{2} \right)^n \exp \left(-\lambda_m \sum_{i=1}^n |Y_i - m_i(\alpha_m)| \right),$$

with $\theta_m = (\alpha_m, \lambda_m) \in \mathbb{R}^{|m|-1} \times (0, \infty)$. A particular and well known case, is the least absolute deviation regression where $m_i(\alpha_m) = \alpha_0 + \alpha_1 X_i^{(1)} + \dots + \alpha_p X_i^{(p)}$ for exogenous variables $X^{(j)}$.

The least absolute value estimation offers a very relevant alternative to the usual least squares estimation. Its two main advantages are, first, that it does not require strong moment conditions to converge or to be asymptotically Gaussian, and second, that it is significantly more robust (see for instance [9]), especially to the presence of possible outliers. It has for example been studied in the context of the detection of breaks in [4] and [2] (for shifts) and [5] (for linear models), for the estimation of parameters of time series, as in [7] (for ARMA processes) or in [1] (for affine causal processes). It also opens the way to more complex procedures, such as robust regression with Huber function (see [8]) or quantile regression (see [11]).

Thus we will start in the following section 2 by proving Laplace's formula in the case of a least absolute values likelihood. The main terms of the (1.3) expansion obtained for differentiable likelihood will also be present, except for one multiplicative term, a random variable bounded in probability. This will allow us to recover in section 3 the classical BIC formula (1.5) with penalty term in $\log(n)$ also for a likelihood of independent Laplace variables: in this case, the fact of considering a non-differentiable likelihood at any point does not affect the asymptotic behavior of this criterion. The proofs of the different results are established in section 4.

2 A Laplace formula for least absolute deviation estimators

2.1 Notation and assumptions

Consider the following notations: for $(U_n)_{n \in \mathbb{N}}$ and $(V_n)_{n \in \mathbb{N}}$ two sequences of r.v., denote:

- $(U_n) = O_P(1)$ when for any $\varepsilon > 0$, there exists $M > 0$ such as $\sup_{n \in \mathbb{N}} \mathbb{P}(|U_n| > M) \leq \varepsilon$;
- $U_n \underset{n \rightarrow \infty}{\overset{a.s.}{\sim}} V_n$ when $\frac{U_n}{V_n} \underset{n \rightarrow \infty}{\overset{a.s.}{\rightarrow}} 1$;
- $U_n \underset{n \rightarrow \infty}{\overset{P}{\sim}} V_n$ when $\mathbb{P}\left(\left|\frac{U_n}{V_n} - 1\right| \geq \varepsilon\right) \underset{n \rightarrow \infty}{\rightarrow} 0$ for any $\varepsilon > 0$.

We begin with the definition of a class of Laplace white noises:

Class LWN: *The sequence $(\varepsilon_i)_{i \in \mathbb{Z}}$ of i.i.d.r.v. is said to belong to Class LWN if there exists $\lambda > 0$ such as the density f_ε of ε_0 satisfies $f_\varepsilon(x) = \frac{1}{2} \lambda e^{-\lambda|x|}$ for any $x \in \mathbb{R}$.*

We extend this class to a much more general class of symmetric stationary processes:

Class SEN: *The stationary sequence $(\varepsilon_i)_{i \in \mathbb{N}}$ is said to belong to Class SEN (Symmetric Ergodic Noise) if:*

1. *For any $x \in \mathbb{R}$, $F_\varepsilon(-x) = 1 - F_\varepsilon(x)$ where F_ε is the probability cumulative function of ε_0 (symmetric distribution);*
2. *There exists $\rho > 0$ such as F_ε is a $\mathcal{C}^2([-r, r])$ function and $F'_\varepsilon(0) = f_\varepsilon(0)$;*

3. We have $E[|\varepsilon_0|] < \infty$, $E[\varepsilon_0] = 0$ and $\frac{1}{n} \sum_{i=1}^n \varepsilon_i \xrightarrow[n \rightarrow \infty]{a.s.} 0$.

Many examples of such symmetric noises are well known: symmetric white noise, *i.e.* (ε_i) is a sequence of independent identically distributed symmetric random variables, α -mixing or τ -weak dependent stationary sequence of centered symmetric random variables, stationary symmetric martingale differences,... Note that it is not assumed here that the sequence admits a moment of order 2: this is one of the well-known advantages of least absolute values estimation over least squares estimation.

2.2 Laplace's method for a median

We prove a first lemma which will be a key result for obtaining a Laplace's method for least absolute values:

Lemma 1. *Let $(\varepsilon_i)_{i \in \mathbb{N}}$ belong to the class SEN. Then we have:*

$$\int_{-\infty}^{+\infty} \exp\left(-\lambda \sum_{i=1}^n (|\varepsilon_i + c| - |\varepsilon_i|)\right) dc \xrightarrow[n \rightarrow \infty]{a.s.} \sqrt{\frac{\pi}{f_\varepsilon(0)}} \frac{1}{\sqrt{n}}. \quad (2.1)$$

The proof of this lemma, as well as all the other proofs, is available in section 4.

Before establishing a new Laplace's method for least absolute values contrasts from Lemma 1, we have to restrict a little the assumptions on the noise $(\varepsilon_i)_{i \in \mathbb{N}}$ by considering a subclass of Class SEN:

Class WDSN: *The stationary sequence $(\varepsilon_i)_{i \in \mathbb{N}}$ is said to belong to Class WDSN (Weak Dependent Symmetric Noise) if:*

1. $(\varepsilon_i)_{i \in \mathbb{N}}$ belongs to the Class SEN;
2. $(\varepsilon_i)_{i \in \mathbb{N}}$ is strong mixing with mixing coefficients $(\alpha_k)_k$ that satisfy $\sum_k \alpha_k < \infty$.

OR

$(\varepsilon_i)_{i \in \mathbb{N}}$ is a linear process such as there exists a white noise $(\xi_t)_{t \in \mathbb{Z}}$ satisfying $\varepsilon_t = \sum_{i=0}^{\infty} a_i \xi_{t-i}$ with $\sum_{i=0}^{\infty} |a_i|^{1/2} < \infty$ and $(a_i)_{i \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$.

Other weak or short dependent processes could also be considered but those both examples are representative and easy to be presented. The case of strong mixing processes is based on results obtained in [16], while the case of linear processes is induced by the important paper of Wu, [20]. Note that the case of long-range dependent processes could also be studied following the paper [12].

Now we consider \hat{m}_n as a median of the sample (Y_1, \dots, Y_n) . Using Lemma 1, we establish an univariate Laplace's method for least absolute values estimation, *i.e.* for \hat{m}_n :

Lemma 2. Consider $m^* \in \mathbb{R}$ and $Y_i = m^* + \varepsilon_i$ for $i \in \mathbb{N}^*$ with (ε_i) belonging to the class WDSN. Define also:

$$\hat{m}_n \in \operatorname{argmin}_{m \in \mathbb{R}} \sum_{i=1}^n |Y_i - m|. \quad (2.2)$$

Then, there exists a sequence of random variables $(A_n)_{n \in \mathbb{N}^*}$ satisfying $(A_n) = O_{\mathbb{P}}(1)$ such as:

$$\int_{-\infty}^{\infty} \exp\left(-\sum_{i=1}^n |Y_i - m|\right) dm \underset{n \rightarrow \infty}{\underset{\mathbb{P}}{\sim}} \exp\left(-\sum_{i=1}^n |Y_i - \hat{m}_n|\right) \sqrt{\frac{\pi}{f_{\varepsilon}(0)}} \frac{1}{\sqrt{n}} e^{-A_n}. \quad (2.3)$$

A straightforward application of this result can be obtain in the particular case where $Y_i = m^* + \varepsilon_i$ where (ε_i) belongs to the class LWN with a known parameter $\lambda^* > 0$. Then after a change of variable, (2.3) induces that:

$$\int_{-\infty}^{\infty} \exp\left(-\lambda^* \sum_{i=1}^n |Y_i - m|\right) dm \underset{n \rightarrow \infty}{\underset{\mathbb{P}}{\sim}} \exp\left(-\lambda^* \sum_{i=1}^n |Y_i - \hat{m}_n|\right) \sqrt{\frac{2\pi}{n}} \lambda^* e^{-\lambda^* A_n}.$$

Therefore, if we denote $\ell_n(m)$ the log-likelihood of (Y_1, \dots, Y_n) , we obtain:

$$\int_{-\infty}^{\infty} \exp(\ell_n(m)) dm \underset{n \rightarrow \infty}{\underset{\mathbb{P}}{\sim}} \exp(\ell_n(\hat{m}_n)) \sqrt{\frac{2\pi}{n}} \times (\lambda^* e^{-\lambda^* A_n}). \quad (2.4)$$

This asymptotic development is thus quite close to the general one obtained in the case of a sufficiently differentiable log-likelihood (1.3), except for the additional multiplicative random term.

2.3 Laplace's method for least absolute value regression

Now we consider a classical linear model with a random design:

1. Assume that $(X_i)_{i \in \mathbb{N}}$ is a sequence of i.i.d. \mathbb{R}^{p+1} -valued where $X_0 = (1, X_0^{(1)}, \dots, X_0^{(p)})'$ and $E(X_0 X_0') = \Gamma$ a definite positive matrix. Moreover $(X_i)_{i \in \mathbb{N}}$ is independent to $(\varepsilon_i)_{i \in \mathbb{N}}$.
2. Let $\theta^* = (\theta_0^*, \theta_1^*, \dots, \theta_p^*) \in \mathbb{R}^{p+1}$ be unknown parameters such as

$$Y_i = \theta_0^* + \sum_{j=1}^p \theta_j^* X_i^{(j)} + \varepsilon_i \quad \text{for any } i \in \mathbb{Z}. \quad (2.5)$$

Now, from an observed sample (Y_1, \dots, Y_n) , define the least absolute value regression estimation of θ^* by:

$$\hat{\theta} = {}^t(\hat{\theta}_0, \dots, \hat{\theta}_p) \in \operatorname{argmin}_{\theta \in \mathbb{R}^{p+1}} \left\{ \sum_{i=1}^n \left| Y_i - \theta_0 - \sum_{j=1}^p \theta_j X_i^{(j)} \right| \right\}. \quad (2.6)$$

Then, we can prove the following proposition:

Proposition 1. If $(Y_i)_{i \in \mathbb{N}}$ satisfies (2.5) and (ε_i) belongs to the class WDSN, then, with $\hat{\theta}$ defined in (2.6):

$$\int_{\mathbb{R}^{p+1}} \exp\left(-\sum_{i=1}^n \left| Y_i - \theta_0 - \sum_{j=1}^p \theta_j X_i^{(j)} \right|\right) d\theta_0 \dots d\theta_p \underset{n \rightarrow \infty}{\underset{\mathbb{P}}{\sim}} \exp\left(-\sum_{i=1}^n \left| Y_i - \hat{\theta}_0 - \sum_{j=1}^p \hat{\theta}_j X_i^{(j)} \right|\right) \left(\left(\frac{\pi}{f_{\varepsilon}(0)} \right)^{\frac{p+1}{2}} \det(\Gamma)^{-1/2} \right) \frac{e^{-B_n}}{n^{\frac{p+1}{2}}}, \quad (2.7)$$

where $(B_n)_{n \in \mathbb{N}^*}$ is a sequence of random variables defined in (4.16) and such as $(B_n) = O_{\mathbb{P}}(1)$.

Hence the Laplace's method for a non differentiable contrast such as least absolute deviation provides almost same type of asymptotic development as in the differentiable case (see 1.3) except for an additional random multiplicative term $\exp(B_n)$.

3 BIC for least absolute deviation regression

In this section, we return to the framework of model selection using the BIC criterion. To do so, we consider the case of the linear model (2.5) in the parametric framework where (ε_i) belongs to the Class LWN, *i.e.* (ε_i) is a sequence of centered Laplace i.i.d.r.v. Then the log-likelihood of (Y_1, \dots, Y_n) is a function $\ell_n(\theta)$ of $\theta \in \mathbb{R}^{p+1}$ and is written:

$$\ell_n(\theta) = n \log(\lambda^*/2) - \lambda^* \sum_{i=1}^n \left| Y_i - \theta_0 - \sum_{j=1}^p \theta_j X_i^{(j)} \right| \quad \text{for any } \theta \in \mathbb{R}^{p+1}.$$

Following its definition given in [19] and recalled in the introduction, the BIC of such a model is written:

$$\begin{aligned} \widehat{BIC} &= -2 \log \left(\int_{\mathbb{R}^{p+1}} \exp(\ell_n(\theta)) d\theta \right) \\ &= -2 \log \left(\int_{\mathbb{R}^{p+1}} \left(\frac{\lambda^*}{2} \right)^n \exp \left(-\lambda^* \sum_{i=1}^n \left| Y_i - \theta_0 - \sum_{j=1}^p \theta_j X_i^{(j)} \right| \right) d\theta_0 \dots d\theta_p \right). \end{aligned}$$

Therefore using Proposition 2.7, we obtain:

$$\begin{aligned} \widehat{BIC} &\underset{n \rightarrow \infty}{\underset{\mathcal{P}}{\simeq}} -2 \ell_n(\widehat{\theta}) + (p+1) \log(n) + 2 \lambda^* B_n + \log(\det(\Gamma)) + (p+1) \log(f_\varepsilon(0)/\pi) \\ &\underset{n \rightarrow \infty}{\underset{\mathcal{P}}{\simeq}} -2 \ell_n(\widehat{\theta}) + (p+1) \log(n), \end{aligned}$$

which is the usual BIC formula.

This result, the fact that we find the usual BIC formula, is also true when we consider the more realistic case where the parameter λ^* is unknown:

Proposition 2. *If $(Y_i)_{i \in \mathbb{N}}$ satisfies (2.5) and (ε_i) belongs to the class WDSN with an unknown parameter $\lambda^* > 0$, the log-likelihood $\ell_n(\theta, \lambda)$ of (Y_1, \dots, Y_n) is defined by*

$$\ell_n(\theta, \lambda) = n \log(\lambda/2) - \lambda \sum_{i=1}^n \left| Y_i - \theta_0 - \sum_{j=1}^p \theta_j X_i^{(j)} \right| \quad \text{for any } \theta \in \mathbb{R}^{p+1}, \lambda > 0. \quad (3.1)$$

With $\widehat{BIC} = -2 \log \left(\int_0^\infty \int_{\mathbb{R}^{p+1}} \exp(\ell_n(\theta, \lambda)) d\theta d\lambda \right)$ and $(\widehat{\theta}, \widehat{\lambda}) \in \underset{\theta \in \mathbb{R}^{p+1}, \lambda > 0}{\operatorname{argmax}} \{ \ell_n(\theta, \lambda) \}$, then

$$\begin{aligned} \widehat{BIC} &\underset{n \rightarrow \infty}{\underset{\mathcal{P}}{\simeq}} -2 \ell_n(\widehat{\theta}, \widehat{\lambda}) + (p+1) \log(n) + 2 \lambda^* B_n + \log(\det(\Gamma)) + (p+1) \log(f_\varepsilon(0)/\pi) \\ &\underset{n \rightarrow \infty}{\underset{\mathcal{P}}{\simeq}} -2 \ell_n(\widehat{\theta}, \widehat{\lambda}) + (p+1) \log(n), \end{aligned}$$

where $(B_n)_{n \in \mathbb{N}^*}$ is a sequence of random variables defined in (4.16) and such as $(B_n) = O_{\mathcal{P}}(1)$.

Thus the classical formula of the BIC criterion is recovered at first order in this framework of Laplace variables, *i.e.* for a contrast by least absolute deviations: its non-differentiability does not change anything.

Remark 1. *The asymptotic expansion of Proposition 2 is also valid if the noise $(\varepsilon)_i$ belongs to the WDSN class. However, we note that in this case the statistical model is no longer parametric but semi-parametric and the likelihood $\ell_n(\theta, \lambda)$ defined in (3.1) is not strictly speaking that of the model.*

4 Proofs

4.1 Proof of Lemma 1

Proof. Let $\delta > 0$. Using the usual Law of Large Numbers, there exists $N_0 \in \mathbb{N}$ such as for any $n \geq N_0$,

$$(1 - \delta)n \mathbf{E}(|\varepsilon_0 + c| - |\varepsilon_0|) \leq \sum_{i=1}^n (|\varepsilon_i + c| - |\varepsilon_i|) \leq (1 + \delta)n \mathbf{E}(|\varepsilon_0 + c| - |\varepsilon_0|) \quad a.s.$$

As a consequence, for $n \geq N_0$,

$$\exp\left(- (1 + \delta)n \mathbf{E}(|\varepsilon_0 + c| - |\varepsilon_0|)\right) \leq \exp\left(- \sum_{i=1}^n (|\varepsilon_i + c| - |\varepsilon_i|)\right) \leq \exp\left(- (1 - \delta)n \mathbf{E}(|\varepsilon_0 + c| - |\varepsilon_0|)\right). \quad (4.1)$$

Moreover, we obtain for $c > 0$:

$$\mathbf{E}(|\varepsilon_0 + c| - |\varepsilon_0|) = \mathbf{E}(c \mathbf{1}_{\varepsilon_0 \geq 0} - c \mathbf{1}_{\varepsilon_0 < -c} - (c + 2\varepsilon_0) \mathbf{1}_{-c \leq \varepsilon_0 < 0}) = c - 2 \int_{-c}^0 F_\varepsilon(t) dt.$$

Therefore, from classical Taylor expansions, we deduce:

$$\mathbf{E}(|\varepsilon_0 + c| - |\varepsilon_0|) \sim f_\varepsilon(0)c^2 \quad \text{when } c \rightarrow 0 \quad (4.2)$$

$$\geq \frac{1}{2} f_\varepsilon(0) c \min(c, c_0) \quad \text{for all } c \geq 0, \quad (4.3)$$

with c_0 depending only on F_ε . For the first expansion, we use an asymptotic expansion of F_ε around 0, and then $\int_{-c}^0 F_\varepsilon(t) dt \sim \int_{-c}^0 (\frac{1}{2} + t f_\varepsilon(0)) dt \sim c - f_\varepsilon(0)c^2$. For the second inequality it is sufficient to study the function $h(c) = c - 2 \int_{-c}^0 F_\varepsilon(t) dt - \frac{1}{2} f_\varepsilon(0) c \min(c, c_0)$. For (u_n) a sequence of positive real numbers such as $u_n \xrightarrow[n \rightarrow \infty]{} 0$, this inequality also implies that for n large enough,

$$\mathbf{E}(|\varepsilon_0 + c| - |\varepsilon_0|) \geq \frac{1}{2} f_\varepsilon(0) c u_n \quad \text{for } c \geq u_n. \quad (4.4)$$

Now assume that $(u_n)_{n \in \mathbb{N}}$ is a sequence such that $u_n \sqrt{n} \xrightarrow[n \rightarrow \infty]{} \infty$ (for instance $u_n = n^{1/4}$), then:

$$\begin{aligned} \int_{-u_n}^{\infty} \exp\left(- n \mathbf{E}(|\varepsilon_0 + c| - |\varepsilon_0|)\right) dc &= \int_{-u_n}^{u_n} \exp\left(- n \mathbf{E}(|\varepsilon_0 + c| - |\varepsilon_0|)\right) dc \\ &\quad + 2 \int_{u_n}^{\infty} \exp\left(- n \mathbf{E}(|\varepsilon_0 + c| - |\varepsilon_0|)\right) dc \\ &= I_1 + I_2. \end{aligned}$$

On the one hand, using (4.2) and with $x = \sqrt{n}c$, and from Lebesgue Theorem,

$$\begin{aligned} I_1 &\underset{n \rightarrow \infty}{\sim} \frac{1}{\sqrt{n}} \int_{-u_n \sqrt{n}}^{u_n \sqrt{n}} \exp(-f_\varepsilon(0)x^2) dx \\ &\underset{n \rightarrow \infty}{\sim} \frac{1}{\sqrt{n}} \int_{-\infty}^{\infty} \exp(-f_\varepsilon(0)x^2) dx \\ &\underset{n \rightarrow \infty}{\sim} \sqrt{\frac{\pi}{f_\varepsilon(0)}} \frac{1}{\sqrt{n}}. \end{aligned}$$

On the other hand, using (4.4),

$$\begin{aligned} |I_2| &\leq 2 \int_{u_n}^{\infty} \exp\left(-\frac{n}{2} f_\varepsilon(0) c u_n\right) dc \\ &\leq \frac{2}{f_\varepsilon(0) n u_n} \exp\left(-\frac{1}{2} n u_n^2\right) = o(I_1), \end{aligned}$$

since $n u_n^2 \xrightarrow[n \rightarrow \infty]{} \infty$. As a consequence,

$$\int_{-\infty}^{\infty} \exp\left(-n \mathbb{E}(|\varepsilon_0 + c| - |\varepsilon_0|)\right) dc \underset{n \rightarrow \infty}{\sim} \sqrt{\frac{\pi}{f_\varepsilon(0)}} \frac{1}{\sqrt{n}}.$$

Finally since this property can be also obtain when we replace n by $(1 + \delta)n$ or $(1 - \delta)n$, we deduce (2.1) from (4.1) \square

4.2 Proof of Lemma 2

Proof. We have:

$$\begin{aligned} &\int_{-\infty}^{\infty} \exp\left(-\sum_{i=1}^n |Y_i - m|\right) dm \\ &= \int_{-\infty}^{\infty} \exp\left(-\sum_{i=1}^n (|Y_i - m| + |\varepsilon_i| - |\varepsilon_i| + |Y_i - \hat{m}_n| - |Y_i - \hat{m}_n|)\right) dm \\ &= \exp\left(-\sum_{i=1}^n |Y_i - \hat{m}_n|\right) \times \exp\left(-\sum_{i=1}^n (|\varepsilon_i| - |\varepsilon_i - (\hat{m}_n - m^*)|)\right) \\ &\quad \times \int_{-\infty}^{\infty} \exp\left(-\sum_{i=1}^n (|\varepsilon_i - (m - m^*)| - |\varepsilon_i|)\right) dm \\ &= \exp\left(-\sum_{i=1}^n |Y_i - \hat{m}_n|\right) \times I \times II \end{aligned} \tag{4.5}$$

On the one hand, using Lemma 1 and a change of variable, we have:

$$II = \int_{-\infty}^{\infty} \exp\left(-\sum_{i=1}^n (|\varepsilon_i - (m - m^*)| - |\varepsilon_i|)\right) dm \underset{n \rightarrow \infty}{\overset{a.s.}{\sim}} \sqrt{\frac{\pi}{f_\varepsilon(0)}} \frac{1}{\sqrt{n}}. \tag{4.6}$$

On the other hand, denote $Z_n = \sqrt{n}(\hat{m}_n - m^*)$. Since (ε_i) belongs to Class WDSN, we know (see for instance [16] in case of strong mixing processes and [20] in case of linear processes) that

$$Z_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{V}{4 f_\varepsilon^2(0)}\right). \tag{4.7}$$

with $f_X(m^*) = f_\varepsilon(0)$ and $V = 1 + 2 \sum_{k=1}^{\infty} \mathbb{E}[\text{sign}(\varepsilon_0) \text{sign}(\varepsilon_k)]$. Then we have $I = \exp(-A_n)$ with

$$A_n := \sum_{i=1}^n (|\varepsilon_i| - |\varepsilon_i - (\hat{m}_n - m^*)|) = \sum_{i=1}^n (|\varepsilon_i| - |\varepsilon_i - n^{-1/2} Z_n|). \quad (4.8)$$

Using the relation: $|x - y| - |x| = -y \text{sign}(x) + 2(y - x)(\mathbf{1}_{0 < x < y} - \mathbf{1}_{y < x < 0})$ for any $(x, y) \in \mathbb{R}^2$, we obtain:

$$\begin{aligned} A_n &= \sum_{i=1}^n (|\varepsilon_i| - |\varepsilon_i - n^{-1/2} Z_n|) = \frac{Z_n}{\sqrt{n}} \sum_{i=1}^n \text{sign}(\varepsilon_i) + 2 \sum_{i=1}^n \left(\varepsilon_i - \frac{Z_n}{\sqrt{n}} \right) (\mathbf{1}_{0 < \varepsilon_i < \frac{Z_n}{\sqrt{n}}} - \mathbf{1}_{\frac{Z_n}{\sqrt{n}} < \varepsilon_i < 0}) \\ &= I_1 + I_2. \end{aligned} \quad (4.9)$$

Let us control the first term $I_1 = \frac{Z_n}{\sqrt{n}} \sum_{i=1}^n \text{sign}(\varepsilon_i)$. From a central limit theorem satisfied by a processes belonging to Class WDSN (strong mixing or weak dependent linear process), we have $\frac{1}{\sqrt{n}} \sum_{i=1}^n \text{sign}(\varepsilon_i) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, V)$, with $0 < V = 1 + 2 \sum_{k=1}^{\infty} \mathbb{E}[\text{sign}(\varepsilon_0) \text{sign}(\varepsilon_k)] < \infty$ (see [16] and [20]). Therefore $\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \text{sign}(\varepsilon_i) \right) = O_P(1)$ and from (4.7), $(Z_n) = O_P(1)$. This induces that

$$I_1 = \left(\frac{Z_n}{\sqrt{n}} \sum_{i=1}^n \text{sign}(\varepsilon_i) \right) = O_P(1). \quad (4.10)$$

Concerning the second right side term of (4.9), first denote for $i \in \mathbb{N}$, the non negative random variable $W_i = \left(\frac{Z_n}{\sqrt{n}} - \varepsilon_i \right) \mathbf{1}_{0 < \varepsilon_i < \frac{Z_n}{\sqrt{n}}}$. Then, for n large enough,

$$\begin{aligned} \mathbb{E}[W_i | Z_n] &= \mathbf{1}_{Z_n \geq 0} \int_0^{\frac{Z_n}{\sqrt{n}}} \left(\frac{Z_n}{\sqrt{n}} - x \right) dF_\varepsilon(x) \\ &\leq \mathbf{1}_{Z_n \geq 0} \frac{Z_n^2}{2n} \sup_{x \in \mathbb{R}} |f_\varepsilon(x)|. \end{aligned}$$

Therefore, with $S_n = \sum_{i=1}^n \left(\frac{Z_n}{\sqrt{n}} - \varepsilon_i \right) \mathbf{1}_{0 < \varepsilon_i < \frac{Z_n}{\sqrt{n}}}$ (implying $S_n \geq 0$) and $M \geq 0$, we have

$$\begin{aligned} \mathbb{E}[S_n | Z_n] &\leq \mathbf{1}_{Z_n \geq 0} \frac{Z_n^2}{2} \sup_{x \in \mathbb{R}} |f_\varepsilon(x)| \\ \mathbb{E}[S_n | |Z_n| \leq M] &\leq \frac{M^2}{2} \sup_{x \in \mathbb{R}} |f_\varepsilon(x)|. \end{aligned}$$

Then, for any $M' > 0$, using a conditional Markov inequality, for any $n \in \mathbb{N}^*$,

$$\begin{aligned} \mathbb{P}[S_n \leq M'] &\geq \mathbb{P}[S_n \leq M' \cap |Z_n| \leq M] \\ &\geq \left(1 - \mathbb{P}[S_n > M' | |Z_n| \leq M] \right) \times \mathbb{P}[|Z_n| \leq M] \\ &\geq \left(1 - \frac{M^2}{2M'} \sup_{x \in \mathbb{R}} |f_\varepsilon(x)| \right) \times \mathbb{P}[|Z_n| \leq M], \end{aligned}$$

from (4.11). Now let $\varepsilon > 0$. Then, since $(Z_n) = O_P(1)$, there exists $M > 0$ such that $\sup_{n \in \mathbb{N}^*} \mathbb{P}[|Z_n| \leq M] \geq 1 - \frac{\varepsilon}{2}$. Now chose $M' > 0$ such as $\frac{M^2}{2M'} \sup_{x \in \mathbb{R}} |f_\varepsilon(x)| \leq \frac{\varepsilon}{2}$. We then deduce that for any $\varepsilon > 0$ then there exists $M' > 0$ such that

$$\sup_{n \in \mathbb{N}^*} \mathbb{P}[S_n \leq M'] \geq \left(1 - \frac{\varepsilon}{2} \right) \geq 1 - \varepsilon \quad \implies \quad (S_n) = O_P(1). \quad (4.11)$$

By symmetry, and this achieves the proof, we finally obtain that

$$I_2 = O_P(1) \quad \implies \quad (A_n) = O_P(1),$$

from (4.9) and (4.10). As a consequence, using the decomposition (4.5) and (4.6), this achieves the proof. \square

4.3 Proof of Proposition 1

Proof. Mutatis mutandis, We will establish a proof following the same procedure as the proofs of Lemmas 1 and 2. Using the same decomposition than in Lemma 2, we have

$$\begin{aligned} & \int_{\mathbb{R}^{p+1}} \exp\left(-\sum_{i=1}^n |Y_i - \theta_0 - \sum_{j=1}^p \theta_j X_i^{(j)}|\right) d\theta \\ &= \exp\left(-\sum_{i=1}^n |Y_i - \hat{\theta}_0 - \sum_{j=1}^p \hat{\theta}_j X_i^{(j)}|\right) \exp\left(-\sum_{i=1}^n \left(|\varepsilon_i| - |\varepsilon_i - (\hat{\theta}_0 - \theta_0^*) - \sum_{j=1}^p (\hat{\theta}_j - \theta_j^*) X_i^{(j)}|\right)\right) \\ & \quad \times \int_{\mathbb{R}^{p+1}} \exp\left(-\sum_{i=1}^n \left(|\varepsilon_i - (\theta_0 - \theta_0^*) - \sum_{j=1}^p (\theta_j - \theta_j^*) X_i^{(j)}| - |\varepsilon_i|\right)\right) d\theta. \end{aligned} \quad (4.12)$$

First, after denoting $\|\cdot\|_\infty$ the norm in \mathbb{R}^{p+1} such as $\|z\|_\infty = \max(z_1, \dots, z_{p+1})$ for $z = (z_1, \dots, z_{p+1}) \in \mathbb{R}^{p+1}$, we obtain a multivariate version of Lemma 1:

$$\begin{aligned} \mathbb{E}\left(|\varepsilon_0 - \theta_0 - \sum_{j=1}^p \theta_j X_0^{(j)}| - |\varepsilon_0|\right) & \underset{\|\theta\|_\infty \rightarrow 0}{\sim} f_\varepsilon(0) \mathbb{E}\left((\theta_0 + \sum_{j=1}^p \theta_j X_0^{(j)})^2\right) \\ & \underset{\|\theta\|_\infty \rightarrow 0}{\sim} f_\varepsilon(0) \theta' \Gamma \theta, \end{aligned} \quad (4.13)$$

where $\theta = (\theta_0, \dots, \theta_p)'$ using the assumptions on sequences (X_i) . As a consequence, using the independence of the different random variable sequences and the almost same cutting of the integral than in Lemma 1, we have:

$$\begin{aligned} \int_{\mathbb{R}^{p+1}} \exp\left(-n \mathbb{E}\left(|\varepsilon_0 - \theta_0 - \sum_{j=1}^p \theta_j X_0^{(j)}| - |\varepsilon_0|\right)\right) d\theta &= \int_{\|\theta\|_\infty \leq u_n} \dots d\theta + \int_{\|\theta\|_\infty > u_n} \dots d\theta \\ & \underset{n \rightarrow \infty}{\sim} \int_{\|\theta\|_\infty \leq u_n} \exp\left(-n f_\varepsilon(0) \theta' \Gamma \theta\right) d\theta \\ & \underset{n \rightarrow \infty}{\sim} \int_{\mathbb{R}^{p+1}} \exp\left(-n f_\varepsilon(0) \theta' \Gamma \theta\right) d\theta \\ & \underset{n \rightarrow \infty}{\sim} \left(\left(\frac{\pi}{f_\varepsilon(0)}\right)^{\frac{p+1}{2}} \frac{1}{\det(\Gamma)^{1/2}}\right) \frac{1}{n^{\frac{p+1}{2}}}. \end{aligned} \quad (4.14)$$

Now, since (ε_i) belongs to Class WDSN, we know that (ε_i) satisfies a strong law of large numbers and following Lemma 2, we deduce:

$$\int_{\mathbb{R}^{p+1}} \exp\left(-\sum_{i=1}^n |\varepsilon_i - \theta_0 - \sum_{j=1}^p \theta_j X_i^{(j)}| - |\varepsilon_i|\right) d\theta \underset{n \rightarrow \infty}{\overset{a.s.}{\sim}} \left(\left(\frac{\pi}{f_\varepsilon(0)}\right)^{\frac{p+1}{2}} \frac{1}{\det(\Gamma)^{1/2}}\right) \frac{1}{n^{\frac{p+1}{2}}}. \quad (4.15)$$

Now consider the second right side term

$$J_2 = \exp(-B_n) \quad \text{with} \quad B_n := \sum_{i=1}^n \left(|\varepsilon_i| - |\varepsilon_i - (\hat{\theta}_0 - \theta_0^*) - \sum_{j=1}^p (\hat{\theta}_j - \theta_j^*) X_i^{(j)}|\right). \quad (4.16)$$

Since (ε_i) belongs to Class WDSN, we know (see for instance [11] in case of white noise, [16] in case of strong mixing processes and [20] in case of linear processes), and since the assumptions of Proposition 1 imply $\frac{1}{n} X X' \xrightarrow[n \rightarrow \infty]{a.s.} \Gamma$ with $X = (X_i)_{1 \leq i \leq n}$, we know that

$$Z_n = \sqrt{n} (\hat{\theta} - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{4 f_\varepsilon^2(0)} \Gamma^{-1} W \Gamma^{-1}\right), \quad (4.17)$$

with $W = \Gamma + 2 \mathbb{E}[X_0] (\mathbb{E}[X_0])' \sum_{k=1}^{\infty} \mathbb{E}[\text{sign}(\varepsilon_0) \text{sign}(\varepsilon_k)]$.

As in the proof of Lemma 2, we use the decomposition:

$$\begin{aligned} B_n &= \sum_{i=1}^n (|\varepsilon_i| - |\varepsilon_i - n^{-1/2} C_i' Z_n|) \\ &= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \text{sign}(\varepsilon_i) C_i' \right) Z_n + 2 \sum_{i=1}^n \left(\varepsilon_i - \frac{C_i' Z_n}{\sqrt{n}} \right) \left(\mathbf{1}_{0 < \varepsilon_i < \frac{C_i' Z_n}{\sqrt{n}}} - \mathbf{1}_{\frac{C_i' Z_n}{\sqrt{n}} < \varepsilon_i < 0} \right) \\ &= I + II \quad (4.18) \end{aligned}$$

Moreover, as it has been established in [16] and [20], the sequence $(\text{sign}(\varepsilon_i) C_i)_{n \in \mathbb{N}}$ satisfies a multidimensional central limit theorem, and therefore:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \text{sign}(\varepsilon_i) C_i \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}_{p+1}(0, W) \implies I = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \text{sign}(\varepsilon_i) C_i' \right) Z_n = O_{\mathbb{P}}(1). \quad (4.19)$$

Now, using the same expansions than in Lemma 2, with $T_n = \sum_{i=1}^n \left(\frac{C_i' Z_n}{\sqrt{n}} - \varepsilon_i \right) \mathbf{1}_{0 < \varepsilon_i < \frac{C_i' Z_n}{\sqrt{n}}}$ we have:

$$\begin{aligned} \mathbb{E}\left[T_n \mid Z_n, (C_i)_{1 \leq i \leq n}\right] &\leq \frac{\sup_{x \in \mathbb{R}} |f_\varepsilon(x)|}{2n} \sum_{i=1}^n \mathbf{1}_{C_i' Z_n \geq 0} (C_i' Z_n)^2 \\ &\leq \frac{\sup_{x \in \mathbb{R}} |f_\varepsilon(x)|}{2} \left(\frac{1}{n} \sum_{i=1}^n \|C_i\|^2 \right) \|Z_n\|^2, \end{aligned}$$

from Cauchy-Schwarz Inequality. Denote $W_n := \left(\frac{1}{n} \sum_{i=1}^n \|C_i\|^2 \right) \|Z_n\|^2$. It is clear from the usual strong law of large numbers that $\frac{1}{n} \sum_{i=1}^n \|C_i\|^2 \xrightarrow[n \rightarrow \infty]{a.s.} \text{Trace}(\Gamma)$ and therefore, with (4.17), $(W_n) = O_{\mathbb{P}}(1)$. Then, for any $M' > 0$, using a conditional Markov inequality, for any $n \in \mathbb{N}^*$,

$$\begin{aligned} \mathbb{P}\left[T_n \leq M'\right] &\geq \mathbb{P}\left[T_n \leq M' \cap Z_n \leq M\right] \\ &\geq \left(1 - \frac{M}{2M'} \sup_{x \in \mathbb{R}} |f_\varepsilon(x)|\right) \times \mathbb{P}\left[W_n \leq M\right], \end{aligned}$$

and we conclude as in Lemma 2 that $(T_n) = O_{\mathbb{P}}(1)$ and therefore $(II) = O_{\mathbb{P}}(1)$ with II defined in (4.16). Using also (4.19), we obtain $(B_n) = O_{\mathbb{P}}(1)$ with B_n defined in (4.16). Therefore, the Proposition 1 is proved from the decomposition (4.12) and (4.15). \square

4.4 Proof of Proposition 2

Proof. We consider

$$I_n = \int_0^\infty \int_{\mathbb{R}^{p+1}} \exp(\ell_n(\theta, \lambda)) d\theta d\lambda \quad \text{with} \quad \ell_n(\theta, \lambda) = n \log(\lambda/2) - \lambda \sum_{i=1}^n \left| Y_i - \theta_0 - \sum_{j=1}^p \theta_j X_i^{(j)} \right|.$$

Since $\lambda \rightarrow \int_{R^{p+1}} \exp(L_n(\theta, \lambda)) d\theta$ is a measurable positive function, using the expansion of Proposition 1, we can write:

$$I_n \underset{n \rightarrow \infty}{\overset{\text{P}}{\sim}} \left(\frac{2\pi}{n}\right)^{\frac{p+1}{2}} \frac{2^{-n}}{\det(\Gamma)^{1/2}} \int_0^\infty \lambda^n \exp\left(-\lambda \sum_{i=1}^n \left|Y_i - \hat{\theta}_0 - \sum_{j=1}^p \hat{\theta}_j X_i^{(j)}\right|\right) e^{-\lambda B_n} d\lambda$$

Therefore, after a change of variable, we deduce that:

$$I_n \underset{n \rightarrow \infty}{\overset{\text{P}}{\sim}} \left(\frac{2\pi}{n}\right)^{\frac{p+1}{2}} \frac{2^{-n}}{\det(\Gamma)^{1/2}} \frac{n!}{\left(B_n + \sum_{i=1}^n \left|Y_i - \hat{\theta}_0 - \sum_{j=1}^p \hat{\theta}_j X_i^{(j)}\right|\right)^{n+1}}$$

Now using the usual maximum likelihood estimator $\hat{\lambda} = n \left(\sum_{i=1}^n \left|Y_i - \hat{\theta}_0 - \sum_{j=1}^p \hat{\theta}_j X_i^{(j)}\right|\right)^{-1}$, and with the Stirling expansion $n! \underset{n \rightarrow \infty}{\sim} n^n e^{-n} \sqrt{2\pi n}$, we obtain:

$$I_n \underset{n \rightarrow \infty}{\overset{\text{P}}{\sim}} \left(\frac{2\pi}{n}\right)^{\frac{p+2}{2}} \frac{2^{-n}}{\det(\Gamma)^{1/2}} \frac{e^{-n}}{\left(\frac{B_n}{n} + \frac{1}{\lambda}\right)^{n+1}}$$

$$\underset{n \rightarrow \infty}{\overset{\text{P}}{\sim}} \left(\frac{2\pi}{n}\right)^{\frac{p+2}{2}} \frac{1}{\det(\Gamma)^{1/2}} \hat{\lambda} e^{-\hat{\lambda} B_n} \left(\frac{\hat{\lambda}}{2}\right)^n e^{-n}$$

and therefore $-2 \log(I_n) \underset{n \rightarrow \infty}{\overset{\text{P}}{\sim}} -2 \ell_n(\hat{\theta}, \hat{\lambda}) + (p+2) \log(n)$

$$-(p+2) \log(2\pi) - 2 \log(\hat{\lambda} e^{-\hat{\lambda} B_n}) + \log(\det(\Gamma))$$

□

References

- [1] Bardet, J.-M., Boularouk, Y. and Djballah, K. (2017) Asymptotic behaviour of the Laplacian quasi-maximum likelihood estimator of affine causal processes. *Electronic Journal of Statistics*, **11**, 452–479.
- [2] Bardet, J.-M. and Dion, C. (2019). Robust semi-parametric multiple change-points detection. *Signal Processing*, **156**, 145–155.
- [3] Bardet, J.-M., Kamila, K. and Kengne, W. (2020). Consistent model selection criteria and goodness-of-fit test for common time series models. *Electronic Journal of Statistics*, **14**, 2009–2052.
- [4] Bai, J. (1995). Least absolute deviation estimation of a shift, *Econometric Theory*, **11**, 403–436.
- [5] Bai, J. (1998). Estimation of multiple-regime regressions with least absolute deviation, *Journal of Statistical Planning and Inference*, **74**, 103–134.
- [6] Chen, C.-F. (1985). On asymptotic normality of limiting density functions with Bayesian implications, *J. R. Statist. Soc. B*, **47**, 540–546.

- [7] R. Davis and W. Dunsmuir. Least absolute deviation estimation for regression with ARMA errors. *Journal of Theoretical Probability*, **10**, 481–497, 1997.
- [8] Huber, P.J. (1964). Robust estimation of a location parameter, *The Annals of Mathematical statistics*, **35**, 73–101.
- [9] Huber, P.J. (1981). *Robust statistics*, Wiley Series in Probability and Mathematical Statistics, Wiley & Sons Inc., New York.
- [10] Kass, R. E., Tierney, L. and Kadane, J. B. (1990). The Validity of Posterior Expansions Based on Laplace’s Method. *Essays in Honor of George Barnard*, eds. S. Geisser, J. S. Hodges, S. J. Press, and A. Zellner, Amsterdam: North-Holland, 473–188.
- [11] Koenker, R. and Bassett, G. (1978). Regression Quantiles, *Econometrica*, **46**, 33–50.
- [12] Koul, H.L. and Mukherjee, K. (1993). Asymptotics of R-, MD- and LAD-estimators in linear regression models with long range dependent errors *Probab. Theory Relat. Fields*, **95**, 535–553.
- [13] Lebarbier, E. and Mary-Huard, T. (2000). Une introduction au critère bic: fondements théoriques et interprétation. *Journal de la Société française de statistique*, **147**, 39–57.
- [14] McQuarrie, A. D. and Tsai, C. L. (1998). Regression and time series model selection. World Scientific.
- [15] Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *Annals of Statistics*, **12**, 758–765.
- [16] Phillips, P. (1991). A Shortcut to LAD Estimator Asymptotics. *Econometric Theory*, **7**, 450–463.
- [17] Pollard, D. (1991). Asymptotics for Least Absolute Deviation Regression Estimators. *Econometric Theory*, **7**, 186–199.
- [18] Raftery, A. (1995) Bayesian model selection in social research, *Sociological Methodology*, **25**, 111–163.
- [19] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464.
- [20] Wu, W.B. (2007). M-Estimation of Linear Models with Dependent Errors. *The Annals of Statistics*, **35**, 495–521.