UNIVERSIDAD NACIONAL DE CÓRDOBA

FACULTAD DE MATEMÁTICA, ASTRONOMÍA, FÍSICA Y COMPUTACIÓN

SERIE “A”

TRABAJOS DE MATEMÁTICA

N.º 121/2020

The BBM-estimator in AR-2D models: a complete study

Grisel M. Britos-Silvia M. Ojeda-Laura A. Rodríguez Astrain-Oscar H. Bustos

Editor: Jorge G. Adrover

CIUDAD UNIVERSITARIA – 5000 CÓRDOBA

REPÚBLICA ARGENTINA

The BMM-estimator in AR-2D models: a complete study

Grisel M. Britos*[,a], Silvia M. Ojeda[a], Laura A. Rodríguez Astrain[a], Oscar H. Bustos[a]

[a]*Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba, Córdoba, Argentina*

Abstract

In this work, we present the BMM 2D estimator, a robust estimator for the parameters of the bidimensional autoregressive model (AR-2D model). The new estimator is a two-dimensional extension of the BMM estimator for the parameters of the autoregressive models used in time series analysis. We demonstrate that the BMM 2D estimator is consistent and asymptotically normal, which provides a valuable tool to carry out inferential studies about the parameters of the AR-2D model. We compare its performance with existing estimators through a Monte Carlo study, considering different levels of additive contamination and window sizes. The results show that the new estimator competes successfully with the other methods, both in accuracy and precision. In the context of image restoration problems, we illustrate the performance of the BMM 2D estimator compared with the least squares estimator.

Key words: AR-2D Models, Robust Estimators, Consistency, Asymptotic Normality, Image Analysis

1. Introduction

In the area of image processing and computer vision, the use and development of techniques called "robust" is frequent, not referring to the term robustness imperiously from a formal statistical perspective. Within this framework, the classic tools do not consider the structure or topology inherent to the images and, in most situations, the models require strong hypotheses about the laws that govern the observed process (Alata & Olivier (2003), Bustos et al. (2009), Dormann et al. (2007), Latha et al. (2014), Ojeda et al. (2010), Quintana et al. (2011), Sahu et al. (2015), Sain & Cressie (2007), Vallejos & Mardesic (2004) and Zielinski et al. (2010)). Two-dimensional autoregressive models were introduced by Whittle (1954) as a class of models capable of capturing spatial correlation in the data collected. These models have proven to be of great importance in several areas that benefit from image processing (Smith et al. (1986), Dormann et al. (2007), Sain & Cressie (2007)) since they make it possible to represent the intensity of an image through a small number of parameters and naturally extend the definition of autoregressive models for time series to $\mathbb{Z}^2$. Consequently, several of the robust tools developed to estimate the parameters of the one-dimensional autoregressive model have been implemented for AR-2D models under contaminated spatial data (see Kashyap & Eom (1988), Allende et al. (1998), Ojeda et al. (2002)).

In the context of robust estimators of the parameters of the AR-2D model with a finite and arbitrary number of parameters, at least three estimators have been defined and studied: the M, GM and RA estimators. In 1988, Kashyap & Eom (1988) presented the M estimators for the AR-2D models. Then, for the same models, Allende et al. (1998) implemented an extension of the M estimators: the Generalized M estimators (GM). They extend the GM estimators defined for unidimensional AR models used to model process in time series. While the performance of these two estimators is acceptable for contaminated data with innovative contamination, there are not known rigorous studies on its asymptotic properties. Similarly, robust Residual Autocovariance (RA) estimators were introduced by Ojeda et al. (2002) for two-dimensional autoregressive models extending the definition for time series of the estimator with the same name (Bustos & Yohai (1986)).

*Corresponding author

Email address: gbritos@famaf.unc.edu.ar (Grisel M. Britos)

The performance of this estimator is better than the M estimator and slightly higher than the GM estimator under innovative and additive contamination. In addition, the RA estimator outperforms its competitors M and GM because its asymptotic properties are known. Indeed, this estimator is strongly consistent and asymptotically normal with known variance-covariance matrix. In contrast, the main disadvantage of the RA estimator compared to the M and GM estimators is its high computational cost, which makes it an ineligible tool in practical applications. Finally, it should be noted that in Britos & Ojeda (2018) an estimator, called BMM 2D, was defined to estimate the parameters of the two-dimensional autoregressive model with three parameters, partially extending the definition given in Muler et al. (2009) for time series. This estimator proved to be a successful tool for estimating the three parameters of the model when the spatial data are contaminated by different contamination schemes, showing good performance (in precision, as well as accuracy and computational time) compared to the estimators mentioned above.

In this paper we present the BMM 2D estimator for the parameters of the unilateral autoregressive spatial processes with p parameters, generalizing the definition established in Britos & Ojeda (2018). This estimator preliminarily estimates the M-scale of the innovation process and then makes an M-estimation of the parameters, relying on an auxiliary model called BIP-AR 2D which allows to control the effect of the outliers in the innovative process. Later, we study the asymptotic behavior of the BMM 2D estimator and we give precise conditions for the strongly consistency and asymptotic normality of the estimator. The paper is organized as follows. Section 2 presents the motivation of this study based on: the adequacy of AR-2D models in the representation of real images; the impact of contamination on the performance of classical least square estimators, and empirical evidence that the BMM version of such estimators is able to resist additive contamination. In section 3, the AR-2D model is formally defined with p parameters and the definition of the auxiliary model BIP-AR 2D is presented. Section 4 defines the BMM estimator of a 2D autoregressive model with p parameters and Section 5 analyzes its performance in an AR-2D model with two parameters under additive contamination compared to other two-dimensional estimators (LS, M, GM, RA). In section 6, the theorems that give strong consistency and asymptotic normality to the BMM estimator are established. Section 7 discusses some final remarks and directions for future work. Finally, some necessary lemmas are enunciated and demonstrated in the Appendix to prove the theorems presented in Section 6 and these theorems are proven.

2. Motivation

Unilateral two-dimensional (AR-2D) autoregressive models have shown a successful performance in the local approximation of digital images. This is due to the great expressiveness these models have to represent a great diversity of textures present in the images. In this section we show graphic examples about the ability of these models to represent texture images of real scenarios. We use the algorithm presented in (Britos & Ojeda (2018)), defined to approximate images through a unilateral AR-2D model with three parameters. Figure 1 shows the results obtained with the said procedure. The images of the Figs. 1 (a) and 1 (g) were obtained from the USC-SIPI image database http://sipi.usc.edu/database/ and contaminated with additive noise, adding a constant value (equal to 50 or 70) in 10 % of the pixels of the images, obtaining the images of the Figures 1 (b) and 1 (h), respectively. From the contaminated images, an AR-2D model was adjusted using a window the size of the entire image. Initially, we restore the original images estimating the parameters from the LS estimator, obtaining the images of the Figures 1 (c) and 1 (i). Later, in a second instance we use the robust BMM-2D estimator. The Figures 1 (e) and 1 (k) show the results of the images restored for this case. In general, the contamination prevents a reliable statistical analysis since the results can be affected by the presence of outliers. However, the similarity between the estimated and the original images suggests that the AR-2D model is suitable for representing real images even in the presence of contamination. Figs. 1 (d) and (j) present the pixel-to-pixel differences between true images and those restored from the LS estimator of the parameters. Similarly, Figures 1 (f) and 1 (l) show the differences using the robust BMM estimator. The better the estimate, the lower the information present in these difference images. When these differences come from the LS estimator, the presence of a residual structure that was not explained by the restoration model is observed; the residual structure is much less noticeable in the differences from the BMM restoration, which contain less information. This fact suggests that analyzing and studying the properties of the robust BMM estimator in detail are relevant tasks.

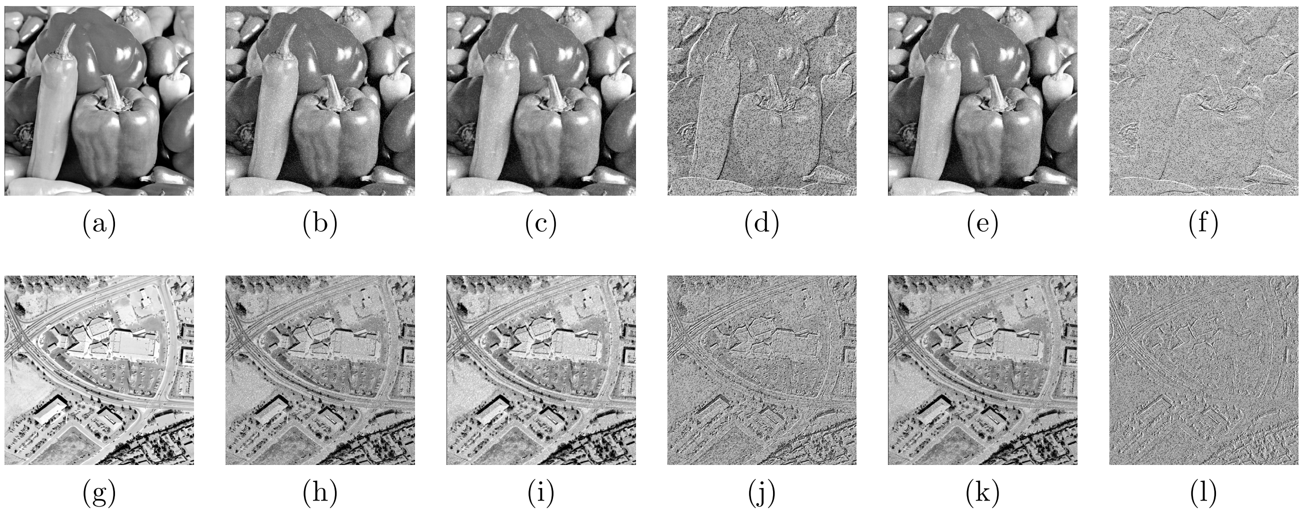


Figure 1: Original and contaminated images, their LS and BMM estimations and their residuals. (a) Original X. (b) Contaminated. (c) LS estimation $\hat{X}_1$. (d) $X - \hat{X}_1$. (e) BMM estimation $\hat{X}_2$. (f) $X - \hat{X}_2$. (g) Original X. (h) Contaminated. (i) LS estimation $\hat{X}_1$. (j) $X - \hat{X}_1$. (k) BMM estimation $\hat{X}_2$. (l) $X - \hat{X}_2$.

3. The model under study

3.1. The Central Two-dimensional Autoregressive Model

We consider a stationary and strongly causal bidimensional AR process Y, with mean μ_0, that can be represented by

$$\Phi_0(B_1, B_2)(Y_{i,j} - \mu_0) = \varepsilon_{i,j},$$

with innovation process $\varepsilon = \{\varepsilon_{i,j}\}$ where $\varepsilon_{i,j}$'s are i.i.d. random variables with symmetric and strictly unimodal distribution and $\mathbf{\Phi}_0(B_1, B_2)$ is a polynomial operator given by $\mathbf{\Phi}_0(B_1, B_2) = 1 - \sum_{(k,l)\in T} \phi^0_{k,l} B_1^k B_2^l$, where $T = T_L = \{(i,j) \in \mathbb{Z}^2 : 0 \leq i,j \leq L, (i,j) \neq (0,0)\}$, with fixed $L \in \mathbb{N}$ and $\phi^0_{k,l} \in \mathbb{R}$, $\forall (k,l) \in T$. $\mathbf{\Phi}_0(B_1, B_2)$ is called "unilateral polynomial with support in T"(Britos (2019)).

We can consider that the Y process is observed in W_M, a strongly causal window: $W_M = \{(i,j) \in \mathbb{Z}^2 : 0 \leq i,j \leq M\}$, with $M \in \mathbb{N}$ and $L \ll M$. We define $(W_M \sim T)_L := \{(i,j) \in W_M : (i-L, j-L) \in W_M\}$. To simplify the notation, we write $(W_M \sim T)$ instead of $(W_M \sim T)_L$. Therefore, $\forall (i,j) \in (W_M \sim T)$, $Y_{i,j}$ can be expressed as

$$Y_{i,j} = \mu_0 + \sum_{(k,l)\in T} \phi^0_{k,l}(Y_{i-k,j-l} - \mu_0) + \varepsilon_{i,j}. \tag{3.1}$$

A sufficient condition for the Y process to be strongly causal is that $\sum_{(k,l)\in T} |\phi_{k,l}| < 1$ (see Guyon (1995)). In that case, Y supports an infinite moving average representation as following:

$$\dot{Y}_{i,j} = \sum_{(k,l)\in I} \lambda_{k,l} \varepsilon_{i-k,j-l}, \tag{3.2}$$

where $I = \{(k,l) \in \mathbb{Z}^2 : k,l \geq 0\}$, $\lambda_{k,l} \in \mathbb{R}$, $\forall (k,l) \in I$ and $\lambda_{0,0} = 1$.

We assume that the parameters of the model (3.1) are unknown. These are the coefficients $\{\phi^0_{k,l}\}_{(k,l)\in T}$ of the unilateral polynomial (indexed in T, subset of $\mathbb{Z}^2$) and the mean μ_0 of the process. Therefore, to obtain a vector representation of the parameters, we define a complete order on $\mathbb{Z}^2$, called spiral order and denoted by $\preceq$. Figure 2 gives a clear idea of this order. For more details it can be consulted Britos (2019).

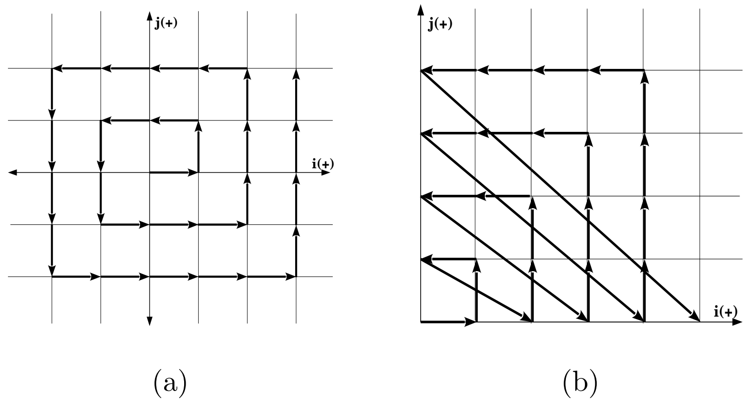


Figure 2: (a) Spiral order in $\mathbb{Z}^2$ ($\preceq$), (b) Spiral order restricted to I

According to the spiral order, we can obtain the following vector expression of the parameters:

$$\beta_0 = (\boldsymbol{\phi}_0, \mu_0) = (\phi^0_{1,0}, \phi^0_{1,1}, \phi^0_{0,1}, \phi^0_{2,0}, \phi^0_{2,1}, \phi^0_{2,2}, \phi^0_{1,2}, \phi^0_{0,2}, ..., \phi^0_{L,0}, ..., \phi^0_{L,L}, \phi^0_{L-1,L}, ..., \phi^0_{0,L}, \mu_0),$$

where

$$\boldsymbol{\phi}^0 = (\phi^0_{1,0}, \phi^0_{1,1}, \phi^0_{0,1}, \phi^0_{2,0}, \phi^0_{2,1}, \phi^0_{2,2}, \phi^0_{1,2}, \phi^0_{0,2}, ..., \phi^0_{L,0}, ..., \phi^0_{L,L}, \phi^0_{L-1,L}, ..., \phi^0_{0,L}).$$

In this way, the parametric space is expressed as

$$\mathcal{B} = \{\beta = (\boldsymbol{\phi}, \mu) : \boldsymbol{\phi} \in \mathcal{B}_0, \mu \in \mathbb{R}\},$$

where $\mathcal{B}_0 = \{\boldsymbol{\phi} \in \mathbb{R}^{(L+1)^2-1} : \sum_{(k,l)\in T} |\phi_{k,l}| \leq 1 - \epsilon\}$ with a fixed $\epsilon > 0$.

For all $\beta \in \mathcal{B}$, the residual function $\varepsilon_{i,j}(\beta)$ is defined as

$$\varepsilon_{i,j}(\beta) = \dot{Y}_{i,j} - \boldsymbol{\phi}' \tilde{\mathbf{Y}}_{i,j}, \tag{3.3}$$

for all $(i,j) \in (W_M \sim T)$ and $\varepsilon_{i,j}(\beta) = 0$ in any other case, where

$$\begin{aligned} \tilde{\boldsymbol{Y}}_{i,j} &= (B^{(1,0)}\dot{Y}_{i,j}, B^{(1,1)}\dot{Y}_{i,j}, B^{(0,1)}\dot{Y}_{i,j}, ..., B^{(L,0)}\dot{Y}_{i,j}, ..., B^{(L,L)}\dot{Y}_{i,j}, ..., B^{(0,L)}\dot{Y}_{i,j}) \\ &= (\dot{Y}_{i-1,j}, \dot{Y}_{i-1,j-1}, \dot{Y}_{i,j-1}, ..., \dot{Y}_{i-L,j}, ..., \dot{Y}_{i-L,j-L}, ..., \dot{Y}_{i,j-L}). \end{aligned}$$

Note that if $(i,j) \in (W_M \sim T)$, (3.3) is equivalent to:

$$\varepsilon_{i,j}(\beta) = (Y_{i,j} - \mu) - \sum_{(k,l)\in T} \phi_{k,l}(Y_{i-k,j-l} - \mu). \tag{3.4}$$

In addition, $\varepsilon_{i,j} = \varepsilon_{i,j}(\beta_0)$, $\forall (i,j) \in (W_M \sim T)$.

From here on we will refer to the model presented in this subsection, as the pure or central two-dimensional autoregressive model and we will denote it AR-2D.

3.2. A Class of Bounded Nonlinear AR Bidimensional Models (BIP-AR 2D)

The robustness is related to the possibility of accurately estimating the parameters of a model that satisfies precise conditions (assumptions of the model), when in fact we observe a contaminated process, where the assumptions are not strictly satisfied. We can model an AR-2D process contaminated with an α fraction of outliers by:

$$Z_{i,j} = (1 - \xi^{\alpha}_{i,j})Y_{i,j} + \xi^{\alpha}_{i,j}W_{i,j},$$

where Y is a pure AR-2D process, W is a replacement process and ξ^{α} is a process of ones and zeros such that $P(\xi^{\alpha}_{i,j} = 1) = \alpha$ and $P(\xi^{\alpha}_{i,j} = 0) = 1 - \alpha$. To estimate the Y parameters in a robust way, when the Z process has been observed instead, we define a new family of models that help control the effect of atypical data on parameter estimation. It is a new class of bounded models, not linear, derived from the AR-2D models, which we will call the AR-2D models of bounded innovative propagation (BIP-AR 2D). This class is defined from the two-dimensional generalization of the analog model presented for time series by Muler et al. (2009). The BIP-AR 2D model arises from the need to estimate the parameters of the pure AR-2D model in the best possible way when a contaminated process is observed, controlling through of a bounded function, the outliers that can be propagated in innovations.

Consider Y as in (3.1) with innovation process ε. Let $\rho : \mathbb{R} \to \mathbb{R}$ be a continuous and not constant function, such that $\rho(0) = 0$, $\rho(x) = \rho(-x)$ and $\rho(x)$ is not decreasing in $|x|$. Let be a constant $b \in \mathbb{R}$ such that $E(\rho(Z)) = b$ when Z is a variable with strictly unimodal symmetric density. Then, the M-scale σ of ε is defined as the solution of the following equation:

$$E\left(\rho\left(\frac{\varepsilon_{i,j}}{\sigma}\right)\right) = b, \tag{3.5}$$

and the family of auxiliary models, called the BIP-AR 2D family, is given by:

$$\dot{X}_{i,j} = \sum_{(k,l)\in I\setminus\{(0,0)\}} \lambda_{k,l}\sigma\eta\left(\frac{\varepsilon_{i-k,j-l}}{\sigma}\right) + \varepsilon_{i,j}, \tag{3.6}$$

where the coefficients $\lambda_{k,l}$ are defined as in (3.2); $\varepsilon_{i,j}$'s are the variables of the ε process, η is an odd and bounded function, and σ is the M-scale of ε. Due to the properties of the constant b, when $\varepsilon \sim N(0, \sigma^2_{\varepsilon})$, we have that $\sigma = \sigma_{\varepsilon}$. In this case, $\varepsilon_{i,j}/\sigma_{\varepsilon} \sim N(0, 1)$ and $E(\rho(\varepsilon_{i,j} \setminus \sigma_{\varepsilon})) = b$; then, σ_{ε} is the M-scale of ε corresponding to b and ρ.

We choose a bounded function η such that it satisfies $\eta(x) = x$ for $|x| \leq k$ for some $k > 0$. So, in the model (3.6), η does not affect the values of the $\varepsilon_{i,j}/\sigma$ when they are in the range $[-k, k]$; but it controls or limits them if they exceed in absolute value to k. In any case, since $\lambda_{i,j} \to 0$ when $(i, j) \to \infty$ (the limit is calculated according to the order restricted to I), it is expected that the effect of atypical observations in $\varepsilon_{i,j}$ disappears in future observations.

Note that (3.6) can be written as:

$$\dot{X}_{i,j} = \sigma\Phi_0^{-1}(B_1, B_2)\eta\left(\frac{\varepsilon_{i,j}}{\sigma}\right) - \sigma\eta\left(\frac{\varepsilon_{i,j}}{\sigma}\right) + \varepsilon_{i,j},$$

and multiplying both members by $\Phi_0(B_1, B_2)$, we get:

$$\Phi_0(B_1, B_2)\dot{X}_{i,j} = \sigma\eta\left(\frac{\varepsilon_{i,j}}{\sigma}\right) - \sigma\Phi_0(B_1, B_2)\eta\left(\frac{\varepsilon_{i,j}}{\sigma}\right) + \Phi_0(B_1, B_2)\varepsilon_{i,j}.$$

The previous expression is equivalent to:

$$X_{i,j} = \mu_0 + \sum_{(k,l)\in T} \phi^0_{(k,l)}(X_{i-k,j-l} - \mu_0) + \sigma \sum_{(k,l)\in T} \phi^0_{(k,l)} \eta \left(\frac{\varepsilon_{i-k,j-l}}{\sigma} \right)$$
$$+ \varepsilon_{i,j} - \sum_{(k,l)\in T} \phi^0_{(k,l)} \varepsilon_{i-k,j-l}.$$

In the strongly causal window W_M, for each $(i,j) \in (W_M \sim T)$, using the equality above, the residual functions $\varepsilon^b_{i,j}(\beta,\sigma)$'s of the BIP-AR 2D model can be recursively defined:

$$\varepsilon^b_{i,j}(\beta,\sigma) = X_{i,j} - \mu - \sum_{(k,l)\in T} \phi_{(k,l)}(X_{i-k,j-l} - \mu) - \sum_{(k,l)\in T} \phi_{(k,l)} \sigma \eta \left(\frac{\varepsilon^b_{i-k,j-l}(\beta,\sigma)}{\sigma} \right) + \sum_{(k,l)\in T} \phi_{(k,l)} \varepsilon^b_{i-k,j-l}(\beta,\sigma), \tag{3.7}$$

and $\varepsilon^b_{i,j}(\beta,\sigma) = 0$ in any other case.

4. BMM 2D Estimator

This section presents a new estimator of the parameters in the AR-2D models (central model) as another alternative to existing robust methods, which is competitive with respect to the other methods and shows desirable asymptotic properties. The new proposal, called BMM estimation calculates, in a first stage M scale estimates obtained from the residual functions of the AR-2D and BIP-AR 2D models, leaving us with the lowest estimated scale. Then, in a second stage, M estimates of the autoregression parameters of the pure AR-2D model are obtained from the scale estimated in the previous stage and considering the residual functions of the two models. Finally, we choose the best estimate. This estimator is a two-dimensional extension of the proposed estimator for time series by Muler et al. (2009) to estimate the parameters of an AR-2D model with p-parameters. In 1964, Huber (1964) introduced the M-scale estimate. According to this proposal, given a sample $\boldsymbol{u} = (u_1, ..., u_N)$ with $u_i \in \mathbb{R}$, an M-estimate of the scale, $S_N(\boldsymbol{u})$, is defined by the value $s \in (0,\infty)$ that satisfies

$$\frac{1}{n} \sum_{i=1}^{n} \rho \left(\frac{u_i}{s} \right) = b, \tag{4.1}$$

where $b = E(\rho(Z))$ when Z has strictly unimodal symmetric density and ρ is a function that satisfies the property **P1**:

P1: $\rho(0) = 0$, $\rho(x) = \rho(-x)$, $\rho(x)$ is continuous, nonconstant and nondecreasing in $|x|$.
Let Y be a pure AR-2D process with innovation process ε; and let $\{y_{i,j}\}$ be a succession of observed data by this process, restricted to the strongly causal window W_M. We start from (3.5) trying to estimate σ through the strategy presented by (4.1) using the data and the two expressions for residuals ((3.4) and (3.7)).

Next, we define the BMM 2D estimate in the AR-2D model following the steps detailed below:

First Step: At this stage, we obtain a robust estimate of σ. To do this, we estimate σ using the residuals of the AR-2D model and the residuals of the BIP-AR 2D model. Finally, we choose the smallest of estimates.

Let ρ_1 be a function that satisfies **P1** and such that $b = E(\rho_1(Z))$ when Z has strictly unimodal symmetric density. Then we define an estimate of $\beta_0 \in \mathcal{B}$ as

$$\hat{\beta}_S = arg \min_{\beta\in\mathcal{B}} S_N(\boldsymbol{\varepsilon}_N(\beta)),$$

and the corresponding estimate of σ is given by

$$s_N = S_N(\boldsymbol{\varepsilon}_N(\hat{\beta}_S)),$$

where $\boldsymbol{\varepsilon}_N(\beta) = (\varepsilon_{M-1,M}(\beta), \varepsilon_{M-1,M-1}(\beta), \varepsilon_{M,M-1}(\beta), \varepsilon_{M-2,M}(\beta), ..., \varepsilon_{L,L}(\beta), \varepsilon_{L+1,L}(\beta), ..., \varepsilon_{M,L}(\beta))$, with $\varepsilon_{i,j}(\beta)$ as in (3.4) and $S_N(\boldsymbol{\varepsilon}_N(\beta))$ is the M-estimate of the scale based on ρ_1, b and the sample $\boldsymbol{\varepsilon}_N(\beta)$, that is,

$$\frac{1}{N} \sum_{(i,j)\in(W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta,\sigma)}{S_N(\boldsymbol{\varepsilon}_N(\beta))} \right) = b.$$

To obtain the estimate of σ for the residuals of the BIP-AR model, $\hat{\beta}^b_S$ was defined by the minimization of $S_N(\boldsymbol{\varepsilon}^b_N(\beta, \hat{\sigma}(\phi)))$ over $\mathcal{B}$, where $\hat{\sigma}(\phi)$ estimates the scale σ for every $\boldsymbol{\phi}$ as if these were the true values of the model and the $\varepsilon_{i,j}$'s were normal. In this case, since the M-scale σ coincides with the standard deviation of $\varepsilon_{i,j}$ and by (3.6), we obtain that

$$\sigma^2 = \frac{\sigma_y^2}{1 + \kappa^2 \sum_{(k,l)\in I\setminus\{(0,0)\}} \lambda^2_{k,l}(\boldsymbol{\phi})},$$

where $\kappa^2 = var(\eta(\varepsilon_{i,j}/\sigma))$ and $\sigma_y^2 = var(Y_{i,j})$. Let $\hat{\sigma}^2_y$ be a robust estimate of σ^2_y such that $\hat{\sigma}_y \to \sigma_y$ *a.e.* and $\kappa^2 = Var(\eta(Z))$ where $Z \sim N(0,1)$. Then we define

$$\hat{\sigma}^2(\boldsymbol{\phi}) = \frac{\hat{\sigma}_y^2}{1 + \kappa^2 \sum_{(k,l)\in I\setminus\{(0,0)\}} \lambda^2_{k,l}(\boldsymbol{\phi})}. \tag{4.2}$$

Later, the estimated scale s^b_N corresponding to the BIP-AR 2D model is defined by

$$\hat{\beta}^b_S = arg \min_{\beta\in\mathcal{B}} S_N(\boldsymbol{\varepsilon}^b_N(\beta, \hat{\sigma}(\boldsymbol{\phi}))),$$

and

$$s^b_N = S_N(\boldsymbol{\varepsilon}^b_N(\hat{\beta}^b_S, \hat{\sigma}(\hat{\boldsymbol{\phi}}^b_S)))$$

where, for simplicity, we denote $\tilde{\sigma} = \hat{\sigma}(\boldsymbol{\phi})$ and

$$\boldsymbol{\varepsilon}^b_N(\beta,\tilde{\sigma}) = (\varepsilon^b_{M-1,L}(\beta,\tilde{\sigma}), \varepsilon^b_{M-1,M-1}(\beta,\tilde{\sigma}), \varepsilon^b_{M,M-1}(\beta,\tilde{\sigma}), ..., \varepsilon^b_{L,L}(\beta,\tilde{\sigma}), \varepsilon^b_{L+1,L}(\beta,\tilde{\sigma}), ..., \varepsilon^b_{M,L}(\beta,\tilde{\sigma})),$$

with $\varepsilon^b_{i,j}(\beta,\tilde{\sigma})$ defined as in (3.7) and $S_N(\boldsymbol{\varepsilon}^b_N(\beta, \hat{\sigma}(\phi)))$ is the M-estimate of the scale based on ρ_1, b and the sample $\boldsymbol{\varepsilon}^b_N(\beta, \hat{\sigma}(\phi))$, that is

$$\frac{1}{N} \sum_{(i,j)\in(W_M \sim T)} \rho_1 \left(\frac{\varepsilon^b_{i,j}(\beta,\tilde{\sigma})}{S_N(\boldsymbol{\varepsilon}^b_N(\beta, \hat{\sigma}(\phi)))} \right) = b.$$

Finally, our estimate of σ is

$$s^*_N = \min(s_N, s^b_N).$$

Second Step: Consider a function ρ_2 that satisfies **P1** and such that $\rho_2 \leq \rho_1$. Let M_N and M^b_N functions defined on $\mathcal{B}$ given by:

$$M_N(\beta) = \frac{1}{N} \sum_{(i,j)\in(W_M \sim T)} \rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s^*_N} \right), \tag{4.3}$$

and

$$M^b_N(\beta) = \frac{1}{N} \sum_{(i,j)\in(W_M \sim T)} \rho_2 \left(\frac{\varepsilon^b_{i,j}(\beta, s^*_N)}{s^*_N} \right).$$

The corresponding M-estimates of the parameters for each function are:

$$\hat{\beta}_M = arg \min_{\beta\in\mathcal{B}} M_N(\beta) \qquad \text{and} \qquad \hat{\beta}^b_M = arg \min_{\beta\in\mathcal{B}} M^b_N(\beta).$$

Then, we define the BMM 2D estimate as

$$\hat{\beta}_M^* = \begin{cases} \hat{\beta}_M, & \text{if } M_N(\hat{\beta}_M) \leq M_N^b(\hat{\beta}_M^b) \\ \hat{\beta}_M^b, & \text{if } M_N(\hat{\beta}_M) > M_N^b(\hat{\beta}_M^b). \end{cases}$$

Remark 1. *In the strongly causal AR-2D model with three parameters, the $\lambda_{k,l}$'s have a known expression and can be calculated according to the work of Basu & Reinsel (1993). A general expression of $\lambda_{k,l}$'s for a model with more parameters can be obtained by calculating the coefficients of Taylor's multinomial expansion of $\mathbf{\Phi}(B_1, B_2)^{-1}$.*

5. Experiments

In this section, we will analyze the performance of the new BMM 2D estimator to estimate the parameters in an AR-2D model under additive contamination compared to other two-dimensional estimators (LS, M, GM, RA). For this purpose, we carried out some Monte Carlo experiments where the parameters were estimated when the process is affected by the additive contamination that adds a constant in a certain percentage of locations. All simulations were implemented in the statistical software R (R Core Team (2017)). The code was presented in the electronic supplementary material 1 of the work Britos & Ojeda (2018).

We considered the AR-2D model with two parameters and a mean of 0 given by:

$$Y_{i,j} = \phi_{1,0}^0 Y_{i-1,j} + \phi_{0,1}^0 Y_{i,j-1} + \varepsilon_{i,j}, \tag{5.1}$$

with $\phi_{1,0}^0 = 0.15$, $\phi_{0,1}^0 = 0.17$ and $\varepsilon = \{\varepsilon_{i,j}\}$ a white noise process where $\varepsilon_{i,j}$'s are identically distributed independent random variables with distribution $N(0,1)$. The unilateral polynomial associated with the model is:

$$\Phi_0(z_1, z_2) = 1 - \phi_{1,0}^0 z_1 - \phi_{0,1}^0 z_2.$$

It is important to mention that the set of parameters was chosen randomly satisfying the condition $|\phi_{1,0}^0| + |\phi_{0,1}^0| < 1$.

The study was conducted with contaminated observations under the model:

$$Z_{i,j} = (1 - \xi_{i,j}^\alpha) Y_{i,j} + \xi_{i,j}^\alpha (Y_{i,j} + k)$$

where $0 < \alpha < 1$, ξ^α is a process of ones and zeros such that $P(\xi_{i,j}^\alpha = 1) = \alpha$ and $P(\xi_{i,j}^\alpha = 0) = 1 - \alpha$, and k is a constant value.

5.1. Estimation of parameters

In this subsection we present some technical details for the computational calculation of the BMM 2D estimator as well as for the computation of the other bidimensional estimators.

According to the definition (4.1), we can write $S_N^2(\boldsymbol{\varepsilon}_N(\beta)) = \sum_{(i,j)\in(W_M \sim T)} r_{i,j}^2(\beta)$ where

$$r_{i,j}(\beta) = \frac{S_N(\boldsymbol{\varepsilon}_N(\beta))}{N^{1/2} b^{1/2}} \rho_1^{1/2} \left(\frac{\varepsilon_{i,j}(\beta)}{S_N(\boldsymbol{\varepsilon}_N(\beta))} \right).$$

Then,

$$\sum_{(i,j)\in(W_M\sim T)} r_{i,j}^2(\beta) = \sum_{(i,j)\in(W_M\sim T)} \frac{S_N^2(\boldsymbol{\varepsilon}_N(\beta))}{N.b}\rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{S_N(\boldsymbol{\varepsilon}_N(\beta))}\right)$$
$$= \frac{S_N^2(\boldsymbol{\varepsilon}_N(\beta))}{N.b} \sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{S_N(\boldsymbol{\varepsilon}_N(\beta))}\right)$$
$$= \frac{S_N^2(\boldsymbol{\varepsilon}_N(\beta))}{N.b}.N.b = S_N^2(\boldsymbol{\varepsilon}_N(\beta)).$$

Then, to calculate $\hat{\beta}_S$ any non-linear least squares algorithm can be used; in our case we use the Levenberg-Marquardt algorithm implemented in the *nls.lm* function of R package *minpack.lm*. This algorithm interpolates between the Gauss-Newton algorithm and the descent method (Marquardt (1963)).

Similarly we transform the problem of minimizing $S_N(\boldsymbol{\varepsilon}_N^b(\beta, \hat{\sigma}(\boldsymbol{\phi})))$ into a problem of nonlinear least squares.

For the calculation of $\hat{\beta}_M$ and $\hat{\beta}_M^b$ in the second stage, we used again the Levenberg-Marquardt algorithm using an idea similar to the previous one and taking as initial estimate the best estimate calculated in the first stage.

In the simulations we consider the following functions:

$$\rho_2(x) = \begin{cases} 0.5x^2, & \text{if } |x| \le 2; \\ 0.002x^8 - 0.052x^6 + 0.432x^4 - 0.972x^2 + 1.792, & \text{if } 2 < |x| \le 3; \\ 3.25, & \text{if } 3 < |x| \end{cases},$$

$\rho_1(x) = \rho_2(\frac{x}{0.405})$ and $\eta = \rho_2'$. The function ρ_1 was chosen such that if $b = max(\rho_1)/2$, then $b = E(\rho_1(Z))$ when $Z \sim \mathcal{N}(0,1)$ and so the scale matches the standard deviation for normal samples.

The same function ρ_2 was used to calculate M-estimate. In addition, for the implementation of the GM estimator the weights were set according to Allende et al. (2001) as:

$$l_{i,j} = 1, \quad \forall\ i,j \quad \text{y}$$
$$t_{i,j} = \frac{\psi_H((Y_{i-1,j}^2 + Y_{i-1,j-1}^2 + Y_{i,j-1}^2)/3)}{(Y_{i-1,j}^2 + Y_{i-1,j-1}^2 + Y_{i,j-1}^2)/3},$$

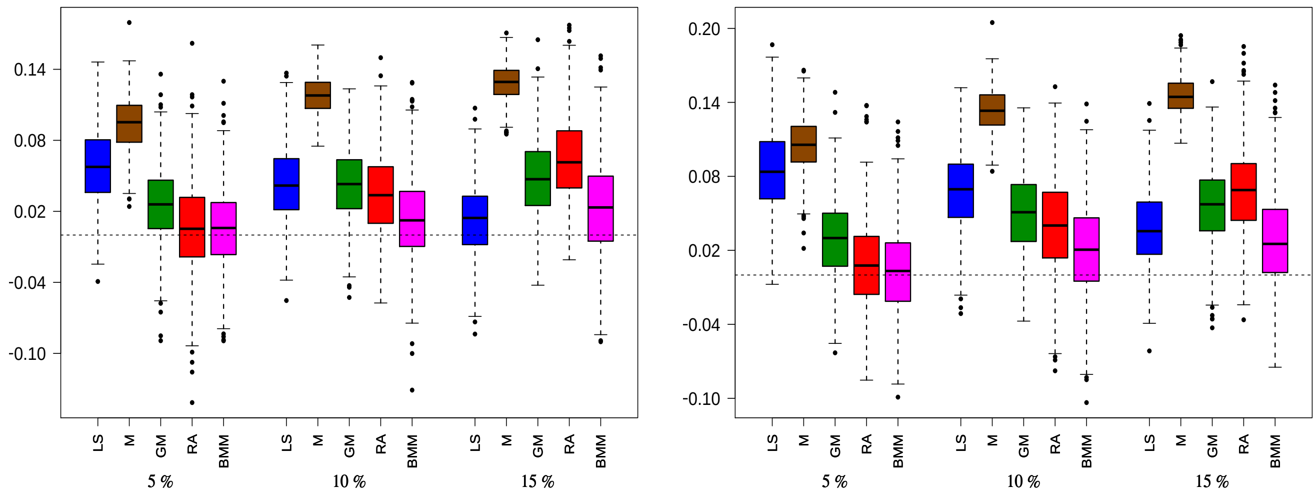
where ψ_H is the following Huber function (Kashyap & Eom (1988)):

$$\psi_H(x) = \begin{cases} x, & \text{if } |x| \le 1.5; \\ 1.5, & \text{if } 1.5 < x; \\ -1.5, & \text{if } x < -1.5. \end{cases}$$

5.2. Simulations

In each experiment, 500 simulations of the model were generated, and the mean sample value, the mean square error (MSE) and the sample variance were calculated. Three levels of contamination (5%, 10% and 15%) and different window sizes were considered: 8×8, 16×16, 32×32 and 57×57.

In the first experiment the capacity of the BMM 2D method was analyzed to estimate the parameters of the model when it is additively contaminated as in (5) with $k = 6$, considering a window of size 32×32 and varying the percentage of contamination among 5%, 10% and 15%, in comparison with the LS, M, GM and RA methods. Table 1 shows the estimated values for $\phi^0_{1,0}$ and $\phi^0_{0,1}$, using the five different procedures analyzed. Figure 3 shows the corresponding boxplots of the residues. It can be seen how as the level of contamination increases the estimates get worse. However, the BMM estimate remains close to the true parameter values even at the highest contamination.



(a) (b)

Figure 3: Boxplots of estimated residuals LS, M, GM, RA and BMM for the AR-2D model with two parameters and contaminated:(a) $\hat{\phi}_{1,0} - \phi_{1,0}$ with $\phi_{1,0} = 0.15$ and (b) $\hat{\phi}_{0,1} - \phi_{0,1}$ with $\phi_{0,1} = 0, 17$. The contamination is additive with $k = 6$, the window size is 32×32 and the percentage of contamination varies among 5, 10 and 15.

Table 1: Estimates of $\phi_{1,0} = 0.15$ and $\phi_{0,1} = 0.17$ in an AR-2D model with additive contamination ($k = 6$) and window size 32×32.

		$\phi_{1,0}$					$\phi_{0,1}$				
%		LS	M	GM	RA	BMM	LS	M	GM	RA	BMM
5%	$\hat{\phi}$	0.0923	0.0556	0.1238	0.1435	**0.1447**	0.0846	0.0647	0.1404	0.1623	**0.1669**
	mse_{ϕ}	0.0044	0.0094	0.0017	0.0016	**0.0013**	0.0084	0.0116	0.0018	0.0014	**0.0013**
	$\hat{\sigma}^2_{\hat{\phi}}$	0.0010	**0.0005**	0.0011	0.0016	0.0012	0.0011	**0.0005**	0.0010	0.0014	0.0013
10%	$\hat{\phi}$	0.1073	0.0320	0.1092	0.1179	**0.1366**	0.1020	0.0360	0.1206	0.1306	**0.1491**
	mse_{ϕ}	0.0028	0.0142	0.0026	0.0027	**0.0015**	0.0056	0.0182	0.0035	0.0034	**0.0020**
	$\hat{\sigma}^2_{\hat{\phi}}$	0.0010	**0.0003**	0.0009	0.0017	0.0014	0.0010	**0.0003**	0.0010	0.0018	0.0016
15%	$\hat{\phi}$	**0.1367**	0.0213	0.1026	0.0855	0.1268	0.1327	0.0251	0.1138	0.1000	**0.1427**
	mse_{ϕ}	**0.0011**	0.0168	0.0033	0.0054	0.0022	0.0024	0.0212	0.0041	0.0062	**0.0023**
	$\hat{\sigma}^2_{\hat{\phi}}$	0.0024	0.0212	0.0041	0.0062	**0.0023**	0.0010	**0.0002**	0.0010	0.0013	0.0016

In the second experiment, the percentage of additive contamination (with $k = 6$) was set at 10% and the window size of the observation varied among 8×8, 16×16, 32×32 and 54×54. In the windows of size 32×32 and 54×54, the BMM method was the best in the sense that its estimates were found closer to the true values. In the size 16×16 the best estimates came from the methods BMM and RA while in the size 8×8 the best were GM, RA and BMM although the estimator RA presented greater dispersion than the others. This behavior was deduced from the comparison of the values estimated by the BMM method with the respective estimates obtained by the other procedures. The results can be seen in Table 2 and in Figure 4.

Finally, in the third experiment, only the LS, RA and BMM estimates were compared (since they were

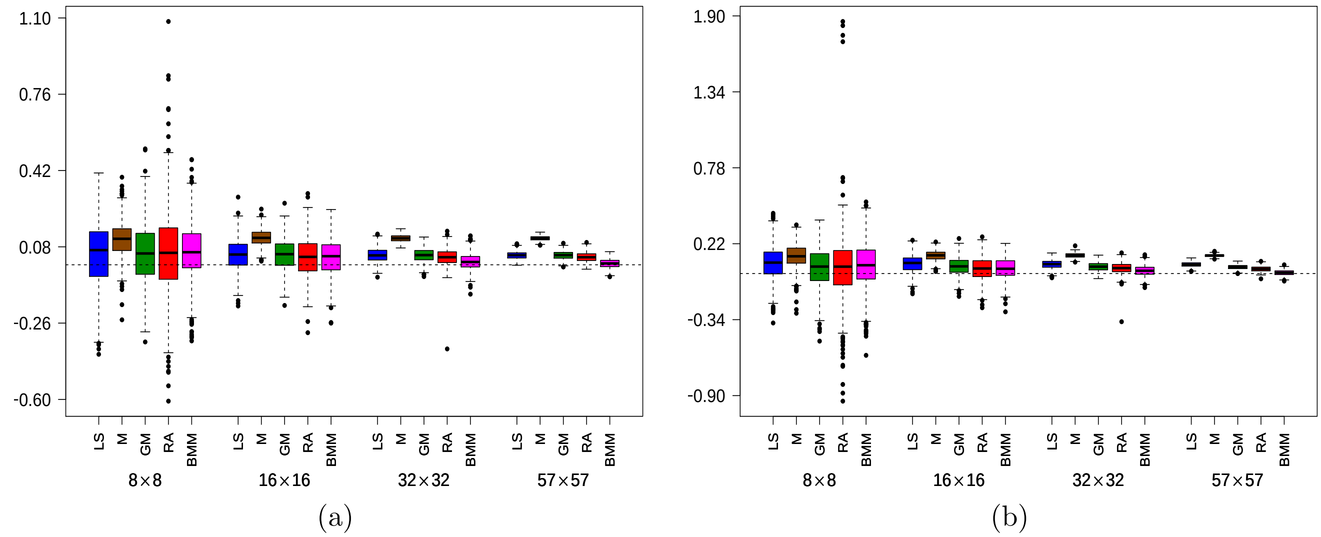
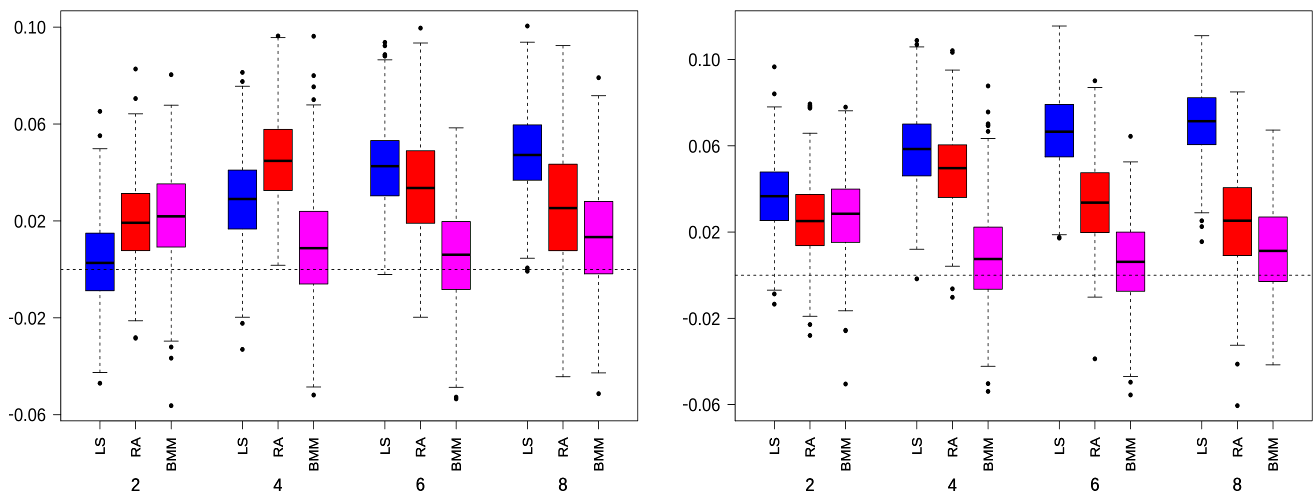


Figure 4: Boxplots of the estimated residuals LS, M, GM, RA and BMM for the AR-2D model with two parameters and contaminated:(a) $\hat{\phi}_{1,0} - \phi_{1,0}$ with $\phi_{1,0} = 0.15$ and (b) $\hat{\phi}_{0,1} - \phi_{0,1}$ with $\phi_{0,1} = 0, 17$. The contamination is additive with $k = 6$ at 10% and the window size varies among 8, 16, 32 and 57.

Table 2: Estimates of $\phi_{1,0} = 0.15$ and $\phi_{0,1} = 0.17$ in an AR-2D model with additive contamination to 10% ($k = 6$).

		$\phi_{1,0}$					$\phi_{0,1}$				
N		LS	M	GM	RA	BMM	LS	M	GM	RA	BMM
8	$\hat{\phi}$	**0.1028**	0.0373	0.1014	0.0935	0.0904	0.0929	0.0425	**0.1279**	0.1209	0.1150
	mse_{ϕ}	0.0227	**0.0188**	0.0210	0.0429	0.0219	**0.0223**	0.0244	0.0249	0.0752	0.0332
	$\hat{\sigma}^2_{\hat{\phi}}$	0.0205	**0.0061**	0.0187	0.0398	0.0184	0.0164	**0.0082**	0.0232	0.0729	0.0302
16	$\hat{\phi}$	0.1062	0.0300	0.1056	0.1181	**0.1206**	0.0972	0.0379	0.1185	**0.1347**	0.1327
	mse_{ϕ}	**0.0065**	0.0158	0.0070	0.0088	0.0075	0.0095	0.0188	**0.0070**	0.0084	0.0074
	$\hat{\sigma}^2_{\hat{\phi}}$	0.0046	**0.0014**	0.0050	0.0078	0.0067	0.0042	**0.0014**	0.0043	0.0072	0.0061
32	$\hat{\phi}$	0.1073	0.0320	0.1092	0.1179	**0.1366**	0.1020	0.0360	0.1206	0.1306	**0.1491**
	mse_{ϕ}	0.0028	0.0142	0.0026	0.0027	**0.0015**	0.0056	0.0182	0.0035	0.0034	**0.0020**
	$\hat{\sigma}^2_{\hat{\phi}}$	0.0010	**0.0003**	0.0009	0.0017	0.0014	0.0010	**0.0003**	0.0010	0.0018	0.0016
57	$\hat{\phi}$	0.1077	0.0319	0.1080	0.1169	**0.1447**	0.1032	0.0370	0.1222	0.1355	**0.1642**
	mse_{ϕ}	0.0021	0.0140	0.0021	0.0015	**0.0004**	0.0048	0.0178	0.0026	0.0016	**0.0004**
	$\hat{\sigma}^2_{\hat{\phi}}$	**3.1e-04**	0.0001	3.2e-04	4.6e-04	4.2e-04	3.2e-04	**9.1e-05**	2.8e-04	3.8e-04	3.9e-04

the ones that presented the best performance).The value of k varied among 2, 4, 6 and 8, the window size was set at 57×57 and the percentage of contamination at 10%. Table 3 and Figure 5 show the results. The LS estimate worsens as the k value increases. In all cases the RA estimates were far from the true parameter values. On the other hand, the BMM estimate is bad when $k = 2$ but for values of k greater than 2 this estimate improves substantially despite being confronted with more extreme contamination.



(a) (b)

Figure 5: Boxplots of the estimated residuals LS, RA and BMM for the AR-2D model with two parameters and contaminated:(a) $\hat{\phi}_{1,0} - \phi_{1,0}$ with $\phi_{1,0} = 0.15$ and (b) $\hat{\phi}_{0,1} - \phi_{0,1}$ with $\phi_{0,1} = 0, 17$. The contamination is additive with 10% of outliers, window size 57×57 and the value of k varies among $\{2, 4, 6, 8\}$.

Table 3: Estimates of $\phi_{1,0} = 0.15$ and $\phi_{0,1} = 0.17$ in an AR-2D model with additive contamination at 10% and window size 57×57.

		$\phi_{1,0}$			$\phi_{0,1}$		
k		LS	RA	BMM	LS	RA	BMM
2	$\hat{\phi}$	**0.1467**	0.1308	0.1279	0.1333	**0.1445**	0.1419
	mse_{ϕ}	**0.0003**	0.0007	0.0008	0.0016	**0.0010**	0.0011
	$\hat{\sigma}^2_{\hat{\phi}}$	**0.0003**	**0.0003**	**0.0003**	**0.0003**	**0.0003**	**0.0003**
4	$\hat{\phi}$	0.1211	0.1050	**0.1404**	0.1113	0.1214	**0.1619**
	mse_{ϕ}	0.0011	0.0023	**0.0006**	0.0037	0.0027	**0.0006**
	$\hat{\sigma}^2_{\hat{\phi}}$	**0.0003**	**0.0003**	0.0005	**0.0003**	**0.0003**	0.0005
6	$\hat{\phi}$	0.1077	0.1169	**0.1447**	0.1032	0.1355	**0.1642**
	mse_{ϕ}	0.0021	0.0015	**0.0004**	0.0048	0.0016	**0.0004**
	$\hat{\sigma}^2_{\hat{\phi}}$	**3.1e-04**	4.6e-04	4.2e-04	**3.2e-04**	3.8e-04	3.9e-04
8	$\hat{\phi}$	0.1018	0.1249	**0.1369**	0.0990	0.1454	**0.1575**
	mse_{ϕ}	0.0026	0.0012	**0.0006**	0.0053	0.0012	**0.0006**
	$\hat{\sigma}^2_{\hat{\phi}}$	3.1e-04	6.2e-04	4.4e-04	2.6e-04	5.6e-04	4.4e-04

6. Asymptotic Results

In this section we will present the main results of this work: Theorem 4 and Theorem 6, which demonstrate the consistency and asymptotic normality (respectively) of the BMM 2D estimator for pure AR-2D processes. To prove Theorem 4 we first show the consistency of the estimator $\hat{\beta}_S$ and then the consistency of the estimator $\hat{\beta}_M$ (Theorems 1 and 3, respectively). Theorem 2 relates the properties of the $\hat{\beta}_S$ estimator with the BMM 2D estimator, $\hat{\beta}^*_M$. Finally, to prove Theorem 6 we will need to prove the asymptotic normality of the estimator $\hat{\beta}_M$ (Theorem 5) before.

These results depend on several lemmas that will be enunciated in this chapter and whose demonstrations can be found in the Appendix. The demonstration strategies are inspired by some ideas presented in Muler et al. (2009), providing in this work demonstrations of the statements for the two-dimensional version.

Consider the following properties:

P1 $\rho(0) = 0$, $\rho(x) = \rho(-x)$, $\rho(x)$ is continuous, nonconstant and nondecreasing in $|x|$.

P2 The $Y = \{Y_{i,j}\}_{(i,j)\in\mathbb{Z}^2}$ process is a strictly stationary AR-2D and ergodic process defined over $(\Omega, \mathcal{A}, P)$ with parameter $\beta_0 = (\boldsymbol{\phi}_0, \mu_0) \in \mathcal{B}$ and innovation process $\varepsilon = \{\varepsilon_{i,j}\}_{(i,j)\in\mathbb{Z}^2}$.

P3 The random variables $\varepsilon_{i,j}$, of the innovation process ε, have an absolutely continuous distribution with a symmetric and strictly unimodal density.

P4 $P(\varepsilon_{i,j} \in C) < 1$ for any compact C.

P5 The η function is continuous, odd and bounded.

The following theorem establishes the consistency of the $\hat{\beta}_S$ estimator obtained in the first stage of the BMM 2D estimator definition.

Theorem 1. *Let Y be a process that satisfies* **P2** *with an innovation process ε satisfying* **P3**. *Let ρ_1 be a bounded function that satisfies* **P1** *with $\sup \rho_1 > b$ and $\psi_1 := \rho_1'$ bounded and continuous function. Then:*

i) $\hat{\beta}_S$ *is strongly consistent to estimate β_0, i.e.,* $\hat{\beta}_S \underset{N\to\infty}{\longrightarrow} \beta_0$ *a.e ..*

ii) $s_N \underset{N\to\infty}{\longrightarrow} s_0$ *a.e. where s_0 is defined by $E(\rho_1(\varepsilon_{i,j}/s_0)) = b$.*

The next theorem states that the scale estimator, s^*_N, converges almost everywhere to s_0.

Theorem 2. *Let Y be a process that satisfies* **P2** *with an innovation process ε satisfying* **P3** *and* **P4**. *Let ρ_1 be a bounded function that satisfies* **P1** *with $\sup \rho_1 > b$. Suppose that $\psi_1 := \rho_1'$ is bounded, continue and that η satisfies* **P5**. *Then, if Y is not a white noise,*

$$s^*_N = min(s_N, s^b_N) \to s_0 \quad a.e..$$

The next theorem demonstrates that the $\hat{\beta}_M$ estimator, obtained in the second stage of the BMM 2D estimator definition, is consistent to estimate β_0.

Theorem 3. *Let Y be a process that satisfies* **P2** *with an innovation process ε satisfying* **P3**. *Let ρ_1 and ρ_2 bounded functions satisfying* **P1**. *Let $\psi_i := \rho_i'$ be bounded and continuous functions with $i = 1, 2$ and such that $\sup \rho_1 > b$. Then,*

$$\hat{\beta}_M \longrightarrow \beta_0 \quad a.e..$$

Finally, the consistency of the BMM 2D estimator, $\hat{\beta}^*_M$, is demonstrated in the following theorem.

Theorem 4. *Suppose that the assumptions of Theorem 3,* **P4** *and* **P5** *hold. Then if the Y process is not a white noise, with probability 1, there exits N_0 such that $\hat{\beta}^*_M = \hat{\beta}_M$ $\forall N \geq N_0$ and then*

$$\hat{\beta}^*_M \longrightarrow \beta_0 \quad a.e..$$

In the following, Theorem 5 establishes the asymptotic normality of the $\hat{\beta}_M$ estimator obtained in the second stage of the BMM 2D estimator definition.

Theorem 5. *Suppose that the assumptions of Theorem 3 hold. Moreover, suppose that ψ_2' is a continuous and bounded function, $\sigma_\varepsilon^2 = E(\varepsilon_{i,j}^2) < \infty$ and the symmetric matrix $C = (C_{i,j})$ of dimension $[(L+1)^2 - 1] \times [(L+1)^2 - 1]$ defined by*

$$C = \begin{pmatrix} \sum_{(k,l)\in I} \lambda_{k,l}^2 & \sum_{(k,l)\in I} \lambda_{k,l}\lambda_{k,l-1} & \sum_{(k,l)\in I} \lambda_{k,l}\lambda_{k+1,l-1} & \cdots & \sum_{(k,l)\in I} \lambda_{k,l}\lambda_{k+1,l-L} \\ \sum_{(k,l)\in I} \lambda_{k,l}\lambda_{k,l-1} & \sum_{(k,l)\in I} \lambda_{k,l}^2 & \sum_{(k,l)\in I} \lambda_{k,l}\lambda_{k-1,l} & \cdots & \sum_{(k,l)\in I} \lambda_{k,l}\lambda_{k+1,l+1-L} \\ \sum_{(k,l)\in I} \lambda_{k,l}\lambda_{k+1,l-1} & \sum_{(k,l)\in I} \lambda_{k,l}\lambda_{k-1,l} & \sum_{(k,l)\in I} \lambda_{k,l}^2 & \cdots & \sum_{(k,l)\in I} \lambda_{k,l}\lambda_{k,l+1-L} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{(k,l)\in I} \lambda_{k,l}\lambda_{k+1,l-L} & \cdots & & & \sum_{(k,l)\in I} \lambda_{k,l}^2 \end{pmatrix}$$

is not singular. Then,

$$\sqrt{N}(\hat{\beta}_M - \beta_0) \xrightarrow{D} \mathcal{N}(0, D)$$

where

$$D = \frac{s_0^2 . E\left(\psi_2^2\left(\frac{\varepsilon_{i,j}}{s_0}\right)\right)}{E^2\left(\psi_2'\left(\frac{\varepsilon_{i,j}}{s_0}\right)\right)} . \begin{pmatrix} \sigma_\varepsilon^{-2} C^{-1} & 0 \\ 0 & \xi_0^{-2} \end{pmatrix},$$

with $\xi_0 = -1 + \sum_{(k,l)\in T} \phi_{k,l}^0$.

Finally, the following theorem proves the asymptotic normality of the BMM 2D estimator $\hat{\beta}_M^*$:

Theorem 6. *Suppose that the assumptions of Theorem 5,* **P4** *and* **P5** *hold. Then,*

$$\sqrt{N}(\hat{\beta}_M^* - \beta_0) \xrightarrow{D} \mathcal{N}(0, D)$$

where D is defined as in Theorem 5.

7. Concluding remarks

In this work, we have presented the new estimator for the parameters of the two-dimensional autoregressive model with p-parameters (BMM 2D estimator). It was initially developed for time series (Muler et al. (2009)) and extended to two-dimensional processes modeled by AR-2D models with three-parameter (Britos & Ojeda (2018)). The main result of this paper is that, we have established and demonstrated that under certain conditions, the BMM 2D estimator is consistent and asymptotically normal.

In Britos & Ojeda (2018), it was verified that the BMM 2D estimator competes successfully with other estimators known in the literature (LS, M, GM and RA 2D) under different replacement contamination schemes, both in accuracy and precision. In this paper, we study the performance of our estimator under additive contamination that adds a fixed constant compared to other existing estimators. The results of these experiments confirm the conclusion of Britos & Ojeda (2018). Among the robust estimators (M, GM and RA), the only one that has asymptotic properties is the RA 2D estimator; but it presents computational implementation difficulties and disadvantages related to its high computational cost, compared with the BMM 2D estimator.

In the following we outline future lines of research. First, we propose to study the theoretical properties of robustness of the BMM 2D estimator: breakpoint, maximum asymptotic bias and influence curve. These concepts have not been addressed for the BMM 2D estimator, and there are not studies of these properties known for proposals M, GM and RA.

Second, within the analysis and image processing, it would be very interesting to study in depth the potential and limitations of the BMM 2D estimator for edge detection, classification and restoration of digital images in combination with other techniques known in the area and the possible contribution to specific problems such as detection of burning areas.

Appendix

Suppose that we have Y a causal strongly stationary AR-2D process defined over $(\Omega, \mathcal{A}, P)$ with mean μ_0, innovation process $\varepsilon = \{\varepsilon_{i,j}\}$ and unilateral polynomial $\mathbf{\Phi}_0$ with support in T as in 3.1. Consider $\beta = (\boldsymbol{\phi}, \mu)$ in $\mathcal{B}$. Through simple accounts, the following expressions are obtained for the first and second derivatives of $\varepsilon_{i,j}(\beta)$:

$$\frac{\partial \varepsilon_{i,j}(\beta)}{\partial \phi_{k,l}} = -\dot{Y}_{i-k,j-l}, \qquad \forall (k,l) \in T, \tag{7.1}$$

$$\frac{\partial \varepsilon_{i,j}(\beta)}{\partial \mu} = -1 + \sum_{(k,l)\in T} \phi_{k,l} =: \xi, \tag{7.2}$$

$$\frac{\partial^2 \varepsilon_{i,j}(\beta)}{\partial \phi_{m,n} \partial \phi_{k,l}} = 0, \quad \forall (k,l), (m,n) \in T, \tag{7.3}$$

$$\frac{\partial^2 \varepsilon_{i,j}(\beta)}{\partial \phi_{k,l} \partial \mu} = 1, \quad \forall (k,l) \in T \quad \text{and} \tag{7.4}$$

$$\frac{\partial^2 \varepsilon_{i,j}(\beta)}{\partial^2 \mu} = 0. \tag{7.5}$$

We define the vector $\bigtriangledown(\varepsilon_{i,j}(\beta))$ that contains the first-order derivatives of $\varepsilon_{i,j}(\beta)$ as:

$$\begin{aligned} \bigtriangledown(\varepsilon_{i,j}(\beta)) &= \left(\frac{\partial \varepsilon_{i,j}(\beta)}{\partial \phi_{1,0}}, \frac{\partial \varepsilon_{i,j}(\beta)}{\partial \phi_{1,1}}, \frac{\partial \varepsilon_{i,j}(\beta)}{\partial \phi_{0,1}}, \ldots, \frac{\partial \varepsilon_{i,j}(\beta)}{\partial \mu} \right)^t \\ &= (-\dot{Y}_{i-1,j}, -\dot{Y}_{i-1,j-1}, -\dot{Y}_{i,j-1}, \ldots, \zeta)^t. \end{aligned}$$

The following definition will allow us to obtain some necessary results to prove the theorems of this work.

Definition 1. *Given the function ρ_1 that satisfies* **P1**, *$s(\beta)$ is defined as the function $s : \mathcal{B} \to \mathbb{R}$ given by*

$$E_{\beta_0}\left(\rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)} \right) \right) = b, \tag{7.6}$$

where b is such that $b = E(\rho_1(Z))$ when Z has symmetric and strictly unimodal density.

Note that with this new definition and because $\varepsilon_{i,j} = \varepsilon_{i,j}(\beta_0)$, it turns out that the value s_0, defined in the Theorem 1, is equal to $s(\beta_0)$.

Given the function ρ_2 used in the second stage of the definition of the BMM 2D estimator, we define the function ψ_2 as $\psi_2 := \rho_2'$. Below we will present some simple results that are obtained with the functions ρ_1, ρ_2 and ψ_2, which will allow us to prove the theorems.

Result 1.

$$\bigtriangledown \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \right) = \frac{1}{s_0} \psi_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) . \bigtriangledown (\varepsilon_{i,j}(\beta)) . \tag{7.7}$$

This result is a consequence of

$$\frac{\partial}{\partial \phi_{k,l}}\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right) = -\frac{1}{s_0}\psi_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\dot{Y}_{i-k,j-l}, \qquad \forall (k,l)\in T \quad \text{and} \tag{7.8}$$

$$\frac{\partial}{\partial \mu}\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right) = \frac{1}{s_0}\psi_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\xi. \tag{7.9}$$

Result 2.

$$\bigtriangledown^2\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right) = \frac{1}{s_0^2}\psi_2'\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right).\bigtriangledown\left(\varepsilon_{i,j}(\beta)\right)\bigtriangledown\left(\varepsilon_{i,j}(\beta)\right)^t + \frac{1}{s_0}\psi_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right).\bigtriangledown^2\left(\varepsilon_{i,j}(\beta)\right). \tag{7.10}$$

This result arises from the facts:

$$\frac{\partial^2}{\partial \phi_{m,n}\partial \phi_{k,l}}\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right) = \frac{1}{s_0^2}\psi_2'\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\dot{Y}_{i-k,j-l}\dot{Y}_{i-m,j-n}, \qquad \forall (k,l),(m,n)\in T, \tag{7.11}$$

$$\frac{\partial^2}{\partial \mu\partial \phi_{k,l}}\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right) = -\frac{\xi}{s_0^2}\psi_2'\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\dot{Y}_{i-k,j-l} + \frac{1}{s_0}\psi_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right), \qquad \forall (k,l)\in T \quad \text{and} \tag{7.12}$$

$$\frac{\partial^2}{\partial \mu}\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right) = \frac{\xi^2}{s_0^2}\psi_2'\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right). \tag{7.13}$$

Result 3. .

a)

$$E\left(\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right) = 0. \tag{7.14}$$

b)

$$E\left[\bigtriangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right)\right] = E\left(\frac{1}{s_0}\psi_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right).E\left[\bigtriangledown\left(\varepsilon_{i,j}(\beta)\right)\right] = 0. \tag{7.15}$$

The Result 3-(a) is because ψ_2 is odd and the distribution of $\varepsilon_{i,j}$ is symmetric. The Result 3-(b) arises from (7.7) and (7.14) and the fact that $\bigtriangledown\left(\varepsilon_{i,j}(\beta)\right)$ is independent of $\psi_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)$.

Result 4. *If V_0 is defined as*

$$V_0 = E\left[\bigtriangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right).\bigtriangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)^t\right], \tag{7.16}$$

then V_0 can be rewritten as

$$V_0 = E\left[\frac{1}{s_0^2}\psi_2^2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right).\bigtriangledown\left(\varepsilon_{i,j}(\beta)\right).\bigtriangledown\left(\varepsilon_{i,j}(\beta)\right)^t\right] = E\left(\frac{1}{s_0^2}\psi_2^2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right).E\left[\bigtriangledown\left(\varepsilon_{i,j}(\beta)\right)\bigtriangledown\left(\varepsilon_{i,j}(\beta)\right)^t\right] \tag{7.17}$$

where

$$\begin{aligned} E\left[\bigtriangledown\left(\varepsilon_{i,j}(\beta)\right)\bigtriangledown\left(\varepsilon_{i,j}(\beta)\right)^t\right] &= E\begin{pmatrix} (\dot{Y}_{i-s,j-t}\dot{Y}_{i-m,j-n})_{(s,t),(m,n)\in T} & (-\xi\dot{Y}_{i-s,j-t})_{(s,t)\in T} \\ (-\xi\dot{Y}_{i-s,j-t})_{(s,t)\in T} & \xi^2 \end{pmatrix} \\ &= \begin{pmatrix} (E(\dot{Y}_{i-s,j-t}\dot{Y}_{i-m,j-n}))_{(s,t),(m,n)\in T} & 0 \\ 0 & \xi^2 \end{pmatrix} \\ &= \begin{pmatrix} \tilde{C} & 0_{L\times 1} \\ 0_{1\times L} & \xi^2 \end{pmatrix} \end{aligned} \tag{7.18}$$

with $\tilde{C} = (E(\dot{Y}_{i-s,j-t}\dot{Y}_{i-m,j-n}))_{(s,t),(m,n)\in T}$ symmetric.

This last result was obtained by (7.7) and the independence of $\bigtriangledown(\varepsilon_{i,j}(\beta))$ with $\psi_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)$.

Remark 2. *It should be noted that*

$$E(\dot{Y}_{i-s,j-t}\dot{Y}_{i-m,j-n}) = \sum_{(k,l)\in I}\sum_{(q,r)\in I}\lambda_{k,l}\lambda_{q,r}E(\varepsilon_{i-s-k,j-t-l}\varepsilon_{i-m-q,j-n-r}). \tag{7.19}$$

Since the variables $\varepsilon_{i,j}$ś are i.i.d. with mean 0, the sum (7.19) becomes

$$E(\dot{Y}_{i-s,j-t}\dot{Y}_{i-m,j-n}) = \sigma_\varepsilon^2\sum_{(k,l)\in I}\lambda_{k,l}\lambda_{k+s-m,l+t-n}.$$

Therefore,

$$\tilde{C} = \sigma_\varepsilon^2 C$$

where C is like in Theorem 5.

The following lemma allow us to demonstrate Lemmas 4 and 11.

Lemma 1. *Let Y be a process that satisfies* **P2** *with an innovation process ε satisfying* **P3***. Then, for any $d>0$ we have there exists a stationary and ergodic process $W^0 = \{W^0_{i,j}\}_{(i,j)\in\mathbb{Z}^2}$ defined on $(\Omega,\mathcal{A},P)$ such that*

$$sup_{\beta\in\mathcal{B}_0\times[-d,d]}|\varepsilon_{i,j}(\beta)| \le W^0_{i,j}, \quad \forall (i,j)\in\mathbb{Z}^2.$$

Further, $E(|W^0_{i,j}|^2)<\infty$.

Proof of Lemma 1:

An equivalent expression to residual functions $\varepsilon_{i,j}(\beta)$ (3.4) is

$$\varepsilon_{i,j}(\beta) = Y_{i,j} - \sum_{(k,l)\in T}\phi_{k,l}Y_{i-k,j-l} + \mu\xi, \tag{7.20}$$

where $\xi = -1 + \sum_{(k,l)\in T}\phi_{k,l}$ (see (7.2)). Then

$$|\varepsilon_{i,j}(\beta)| \le |Y_{i,j}| + \sum_{(k,l)\in T}|\phi_{k,l}||Y_{i-k,j-l}| + |\mu\xi|.$$

Since $\beta\in\mathcal{B}_0\times[-d,d]$, $\sum_{(k,l)\in T}|\phi_{k,l}|<1$ and $|\mu|\le d$, then $|\xi|<1$. Later, as $\#(T)<\infty$, we have that

$$|\varepsilon_{i,j}(\beta)| \le |Y_{i,j}| + \sum_{(k,l)\in T}|Y_{i-k,j-l}| + 2d < \infty.$$

We define $W^0 = \{W^0_{i,j}\}_{(i,j)\in\mathbb{Z}^2}$ such that $W^0_{i,j} := |Y_{i,j}| + \sum_{(k,l)\in T}|Y_{i-k,j-l}| + 2d$. Further, as Y is a stationary process with finite first-order moment, W^0 is a stationary process and $E(({W^0}_{i,j})^2)<\infty$.

Let us prove that W^0 is ergodic. Let $g:(\mathbb{R}^{\mathbb{Z}^2},\mathcal{B}^{\mathbb{Z}^2})\to(\mathbb{R}^{\mathbb{Z}^2},\mathcal{B}^{\mathbb{Z}^2})$ measurable function defined by $g(X) = |B^{(0,0)}(X)| + \sum_{(k,l)\in T}|B^{(k,l)}(X)| + 2d$, where $|B^{(k,l)}(X)|(i,j) = |B^{(k,l)}(X)(i,j)|$. Then $W^0 = g(Y)$. Given $A\in\mathcal{I} = \{A'\in\mathcal{B}^{\mathbb{Z}^2} : A' \text{ is } B^{(k,l)} - \text{invariant } \forall (k,l)\in\mathbb{Z}^2\}$, we want to show that $P_{W^0}(A) = 0$ or 1.

Due to the fact that $P_{W^0}(A) = P(W^0\in A) = P(g(Y)\in A) = P(Y^{-1}(g^{-1}(A))) = P(Y\in g^{-1}(A))$, it is enough to prove that $g^{-1}(A)\in\mathcal{I}$.

First let us show that $g^{-1}(A)$ is contained in $B^{(s,t)}(g^{-1}(A)) = \{B^{(s,t)}(X) : g(X) \in A\}$, $\forall (s,t) \in \mathbb{Z}^2$. Let $X \in g^{-1}(A)$, later $g(X) \in A$, i.e., $|B^{(0,0)}(X)| + \sum_{(k,l)\in T} |B^{(k,l)}(X)| + 2d \in A$. Applying $B^{(-s,-t)}$ to the above, we have that $B^{(-s,-t)}|B^{(0,0)}(X)| + \sum_{(k,l)\in T} B^{(-s,-t)}|B^{(k,l)}(X)| + 2d \in B^{(-s,-t)}(A) = A$ (because $A \in \mathcal{I}$). Then,

$$|B^{(0,0)}(B^{(-s,-t)}(X))| + \sum_{(k,l)\in T} |B^{(k,l)}(B^{(-s,-t)}(X))| + 2d \in A,$$

that is, $g(B^{(-s,-t)}(X)) \in A$. Therefore, $X = B^{(s,t)}(B^{(-s,-t)}(X)) \in B^{(s,t)}(g^{-1}(A))$. Later, $g^{-1}(A) \subseteq B^{(s,t)}(g^{-1}(A))$. Similarly, we can prove that $B^{(s,t)}(g^{-1}(A)) \subseteq g^{-1}(A)$. Hence $g^{-1}(A) \in \mathcal{I}$ and $P(Y \in g^{-1}(A)) = 0$ or 1.

□

The next three lemmas prove Theorem 1. The first lemma sets properties on the function $s(\beta)$. The following two lemmas establish relations between the function $s(\beta)$ and the estimator of scale with the residuals of the AR-2D model ($S_N(\boldsymbol{\varepsilon}_N(\beta))$).

Lemma 2. *Let Y be a process that satisfies* **P2** *with an innovation process ε satisfying* **P3**. *Suppose that ρ_1 is a function satisfying* **P1** *and that the function s is as in (7.6). Then,*

i) if $\beta \neq \beta_0$, we have that $s_0 = s(\beta_0) < s(\beta)$.

ii) s is continuous.

Proof of Lemma 2:

Let us prove (i). Note that we always can choose a positive solution of the equation (7.6) because if s is a solution, $|s|$ is also a solution (since **P1** is satisfied). Later, $s(\beta)$ is positive.

Let $\beta = (\boldsymbol{\phi}, \mu) \neq \beta_0 = (\boldsymbol{\phi}_0, \mu_0)$. We have that

$$\begin{aligned}
\varepsilon_{i,j}(\beta) &= \Phi(B_1,B_2)(Y_{i,j} - \mu) \\
&= \Phi(B_1,B_2)(Y_{i,j} - \mu_0) + \Phi(B_1,B_2)(\mu_0 - \mu) \\
&= \Phi(B_1,B_2)\Phi_0(B_1,B_2)^{-1}\varepsilon_{i,j} + \left(1 - \sum_{(k,l)\in T} \phi_{k,l}\right)(\mu_0 - \mu) \\
&= \omega(B_1,B_2)\varepsilon_{i,j} + c.(\mu_0 - \mu), \qquad (7.21)
\end{aligned}$$

where $\omega(B_1,B_2) := \Phi(B_1,B_2)\Phi_0(B_1,B_2)^{-1}$ and $c = -\xi$. Since $\beta, \beta_0 \in \mathcal{B}$, then $\Phi(z_1,z_2)\Phi_0(z_1,z_2)^{-1}$ can be written over D^* as a power series sum: $1 + \sum_{(k,l)\in I\setminus\{(0,0)\}} w_{k,l} z_1^k z_2^l$ (see Guyon (1995)). Later,

$$\omega(B_1,B_2) = 1 + \sum_{(k,l)\in I\setminus\{(0,0)\}} w_{k,l} B_1^k B_2^l.$$

Put

$$\triangle_{i,j}(\beta) := \sum_{(k,l)\in I\setminus\{(0,0)\}} w_{k,l}\varepsilon_{i-k,j-l} + c.(\mu_0 - \mu).$$

Then, by the equation (7.21),

$$\varepsilon_{i,j}(\beta) = \varepsilon_{i,j} + \triangle_{i,j}(\beta).$$

Later,

$$\rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right) = \rho_1\left(\frac{\varepsilon_{i,j} + \triangle_{i,j}(\beta)}{s_0}\right)$$

and

$$E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right) = E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j} + \triangle_{i,j}(\beta)}{s_0}\right)\right).$$

Let $S(p,q)$ be defined as

$$S(p,q) = E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}+p}{q}\right)\right), \tag{7.22}$$

for $p, q \in \mathbb{R}$ with $q \neq 0$. Note that $S(p,q)$ is decreasing in $|q|$. Lemma 3.1 of Yohai (1985), shows that if **P1** and **P3** are satisfied, then for all p, $q \neq 0$ we have

$$S(0,q) \leq S(p,q),$$

and equality holds if and only if $p = 0$. Later,

$$\begin{aligned} E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right) &= S(\triangle_{i,j}(\beta), s_0) \\ &\geq S(0, s_0) = E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}}{s_0}\right)\right) = b \end{aligned}$$

and equality is valid if and only if $\triangle_{i,j}(\beta) = 0$ *a.e.*. Due to the identifiability of the AR-2D model, this happens if and only if $\beta = \beta_0$. Then $\beta \neq \beta_0$ implies

$$E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right) = S(\triangle_{i,j}(\beta), s_0) > b = S(\triangle_{i,j}(\beta), s(\beta)) = E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)}\right)\right)$$

and therefore, as $S(p,q)$ is decreasing in $|q|$ and $s(\beta)$ is a positive function, we have $s_0 = s(\beta_0) < s(\beta)$.

Now let us demonstrate (ii).

Let $\epsilon > 0$ be arbitrarily small. Due to the function $S(p,q)$ defined in the equation (7.22) is decreasing in $|q|$, we define $\beta_1 \in \mathcal{B}$ and $s_1 = s(\beta_1) > 0$ such that

$$E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1+\epsilon}\right)\right) = S(\triangle_{i,j}(\beta_1), s_1+\epsilon) < S(\triangle_{i,j}(\beta_1), s_1) = E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1}\right)\right) = b \tag{7.23}$$

and

$$E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1-\epsilon}\right)\right) = S(\triangle_{i,j}(\beta_1), s_1-\epsilon) > S(\triangle_{i,j}(\beta_1), s_1) = E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1}\right)\right) = b. \tag{7.24}$$

Let $q_1(\lambda)$ and $q_2(\lambda)$ be random variables defined by

$$q_1(\lambda) = \sup_{||\beta-\beta_1||\leq\lambda} \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_1+\epsilon}\right),$$

$$q_2(\lambda) = \sup_{||\beta-\beta_1||\leq\lambda} \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_1-\epsilon}\right).$$

Then, the succession of random variables $\{q_1(1/n)\}_{n\geq1}$ and $\{q_2(1/n)\}_{n\geq1}$ converge to $\rho_1\left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1+\epsilon}\right)$ and $\rho_1\left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1-\epsilon}\right)$ respectively.
By the Lebesgue's Dominated Convergence Theorem,

$$\lim_{n\longrightarrow\infty}\int_\Omega q_1(1/n)dP_{\beta_0}(\omega) = \int_\Omega \rho_1\left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1+\epsilon}\right)dP_{\beta_0}(\omega) = E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1+\epsilon}\right)\right) \tag{7.25}$$

and

$$\lim_{n\longrightarrow\infty}\int_\Omega q_2(1/n)dP_{\beta_0}(\omega) = \int_\Omega \rho_1\left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1-\epsilon}\right)dP_{\beta_0}(\omega) = E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1-\epsilon}\right)\right), \tag{7.26}$$

and hence, by (7.23) with (7.25) and by (7.24) with (7.26), respectively, there exists n_0 such that if $n \geq n_0$,

$$\int_\Omega \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_1+\epsilon}\right) dP_{\beta_0} \leq \int_\Omega q_1(1/n) dP_{\beta_0} < b = \int_\Omega \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)}\right) dP_{\beta_0} \tag{7.27}$$

and

$$\int_\Omega \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_1-\epsilon}\right) dP_{\beta_0} \geq \int_\Omega q_2(1/n) dP_{\beta_0} > b = \int_\Omega \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)}\right) dP_{\beta_0} \tag{7.28}$$

$$\forall \beta \text{ such that } ||\beta - \beta_0|| < \frac{1}{n} \leq \frac{1}{n_0}.$$

Later, let $\delta = \frac{1}{n_0}$. By the equations (7.27) and (7.28), if $||\beta - \beta_0|| < \delta$, we have

$$\int_\Omega \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_1+\epsilon}\right) dP_{\beta_0} < \int_\Omega \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)}\right) dP_{\beta_0}$$

and

$$\int_\Omega \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_1-\epsilon}\right) dP_{\beta_0} > \int_\Omega \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)}\right) dP_{\beta_0}.$$

Since ρ_1 is a positive function, we obtain that

$$\rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_1+\epsilon}\right) < \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)}\right) \quad a.e.$$

and

$$\rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_1-\epsilon}\right) > \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)}\right) \quad a.e..$$

Therefore, as $\rho_1(|u|)$ is non-decreasing and $s(\beta)$ is positive, we have $s_1 - \epsilon < s(\beta) < s_1 + \epsilon$, i.e., s is continuous in any $\beta_1 \in \mathcal{B}$. Later, s is continuous.

□

Lemma 3. *Under the assumptions of Theorem 1, for any $d > 0$, we have that*

$$\lim_{N\to\infty} \sup_{\beta \in \mathcal{B}_0 \times [-d,d]} |S_N(\boldsymbol{\varepsilon}_N(\beta)) - s(\beta)| = 0 \ a.e..$$

Proof of Lemma 3:

Let

$$h_1 = \inf_{\beta \in \mathcal{B}_0 \times [-d,d]} s(\beta) \quad \text{ and } \quad h_2 = \sup_{\beta \in \mathcal{B}_0 \times [-d,d]} s(\beta).$$

Then, by definition of $s(\beta)$, we have that $h_1 > 0$ and $h_2 < \infty$.

We consider the continuous function $f(y, \beta, c) = \rho_1\left(\frac{\Phi(B_1,B_2)(y-\mu)}{c}\right) - E_{\beta_0}\left(\rho_1\left(\frac{\Phi(B_1,B_2)(y-\mu)}{c}\right)\right)$ defined over $\mathbb{R} \times C$ with $C = \mathcal{B}_0 \times [-d,d] \times [h_1/2, 2h_2]$ compact. Due to $Y = \{Y_{i,j}\}_{(i,j)\in\mathbb{Z}^2}$ is a ergodic process we have that $E(f(Y,\beta,c)) = 0$ and further $\sup_{(\beta,c)\in C} |f(Y,\beta,c)| \leq K$ (because a continuous function on a compact is bounded). By Lemma 3 of Muler & Yohai (2002) we obtain

$$\lim_{N\longrightarrow\infty} \sup_{\beta\in\mathcal{B}_0\times[-d,d], c\in[h_1/2,2h_2]} \left| \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{c}\right) - E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{c}\right)\right)\right| = 0 \quad a.e., \tag{7.29}$$

where $N = \#(W_M \sim T) = (M - L + 1)^2$.

Let $0 \leq \epsilon \leq h_1/2$. We define the functions $g_i : \mathcal{B} \to \mathbb{R}$ for $i = 1, 2$ as

$$g_1(\beta) = E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)+\epsilon}\right)\right) \quad \text{ and } \quad g_2(\beta) = E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)-\epsilon}\right)\right).$$

By the definition of $s(\beta)$ and due to ρ_1 satisfies **P1**, we have that $g_1(\beta) < b$ and $g_2(\beta) > b$, $\forall \beta \in \mathcal{B}$. Since $\mathcal{B}_0$ is a compact set and, g_1 and g_2 are continuous (because $s(\beta)$ and $\varepsilon_{i,j}(\beta)$ are continuous and ρ_1 satisfies **P1**), we have

$$\kappa_1 := \sup_{\beta \in \mathcal{B}_0 \times [-d,d]} g_1(\beta) < b \quad \text{and} \quad \kappa_2 := \inf_{\beta \in \mathcal{B}_0 \times [-d,d]} g_2(\beta) > b.$$

Let $\delta = min(b - \kappa_1, \kappa_2 - b)$. By Eq. (7.29), there exists N_0 such that

$$\sup_{\beta \in \mathcal{B}_0 \times [-d,d], c \in [h_1/2, 2h_2]} \left| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{c} \right) - E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{c} \right) \right) \right| \leq \frac{\delta}{2} \quad a.e. \quad \forall N \geq N_0. \tag{7.30}$$

Note that $s(\beta) - \epsilon \in [h_1/2, 2h_2]$ since $h_1/2 = h_1 - \frac{h_1}{2} < s(\beta) - \frac{h_1}{2} < s(\beta) - \epsilon$ and also $s(\beta) - \epsilon < h_2 - \epsilon < h_2 + \epsilon < h_2 + h_2 = 2h_2$ due to the condition in ϵ and by the definitions of h_1 and h_2 respectively.

Hence, by (7.30), we obtain that

$$-\frac{1}{N} \sum_{(i,j) \in W_M \sim T} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) - \epsilon} \right) + E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) - \epsilon} \right) \right) \leq \frac{\delta}{2} \; a.e., \quad \forall \beta \in \mathcal{B}_0 \times [-d,d], \quad \forall N \geq N_0,$$

that is,

$$g_2(\beta) - \frac{\delta}{2} = E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) - \epsilon} \right) \right) - \frac{\delta}{2} \leq \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) - \epsilon} \right) \; a.e., \quad \forall \beta \in \mathcal{B}_0 \times [-d,d], \quad \forall N \geq N_0.$$

Later, taking infimum one gets

$$\inf_{\beta \in \mathcal{B}_0 \times [-d,d]} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) - \epsilon} \right) \geq \inf_{\beta \in \mathcal{B}_0 \times [-d,d]} g_2(\beta) - \frac{\delta}{2} = \kappa_2 - \frac{\delta}{2} \quad a.e.. \tag{7.31}$$

Furthermore, by the definition of δ, we know that $\kappa_2 - \frac{\delta}{2} \geq b + \frac{\delta}{2} > b$, and by Eq. (7.31) we have

$$\inf_{\beta \in \mathcal{B}_0 \times [-d,d]} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) - \epsilon} \right) > b = \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{S_N(\boldsymbol{\varepsilon}_N(\beta))} \right) \quad a.e.. \tag{7.32}$$

Similarly, we can observe that $s(\beta) + \epsilon \in [h_1/2, 2h_2]$ since $h_1/2 < h_1 < h_1 + \epsilon < s(\beta) + \epsilon$ and also $s(\beta) + \epsilon < h_2 + \epsilon < h_2 + \frac{h_1}{2} < h_2 + h_2 = 2h_2$.

Therefore, by (7.30), we obtain that

$$\frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) + \epsilon} \right) - E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) + \epsilon} \right) \right) \leq \frac{\delta}{2} \; a.e., \quad \forall \beta \in \mathcal{B}_0 \times [-d,d], \quad \forall N \geq N_1,$$

that is,

$$\frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) + \epsilon} \right) \leq E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) + \epsilon} \right) \right) + \frac{\delta}{2} = g_1(\beta) + \frac{\delta}{2} \; a.e., \quad \forall \beta \in \mathcal{B}_0 \times [-d,d], \quad \forall N \geq N_1, .$$

Later, taking supreme, we have

$$\sup_{\beta \in \mathcal{B}_0 \times [-d,d]} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) + \epsilon} \right) \leq \sup_{\beta \in \mathcal{B}_0 \times [-d,d]} g_1(\beta) + \frac{\delta}{2} = \kappa_1 + \frac{\delta}{2} \quad a.e.. \tag{7.33}$$

In addition, by the definition of δ, one knows that $\kappa_1 + \frac{\delta}{2} \geq b - \frac{\delta}{2} < b$, and by (7.33) we get that

$$\sup_{\beta \in \mathcal{B}_0 \times [-d,d]} \frac{1}{N} \sum_{(i,j)\in(W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)+\epsilon} \right) < b = \frac{1}{N} \sum_{(i,j)\in(W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{S_N(\boldsymbol{\varepsilon}_N(\beta))} \right) \quad a.e.. \tag{7.34}$$

Let us prove that there exists an $(i_0, j_0) \in (W_M \sim T)$ such that $\left|\frac{\varepsilon_{i_0,j_0}(\beta)}{s(\beta)+\epsilon}\right| < \left|\frac{\varepsilon_{i_0,j_0}(\beta)}{S_N(\boldsymbol{\varepsilon}_N(\beta))}\right|$ *a.e.*, $\forall \beta \in \mathcal{B}_0 \times [-d,d]$.

Suppose that $\left|\frac{\varepsilon_{i,j}(\beta)}{s(\beta)+\epsilon}\right| \geq \left|\frac{\varepsilon_{i,j}(\beta)}{S_N(\boldsymbol{\varepsilon}_N(\beta))}\right|$, $\forall (i,j) \in (W_M \sim T)$ and $\forall \beta \in \mathcal{B}_0 \times [-d,d]$. Hence, $s(\beta) + \epsilon > 0$, $S_N(\boldsymbol{\varepsilon}_N(\beta)) > 0$ and by the monotonicity of $\rho_1(|u|)$ we obtain that $\rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)+\epsilon}\right) = \rho_1\left(\frac{|\varepsilon_{i,j}(\beta)|}{s(\beta)+\epsilon}\right) \geq \rho_1\left(\frac{|\varepsilon_{i,j}(\beta)|}{S_N(\boldsymbol{\varepsilon}_N(\beta))}\right) = \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{S_N(\boldsymbol{\varepsilon}_N(\beta))}\right)$, $\forall (i,j) \in (W_M \sim T)$ and $\forall \beta \in \mathcal{B}_0 \times [-d,d]$, which means

$$\sum_{(i,j)\in(W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)+\epsilon} \right) \geq \sum_{(i,j)\in(W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{S_N(\boldsymbol{\varepsilon}_N(\beta))} \right), \quad \forall \beta \in \mathcal{B}_0 \times [-d,d]$$

but this is absurd by (7.34). Later, $\frac{|\varepsilon_{i_0,j_0}(\beta)|}{s(\beta)+\epsilon} < \frac{|\varepsilon_{i_0,j_0}(\beta)|}{S_N(\boldsymbol{\varepsilon}_N(\beta))}$ almost everywhere for any $(i_0, j_0) \in (W_M \sim T)$. Thus, $s(\beta) + \epsilon > S_N(\boldsymbol{\varepsilon}_N(\beta))$, *a.e.* $\forall \beta \in \mathcal{B}_0 \times [-d,d]$.

In the same way, (7.32) demonstrates that $s(\beta) - \epsilon < S_N(\boldsymbol{\varepsilon}_N(\beta))$, *a.e.* $\forall \beta \in \mathcal{B}_0 \times [-d,d]$ and for all $N > N_1$.

Hence, $|S_N(\boldsymbol{\varepsilon}_N(\beta)) - s(\beta)| \leq \epsilon$, *a.e.*, $\forall \beta \in \mathcal{B}_0 \times [-d,d]$ and $\forall N > max(N_0, N_1)$, that is,

$$\sup_{\beta \in \mathcal{B}_0 \times [-d,d]} |S_N(\boldsymbol{\varepsilon}_N(\beta)) - s(\beta)| \leq \epsilon \quad a.e. \quad \forall N > max(N_0, N_1).$$

Then,

$$\lim_{N\to\infty} \sup_{\beta \in \mathcal{B}_0 \times [-d,d]} |S_N(\boldsymbol{\varepsilon}_N(\beta)) - s(\beta)| = 0 \; a.e.$$

and the lemma is proven. □

Lemma 4. *Under the assumptions of Theorem 1, there exits $d > 0$ satisfying*

$$\liminf_{N\to\infty} \inf_{|\mu|>d, \beta \in \mathcal{B}} S_N(\boldsymbol{\varepsilon}_N(\beta)) > s_0 + 1 \quad a.e..$$

Proof of Lemma 4:

Given $\beta = (\boldsymbol{\phi}, \mu)$ with $\boldsymbol{\phi} \in \mathcal{B}_0$, let us call $\vartheta_{i,j}(\beta) := \varepsilon_{i,j}(\beta) - \varepsilon_{i,j}(\boldsymbol{\phi}, 0)$. By the definition of $\varepsilon_{i,j}(\beta)$ (7.20), we have

$$\vartheta_{i,j}(\beta) = \mu\xi = -\mu \left(1 - \sum_{(k,l)\in T} \phi_{k,l} \right), \qquad \forall (i,j) \in (W_M \sim T). \tag{7.35}$$

Furthermore, it is easy to see that

$$\vartheta_{i,j}(\beta) = \mu . \vartheta_{i,j}(\boldsymbol{\phi}, 1).$$

Using the compactness of $\mathcal{B}_0$, there exist $\delta > 0$ and $K_1 > 0$ such that for all $\boldsymbol{\phi} \in \mathcal{B}_0$, one obtains that

$$\delta \leq 1 - \sum_{(k,l)\in T} \phi_{k,l} \leq K_1. \tag{7.36}$$

Later, by (7.35) and using (7.36), we have

$$\inf_{\boldsymbol{\phi} \in \mathcal{B}_0} |\vartheta_{i,j}(\beta)| = \inf_{\boldsymbol{\phi} \in \mathcal{B}_0} \left| \mu \left(1 - \sum_{(k,l)\in T} \phi_{k,l} \right) \right| \geq \frac{\delta}{2} |\mu| \tag{7.37}$$

and by Lemma 1 we obtain

$$\sup_{\phi \in \mathcal{B}_0} |\varepsilon_{i,j}(\phi, 0)| \leq W^0_{i,j} \tag{7.38}$$

where $W^0 = \{W^0_{i,j}\}_{(i,j)\in\mathbb{Z}^2}$ is a stationary process.

Due to the fact that $\sup \rho_1 > b$ and $lim_{x\longrightarrow\infty}\rho_1(|x|) = \sup \rho_1$, there exist $k_0 > 0$ and $\lambda > 1$ such that for all $|x| \geq k_0$, we get that

$$\rho_1(x) \geq \lambda b. \tag{7.39}$$

Since $\{W^0_{i,j}\}_{(i,j)\in\mathbb{Z}^2}$ is strictly stationary, for each (i,j), the variables $W^0_{i,j}$'s have the same distribution, then there exits m such that

$$P(W^0_{i,j} < m/2) > \frac{1}{\lambda}. \tag{7.40}$$

Let k be defined by

$$k = max\left(\frac{m}{s_0+1}, k_0\right) \tag{7.41}$$

and let d be a constant such that $d \geq \max(4(s_0+1)k/\delta, |\mu_0|)$. Then, using (7.37) we obtain that

$$\inf_{\phi\in\mathcal{B}_0, |\mu|>d} |\vartheta_{i,j}(\beta)| \geq \frac{\delta}{2}d \geq 2(s_0+1)k. \tag{7.42}$$

Because ρ_1 satisfies **P1**, we get that

$$\inf_{\phi\in\mathcal{B}_0, |\mu|>d} \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_0+1}\right) \geq \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\inf_{\phi\in\mathcal{B}_0, |\mu|>d}\left|\frac{\varepsilon_{i,j}(\beta)}{s_0+1}\right|\right) I(A_{i,j}) \tag{7.43}$$

where $A_{i,j} = \{W^0_{i,j} < m/2\}$ and $I(A_{i,j})$ denotes the indicator function of $A_{i,j}$. By Eq. (7.38) and the definition of $\vartheta_{i,j}$, we can obtain that

$$|\varepsilon_{i,j}(\beta)| \geq |\vartheta_{i,j}(\beta)| - |\varepsilon_{i,j}(\phi, 0)| \geq |\vartheta_{i,j}(\beta)| - W^0_{i,j}. \tag{7.44}$$

Then Eqs. (7.41), (7.44) and (7.42) imply that

$$A_{i,j} \subset \{W^0_{i,j} < k.(s_0+1)\} \subset \left\{\inf_{|\mu|>d, \phi\in\mathcal{B}_0} |\varepsilon_{i,j}(\beta)| > k.(s_0+1)\right\}. \tag{7.45}$$

Due to the fact that $\rho_1 \geq 0$ and $\rho_1(|u|)$ is non-decreasing, by (7.45), we have

$$\begin{aligned}\frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\inf_{\phi\in\mathcal{B}_0, |\mu|>d}\left|\frac{\varepsilon_{i,j}(\beta)}{s_0+1}\right|\right) I(A_{i,j}) &\geq \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_1(k) I(A_{i,j}) \\ &= \frac{\rho_1(k)}{N} \sum_{(i,j)\in(W_M\sim T)} I(A_{i,j}).\end{aligned} \tag{7.46}$$

Since $W^0_{i,j}$ is an ergodic and stationary process, by the Ergodic Theorem (Guyon (1995)) and by (7.40) we get

$$\lim_{N\longrightarrow\infty} \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} I(A_{i,j}) = E(I(A_{i,j})) = P(A_{i,j}) > \frac{1}{\lambda} \tag{7.47}$$

in $\mathcal{L}^2$ and, hence, converges $a.e.$. Then, by (7.43) and (7.46):

$$\inf_{\phi\in\mathcal{B}_0, |\mu|>d} \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_0+1}\right) \geq \frac{\rho_1(k)}{N} \sum_{(i,j)\in(W_M\sim T)} I(A_{i,j}).$$

Taking lower limit and by expression (7.47) we obtain

$$\liminf_{N\longrightarrow\infty} \inf_{\boldsymbol{\phi}\in\mathcal{B}_0,|\mu|>d} \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_0+1}\right) > \frac{\rho_1(k)}{\lambda} \quad a.e.. \tag{7.48}$$

In addition, Eqs. (7.39) and (7.41) imply

$$\frac{\rho_1(k)}{\lambda} \geq b, \tag{7.49}$$

then, by (7.48) and (7.49):

$$\liminf_{N\longrightarrow\infty} \inf_{\boldsymbol{\phi}\in\mathcal{B}_0,|\mu|>d} \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_0+1}\right) > b = \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{S_N(\boldsymbol{\varepsilon}_N(\beta))}\right) \quad a.e..$$

Later, for N large enough, one has that

$$\frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{s_0+1}\right) > \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{S_N(\boldsymbol{\varepsilon}_N(\beta))}\right) \quad \forall\boldsymbol{\phi}\in\mathcal{B}_0, |\mu|>d \quad a.e..$$

By arguments similar to those used in Lemma 3 we have

$$s_0+1 < S_N(\boldsymbol{\varepsilon}_N(\beta)), \qquad \forall\boldsymbol{\phi}\in\mathcal{B}_0, |\mu|>d \quad a.e..$$

Therefore,

$$\inf_{|\mu|>d,\beta\in\mathcal{B}} S_N(\boldsymbol{\varepsilon}_N(\beta)) > s_0+1 \quad a.e. \text{ for } N \gg 0.$$

Finally, taking lower limit, we obtain

$$\liminf_{N\to\infty} \inf_{|\mu|>d,\beta\in\mathcal{B}} S_N(\boldsymbol{\varepsilon}_N(\beta)) > s_0+1 \quad a.e..$$

Later, the lemma is proven.

□

Next we demonstrate Theorem 1.

Proof of Theorem 1:

First let us prove some preliminary results.

Take $\epsilon > 0$ arbitrarily small and d as in Lemma 4. From Lemma 2, $s(\beta)$ is continuous and reaches an absolute minimum at $\beta_0 \in \mathcal{B}$. Let us see that there exists a $\gamma > 0$ such that

$$\min_{\beta\in\mathcal{B}_0\times[-d,d],||\beta-\beta_0||\geq\epsilon} s(\beta) \geq s_0+\gamma. \tag{7.50}$$

Since $\mathcal{B}_0$ is compact and $s(\beta)$ is continuous, $\forall\ \beta \in \mathcal{B}_0\times[-d,d]$ and $||\beta-\beta_0|| \geq \epsilon$, there exists a $0<\gamma<1$ such that $s(\beta)-s(\beta_0) > \gamma$, that is, $s(\beta) > \gamma + s(\beta_0) = \gamma + s_0$. Then

$$\min_{\beta\in\mathcal{B}_0\times[-d,d],||\beta-\beta_0||\geq\epsilon} s(\beta) \geq s_0+\gamma.$$

By Lemma 3, $\exists N_1$ such that $\sup_{\beta\in\mathcal{B}_0\times[-d,d]} |S_N(\boldsymbol{\varepsilon}_N(\beta)) - s(\beta)| < \gamma/4$ *a.e.* $\forall N > N_1$. Then

$$-S_N(\boldsymbol{\varepsilon}_N(\beta)) + s(\beta) \leq |-S_n(\boldsymbol{\varepsilon}_N(\beta)) + s(\beta)| < \frac{\gamma}{4}, \quad \forall\beta\in\mathcal{B}_0\times[-d,d];$$

later

$$s(\beta) - \frac{\gamma}{4} < S_N(\boldsymbol{\varepsilon}_N(\beta)), \quad \forall \beta \in \mathcal{B}_0 \times [-d, d];$$

and therefore

$$s(\beta) - \frac{\gamma}{4} < S_N(\boldsymbol{\varepsilon}_N(\beta)), \quad \forall \beta \in \mathcal{B}_0 \times [-d, d] \quad \text{and} \quad ||\beta - \beta_0|| \geq \epsilon.$$

Taking minimum in the last expression and by the equation (7.50) we have

$$\begin{aligned} \min_{\beta \in \mathcal{B}_0 \times [-d,d], ||\beta - \beta_0|| \geq \epsilon} S_N(\boldsymbol{\varepsilon}_N(\beta)) &\geq \min_{\beta \in \mathcal{B}_0 \times [-d,d], ||\beta - \beta_0|| \geq \epsilon} s(\beta) - \frac{\gamma}{4} \\ &\geq s_0 + \gamma - \frac{\gamma}{4} \\ &= s_0 + \frac{3}{4}\gamma \\ &> s_0 + \frac{\gamma}{2} \quad a.e.. \end{aligned}$$

Therefore,

$$\min_{\beta \in \mathcal{B}_0 \times [-d,d], ||\beta - \beta_0|| \geq \epsilon} S_N(\boldsymbol{\varepsilon}_N(\beta)) > s_0 + \frac{\gamma}{2} \quad a.e.. \tag{7.51}$$

Because $\beta_0 \in \mathcal{B}_0 \times [-d, d]$ ($\boldsymbol{\phi}_0 \in \mathcal{B}_0$ and $|\mu_0| \leq d$) and by Lemma 3

$$\sup_{\beta \in \mathcal{B}_0 \times [-d,d]} |S_N(\boldsymbol{\varepsilon}_N(\beta)) - s(\beta)| < \frac{\gamma}{4} \quad a.e.,$$

then

$$S_N(\boldsymbol{\varepsilon}_N(\beta_0)) - s(\beta_0) < \frac{\gamma}{4} \quad a.e.;$$

later

$$S_N(\boldsymbol{\varepsilon}_N(\beta_0)) < \frac{\gamma}{4} + s(\beta_0) = \frac{\gamma}{4} + s_0 \quad a.e..$$

Hence

$$S_N(\boldsymbol{\varepsilon}_N(\beta_0)) < \frac{\gamma}{4} + s_0 \quad a.e.. \tag{7.52}$$

By Lemma 4,

$$\sup_{N \geq 0} \left(\inf_{k \geq N} \left(\inf_{|\mu| > d, \boldsymbol{\phi} \in \mathcal{B}_0} (S_k(\boldsymbol{\varepsilon}_k(\beta))) \right) \right) > s_0 + 1 \quad a.e..$$

Due to $B_N := \inf_{k \geq N} \left(\inf_{|\mu| > d, \boldsymbol{\phi} \in \mathcal{B}_0} (S_k(\boldsymbol{\varepsilon}_k(\beta)) \right)$ is an increasing succession, $\exists N_2$ such that $B_N \geq s_0 + \gamma \ a.e. \quad (0 < \gamma < 1), \forall N \geq N_2$. Then

$$B_N := \inf_{k \geq N} \left(\inf_{|\mu| > d, \boldsymbol{\phi} \in \mathcal{B}_0} (S_k(\boldsymbol{\varepsilon}_k(\beta)) \right) \geq s_0 + \gamma \quad a.e., \quad \forall N \geq N_2;$$

later

$$\inf_{|\mu|>d,\phi\in\mathcal{B}_0}(S_k(\boldsymbol{\varepsilon}_k(\beta)))\geq s_0+\gamma \quad a.e., \quad \forall k\geq N \quad \text{and} \quad \forall N\geq N_2.$$

In particular,

$$\inf_{|\mu|>d,\phi\in\mathcal{B}_0}(S_N(\boldsymbol{\varepsilon}_N(\beta)))\geq s_0+\gamma \quad a.e., \quad \forall N\geq N_2. \tag{7.53}$$

Let us show (i), that is, $\hat{\beta}_S\to\beta_0 \quad a.e.$.

Given $\epsilon>0$, let $N_0=\max(N_1,N_2)$. So, if $N\geq N_0$, then (7.50), (7.51), (7.52) and (7.53) are satisfied. From (7.51) and (7.53) one gets that

$$\min_{\beta\in\mathcal{B},||\beta-\beta_0||\geq\epsilon} S_N(\boldsymbol{\varepsilon}_N(\beta))\geq s_0+\frac{\gamma}{2} \quad a.e.. \tag{7.54}$$

Furthermore, by the definition of $\hat{\beta}_S$,

$$S_N(\boldsymbol{\varepsilon}_N(\beta))\geq S_N(\boldsymbol{\varepsilon}_N(\hat{\beta}_S)) \quad \forall\beta\in\mathcal{B}.$$

In particular, when $\beta=\beta_0$,

$$S_N(\boldsymbol{\varepsilon}_N(\beta_0))\geq S_N(\boldsymbol{\varepsilon}_N(\hat{\beta}_S)). \tag{7.55}$$

By (7.52) and (7.55),

$$s_0+\frac{\gamma}{4}>S_N(\boldsymbol{\varepsilon}_N(\beta_0))\geq S_N(\boldsymbol{\varepsilon}_N(\hat{\beta}_S)) \quad a.e.. \tag{7.56}$$

If $||\hat{\beta}_S-\beta_0||\geq\epsilon$, by (7.54) we obtain

$$S_N(\boldsymbol{\varepsilon}_N(\hat{\beta}_S))\geq s_0+\frac{\gamma}{2} \quad a.e., \tag{7.57}$$

and besides Eqs. (7.54) and (7.56) imply

$$s_0+\frac{\gamma}{4}>S_N(\boldsymbol{\varepsilon}_N(\beta_0))\geq S_N(\boldsymbol{\varepsilon}_N(\hat{\beta}_S))\geq s_0+\frac{\gamma}{2} \quad a.e.,$$

which is absurd. Therefore, it must be that $||\hat{\beta}_S-\beta_0||\leq\epsilon \quad a.e.\ \forall N>N_0=max(N_1,N_2)$, i.e.,

$$\hat{\beta}_S \underset{N\to\infty}{\longrightarrow} \beta_0 \quad a.e..$$

Now let us see (ii), i.e., $s_N \underset{N\to\infty}{\longrightarrow} s_0=s(\beta_0) \quad a.e.$.

By the continuity of $s(\beta)$ and as $\hat{\beta}_S \underset{N\to\infty}{\longrightarrow} \beta_0 \quad a.e.$, then $s(\hat{\beta}_S) \underset{N\to\infty}{\longrightarrow} s(\beta_0)=s_0 \quad a.e.$.

In addition, $\hat{\beta}_S\in\mathcal{B}_0\times[-d,d]$ because $\hat{\beta}_S \underset{N\to\infty}{\longrightarrow} \beta_0$ $a.e.$, for $N\gg 0$. Later, by Lemma 3, $|S_N(\boldsymbol{\varepsilon}_N(\hat{\beta}_S))-s(\hat{\beta}_S)| \underset{N\to\infty}{\longrightarrow} 0$ $a.e.$.

Finally, due to the fact that

$$|s_N-s_0|\leq|S_N(\boldsymbol{\varepsilon}_N(\hat{\beta}_S))-s(\hat{\beta}_S)|+|s(\hat{\beta}_S)-s_0|,$$

we have that

$$s_N \underset{N\to\infty}{\longrightarrow} s_0 \quad a.e..$$

□

The following lemma will allow us to demonstrate Lemma 6.

Lemma 5. *Let* $C = \{f : \mathcal{B} \times \mathbb{R}_{>0} \to \mathbb{R}\}$, $A \in M^{(M+1)\times(M+1)}(C)$ *where* $A = [v_{i,j}(\beta, \sigma)]_{0\leq i,j\leq M}$ *such that* $\forall (i, j) \in (W_M \sim T)$,

$$v_{i,j}(\beta, \sigma) = \mu f_1^{i,j}(\boldsymbol{\phi}) + \sigma f_2^{i,j}(\beta, \sigma) + \sum_{(k,l)\in T} \phi_{k,l} v_{i-k,j-l}(\beta, \sigma) \tag{7.58}$$

where for each $(i, j) \in (W_M \sim T)$, $f_1^{i,j}$ *is a polynomial in* $\boldsymbol{\phi}$ *and* $f_2^{i,j}$ *is a bounded function on a compact. Then,* $\forall (i, j) \in (W_M \sim T)$:

$$v_{i,j}(\beta, \sigma) = \mu g_1^{i,j}(\boldsymbol{\phi}) + \sigma g_2^{i,j}(\beta, \sigma) + \sum_{(m,n)\in W_M^c} g_{m,n}^{i,j}(\boldsymbol{\phi}) v_{m,n}(\beta, \sigma) \tag{7.59}$$

where $W_M^c = W_M \setminus (W_M \sim T)$ *and such that* $\forall (i, j) \in (W_M \sim T)$ *and* $\forall (m, n) \in W_M^C$, $g_1^{i,j}$ *and* $g_{m,n}^{i,j}$ *are polynomials in* $\boldsymbol{\phi}$ *and* $\forall (i, j) \in (W_M \sim T)$, $g_2^{i,j}$ *is a bounded function on a compact.*

Proof of Lemma 5:

In the following, we will do an induction demonstration in M.

Suppose that $M = 2$. Then $L = 1$, $(W_M \sim T) = \{(2, 2), (1, 2), (1, 1), (2, 1)\}$ and the matrix A is

$$A = \begin{pmatrix} v_{0,2}(\beta, \sigma) & v_{1,2}(\beta, \sigma) & v_{2,2}(\beta, \sigma) \\ v_{0,1}(\beta, \sigma) & v_{1,1}(\beta, \sigma) & v_{2,1}(\beta, \sigma) \\ v_{0,0}(\beta, \sigma) & v_{1,0}(\beta, \sigma) & v_{2,0}(\beta, \sigma) \end{pmatrix}$$

where $v_{i,j}(\beta, \sigma)$ is as in (7.58). By (7.58), $v_{1,1}(\beta, \sigma)$ satisfies (7.59). Replacing the expression of $v_{1,1}(\beta, \sigma)$ in $v_{1,2}(\beta, \sigma)$ and $v_{2,1}(\beta, \sigma)$, we obtain that (7.59) is satisfied. In the same way, replacing the expression (7.59) of $v_{1,1}(\beta, \sigma)$, $v_{1,2}(\beta, \sigma)$ and $v_{2,1}(\beta, \sigma)$ in $v_{2,2}(\beta, \sigma)$ we have that $v_{2,2}(\beta, \sigma)$ satisfies (7.59). Later, the lemma holds with $M = 2$.

Suppose that the lemma is valid for $M = k$, let us show that for $M = k + 1$ too is valid.

Let A be a matrix of size $(M + 1) * (M + 1)$:

$$A = [v_{i,j}(\beta, \sigma)]_{0\leq i,j\leq(k+1)}.$$

We consider the submatrix of size $M * M$: $A_{1,k} = A[1 : (k + 1), 1 : (k + 1)] = [v_{s,t}^1(\beta, \sigma)]_{0\leq s,t\leq k}$ where $v_{s,t}^1(\beta, \sigma) = v_{s+1,t+1}(\beta, \sigma)$. Later, $v_{i,j}^1(\beta, \sigma)$ satisfies the condition (7.58) and by inductive hypothesis, for all

$$(i, j) \in (W_{1,M-1} \sim T) = \{(m, n) : (L + 1) \leq m \leq (k + 1), (L + 1) \leq n \leq (k + 1)\} :$$

$$v_{i-1,j-1}^1(\beta, \sigma) = v_{i,j}(\beta, \sigma) = \mu g_{1,1}^{i,j}(\boldsymbol{\phi}) + \sigma g_{1,2}^{i,j}(\beta, \sigma) + \sum_{(m,n)\in W_{1,M-1}^c} g_{1,m,n}^{i,j}(\boldsymbol{\phi}) v_{m,n}(\beta, \sigma) \tag{7.60}$$

where $W_{1,M-1}^c := W_{1,M-1} \setminus (W_{1,M-1} \sim T)$.

The idea is to prove:

(a) $v_{i,j}(\beta, \sigma)$ is written as (7.59) for all

$$(i, j) \in H_1 := \{(L, n) : L \leq n \leq (k + 1)\} \cup \{(m, L) : L + 1 \leq m \leq (k + 1)\}.$$

(b) $v_{i,j}(\beta, \sigma)$ is written as (7.59) for all

$$(i, j) \in (W_{1,M-1} \sim T).$$

Let us show (a).

I) Let us take the submatrix of size $M * M$: $A_{2,k} = A[0:k, 0:k] = [\tilde{v}_{i,j}(\beta,\sigma)]_{0\le i,j\le k}$ where $\tilde{v}_{i,j}(\beta,\sigma) = v_{i,j}(\beta,\sigma)$.

Due to the matrix $A_{2,k}$ satisfies the conditions of the lemma, by inductive hypothesis we have that for all $(i,j) \in (W_{2,M-1} \sim T) = \{(m,n) : L \le m \le k, L \le n \le k\}$:

$$\tilde{v}_{i,j}(\beta,\sigma) = v_{i,j}(\beta,\sigma) = \mu g_{2,1}^{i,j}(\boldsymbol{\phi}) + \sigma g_{2,2}^{i,j}(\beta,\sigma) + \sum_{(m,n)\in W_{2,M-1}^c} g_{2,m,n}^{i,j}(\boldsymbol{\phi}) v_{m,n}(\beta,\sigma) \tag{7.61}$$

where $W_{2,M-1}^c = W_{2,M-1} \setminus (W_{2,M-1} \sim T) \subset W_M^c := W_M \setminus (W_M \sim T)$. In particular, $v_{i,j}(\beta,\sigma)$ is written as (7.61) for all $(i,j) \in H_2 := \{(L,n) : L \le n \le k\} \cup \{(m,L) : L+1 \le m \le k\} \subset H_1$.

II) It remains to be seen that $v_{L,k+1}(\beta,\sigma)$ and $v_{k+1,L}(\beta,\sigma)$ are written as (7.59). By definition (7.58), one gets that

$$\begin{aligned}
v_{L,k+1}(\beta,\sigma) &= \mu f_1^{L,k+1}(\boldsymbol{\phi}) + \sigma f_2^{L,k+1}(\beta,\sigma) + \sum_{(m,n)\in T} \phi_{m,n} v_{L-m,k+1-n}(\beta,\sigma) \\
&= \mu f_1^{L,k+1}(\boldsymbol{\phi}) + \sigma f_2^{L,k+1}(\beta,\sigma) + \sum_{(s,t)\in V_{L,k+1}} \phi_{L-s,k+1-t} v_{s,t}(\beta,\sigma) \\
&= \mu f_1^{L,k+1}(\boldsymbol{\phi}) + \sigma f_2^{L,k+1}(\beta,\sigma) + \sum_{(s,t)\in V_{L,k+1}\cap W_M^c} \phi_{L-s,k+1-t} v_{s,t}(\beta,\sigma) \\
&+ \sum_{(s,t)\in V_{L,k+1}\cap(W_M\sim T)} \phi_{L-s,k+1-t} v_{s,t}(\beta,\sigma)
\end{aligned}$$

where $V_{L,k+1} = \{(m,n) : 0 \le m \le L, k+1-L \le n \le k+1, (m,n) \ne (L,k+1)\}$ and since $V_{L,k+1} \cap (W_M \sim T) \subset H_2$, then by (I), $v_{L,k+1}(\beta,\sigma)$ satisfies (7.59). In the same way, we can see that $v_{k+1,L}(\beta,\sigma)$ also satisfies (7.59). Later, (a) is demonstrated.

Let us prove (b). As seen in (7.60), for all $(i,j) \in (W_{1,M-1} \sim T)$,

$$\begin{aligned}
v_{i,j}(\beta,\sigma) &= \mu g_{1,1}^{i,j}(\boldsymbol{\phi}) + \sigma g_{1,2}^{i,j}(\beta,\sigma) + \sum_{(m,n)\in W_{1,M-1}^c} g_{1,m,n}^{i,j}(\boldsymbol{\phi}) v_{m,n}(\beta,\sigma) \\
&= \mu g_{1,1}^{i,j}(\boldsymbol{\phi}) + \sigma g_{1,2}^{i,j}(\beta,\sigma) + \sum_{(m,n)\in W_{1,M-1}^c\cap W_M^c} g_{1,m,n}^{i,j}(\boldsymbol{\phi}) v_{m,n}(\beta,\sigma) \\
&+ \sum_{(m,n)\in W_{1,M-1}^c\cap(W_M\sim T)} g_{1,m,n}^{i,j}(\boldsymbol{\phi}) v_{m,n}(\beta,\sigma) \\
&= \mu g_{1,1}^{i,j}(\boldsymbol{\phi}) + \sigma g_{1,2}^{i,j}(\beta,\sigma) + \sum_{(m,n)\in W_{1,M-1}^c\cap W_M^c} g_{1,m,n}^{i,j}(\boldsymbol{\phi}) v_{m,n}(\beta,\sigma) \\
&+ \sum_{(m,n)\in H_1} g_{1,m,n}^{i,j}(\boldsymbol{\phi}) v_{m,n}(\beta,\sigma).
\end{aligned}$$

Note that $W_{1,M-1}^c \cap W_M^c \subset W_M^c$. Furthermore, as we demonstrated in (a), one gets that

$$\sum_{(m,n)\in H_1} g_{1,m,n}^{i,j}(\boldsymbol{\phi}) v_{m,n}(\beta,\sigma) = \sum_{(m,n)\in H_1} g_{1,m,n}^{i,j}(\boldsymbol{\phi}) \left(\mu g_1^{m,n}(\boldsymbol{\phi}) + \sigma g_2^{m,n}(\beta,\sigma) + \sum_{(s,t)\in W_M^c} g_{s,t}^{m,n}(\boldsymbol{\phi}) v_{s,t}(\beta,\sigma) \right).$$

Later, (b) is proven and, therefore, the lemma is demonstrated.

□

The following three lemmas, allow us to prove Theorem 2. Lemma 6 gets a bound for the residues in the BIP-AR-2D model, which allowed us to demonstrate Lemmas 7 and 12. The results of Lemmas 7 and 8 are used directly in the proof of Theorem 2. These lemmas establish relationships between the M-scale at the true parameters and the M-estimators of scale under residuals of a BIP-AR-2D model.

Lemma 6. *Let Y be a process that satisfies* **P2**. *Given $d > 0$ and $\tilde{\sigma} > 0$, there exit constants $C > 0$ and $D > 0$ such that*

$$\sup_{\beta \in \mathcal{B}_0 \times [-d,d]} \sup_{0<\sigma\leq\tilde{\sigma}} |\varepsilon^b_{i,j}(\beta,\sigma) - Y_{i,j}| \leq C\tilde{\sigma} + D, \quad \forall (i,j) \in (W_M \sim T).$$

Proof of Lemma 6:

For $(i,j) \in (W_M \sim T)$, $\beta \in \mathcal{B}_0 \times [-d,d]$ and $\sigma \leq \tilde{\sigma}$, let

$$v_{i,j}(\beta,\sigma) = \varepsilon^b_{i,j}(\beta,\sigma) - Y_{i,j},$$

$$D_{i,j}(\beta,\sigma) = -\sum_{(k,l)\in T} \phi_{k,l}\eta\left(\frac{\varepsilon^b_{i-k,j-l}(\beta,\sigma)}{\sigma}\right)$$

and $v_{i,j}(\beta,\sigma) = -Y_{i,j}$, $\forall (i,j) \in W_M \setminus (W_M \sim T)$.

From the definition of $\varepsilon^b_{i,j}(\beta,\sigma)$ (see (3.7)), it follows that $\forall (i,j) \in (W_M \sim T)$, $v_{i,j}(\beta,\sigma)$ satisfies the recursive equation

$$v_{i,j}(\beta,\sigma) = \mu\left(-1 + \sum_{(k,l)\in T}\phi_{k,l}\right) + \sigma D_{i,j}(\beta,\sigma) + \sum_{(k,l)\in T}\phi_{k,l}v_{i-k,j-l}(\beta,\sigma).$$

Using Lemma 5, we have that $\forall (i,j) \in (W_M \sim T)$, $v_{i,j}(\beta,\sigma)$ is written as

$$\begin{aligned} v_{i,j}(\beta,\sigma) &= \mu f_1^{i,j}(\boldsymbol{\phi}) + \sigma f_2^{i,j}(\beta,\sigma) + \textstyle\sum_{(m,n)\in W_M\setminus(W_M\sim T)} f^{i,j}_{m,n}(\boldsymbol{\phi})v_{m,n}(\beta,\sigma) \\ &= \mu f_1^{i,j}(\boldsymbol{\phi}) + \sigma f_2^{i,j}(\beta,\sigma) - \textstyle\sum_{(m,n)\in W_M\setminus(W_M\sim T)} f^{i,j}_{m,n}(\boldsymbol{\phi})Y_{m,n} \end{aligned}$$

where

$\forall (i,j)$, $f_1^{i,j}$ is a polynomial,

$\forall (i,j)$ and $\forall (m,n) \in W_M \setminus (W_M \sim T)$, $f^{i,j}_{m,n}$ is a polynomial and

$\forall (i,j)$, $f_2^{i,j}$ is a bounded function on compact sets.

Due to the fact that $\mathcal{B}_0$ is compact, $\mu \in [-d,d]$ and η is bounded, there exist C_1 and C such that

$$\sup_{\beta \in \mathcal{B}_0 \times [-d,d]} \sup_{0<\sigma\leq\tilde{\sigma}} |\mu f_1^{i,j}(\boldsymbol{\phi}) + \sigma f_2^{i,j}(\beta,\sigma)| \leq dC_1 + \tilde{\sigma}C, \quad (i,j) \in (W_M \sim T).$$

In addition, as $\mathcal{B}_0$ is compact, there exists C_3 constant such that

$$\sup_{\beta \in \mathcal{B}_0 \times [-d,d]} \sup_{0<\sigma\leq\tilde{\sigma}} \left|\sum_{(m,n)\in W_M\setminus(W_M\sim T)} f^{i,j}_{m,n}(\boldsymbol{\phi})Y_{m,n}\right| \leq C_3, \quad (i,j) \in (W_M \sim T).$$

Therefore, there exist C and D positive constants such that

$$\sup_{\beta \in \mathcal{B}_0 \times [-d,d]} \sup_{0<\sigma\leq\tilde{\sigma}} |v_{i,j}(\beta,\sigma)| \leq C\tilde{\sigma} + D, \quad (i,j) \in (W_M \sim T).$$

Later, the lemma is proven.

□

Lemma 7. *Under the assumptions of Theorem 2, given $d > 0$, there exists $\delta > 0$ such that*

$$\liminf_{N\to\infty} \inf_{\beta\in\mathcal{B}_0\times[-d,d]} S_N(\boldsymbol{\varepsilon}^b_N(\beta,\hat{\sigma}(\boldsymbol{\phi}))) \geq s_0+\delta \quad a.e..$$

Proof of Lemma 7:

To demonstrate the lemma, we will prove that there exists $\delta > 0$ such that

$$\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s_0+\delta}\right) > \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))}{S_N(\varepsilon^b_N(\beta,\hat{\sigma}(\boldsymbol{\phi})))}\right) = b \quad a.e.. \tag{7.62}$$

As well as $\hat{\sigma}(\boldsymbol{\phi}) \leq \hat{\sigma}_Y$ and $\hat{\sigma}_Y \to \sigma_Y$ *a.e.*, then $\hat{\sigma}(\boldsymbol{\phi}) \leq \sigma_Y$ for all $N > N_0$ and we have

$$\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))}{s_0+\delta}\right) > \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))}{S_N(\varepsilon^b_N(\beta,\hat{\sigma}(\boldsymbol{\phi})))}\right) \; a.e. \quad \forall\beta\in\mathcal{B}_0\times[-d,d].$$

Then, $\exists(i_0,j_0)\in(W_M\sim T)$ such that

$$\frac{|\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))|}{s_0+\delta} > \frac{|\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))|}{S_N(\varepsilon^b_N(\beta,\hat{\sigma}(\boldsymbol{\phi})))} \; a.e. \quad \forall\beta\in\mathcal{B}_0\times[-d,d].$$

Later,

$$S_N(\varepsilon^b_N(\beta,\hat{\sigma}(\boldsymbol{\phi}))) \geq s_0+\delta$$

and taking the smallest, we have that

$$\inf_{\beta\in\mathcal{B}_0\times[-d,d]} S_N(\varepsilon^b_N(\beta,\hat{\sigma}(\boldsymbol{\phi}))) > s_0+\delta \quad a.e..$$

Therefore,

$$\liminf_{N\to\infty} \inf_{\beta\in\mathcal{B}_0\times[-d,d]} S_N(\varepsilon^b_N(\beta,\hat{\sigma}(\boldsymbol{\phi}))) \geq s_0+\delta \quad a.e..$$

Finally, the lemma is proven.

To prove (7.62), we consider two facts:
1) There exists $\delta > 0$ such that

$$\liminf_{N\to\infty} \sup_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq\sigma_Y} \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} E\left(\rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s_0+\delta}\right)\right) \geq b+\delta \quad a.e.. \tag{7.63}$$

2)

$$\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} \rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s_0+\delta}\right) - E\left(\rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s_0+\delta}\right)\right) \longrightarrow 0 \quad a.e..$$

Let us see 1).

By the definition of $\hat{\sigma}(\boldsymbol{\phi})$ (4.2), $\hat{\sigma}(\boldsymbol{\phi}) \leq \hat{\sigma}_Y$, where $\hat{\sigma}_Y$ is a robust estimator of σ_Y such that $lim_{N\to\infty}\hat{\sigma}_Y = \sigma_Y$ *a.e.*. Using Lemma 6, we can find constants $C_1 > 0$ and $C_2 > 0$ such that

$$\sup_{\beta\in\mathcal{B}_0\times[-d,d]} |\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi})) - Y_{i,j}| \leq C_1\hat{\sigma}_Y + C_2, \quad \forall(i,j)\in(W_M\sim T).$$

Since $lim_{N\to\infty}\hat{\sigma}_Y = \sigma_Y$ *a.e.*, with probability 1, $\exists N_0$ such that for $N > N_0$, $\forall(i,j)\in W_M$, there exist constants $\tilde{C}_1$ and $\tilde{C}_2$ such that

$$C_1\hat{\sigma}_Y + C_2 < \tilde{C}_1\sigma_Y + \tilde{C}_2 \quad a.e..$$

Let $D = \tilde{C}_1\sigma_Y + \tilde{C}_2$. Then $\forall\ N > N_0$, $\forall(i,j) \in (W_M \sim T)$ we have

$$\sup_{\beta\in\mathcal{B}_0\times[-d,d]} |\varepsilon^b_{i,j}(\beta,\sigma) - Y_{i,j}| \leq D \quad \forall\sigma \leq \sigma_Y \tag{7.64}$$

and, in particular,

$$\sup_{\beta\in\mathcal{B}_0\times[-d,d]} |\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi})) - Y_{i,j}| \leq D.$$

On the other hand, we can write the process $\{Y_{i,j}\}$ as $Y_{i,j} = \mu_0 + \varepsilon_{i,j} + v_{i,j}$, where $\{v_{i,j}\}$ is a stationary process that depends of $\varepsilon_{k,l}$, when $(k,l) \preceq (i,j)$ and $(k,l) \neq (i,j)$ $(v_{i,j} = \sum_{(k,l)\in T} \phi_{k,l}\varepsilon_{i-k,j-l})$.

As innovation process ε satisfies **P4**, then the distribution of $\varepsilon_{i,j}$ is unbounded, $(\forall\delta > 0,\ P(|\varepsilon_{i,j}| > \delta) > 0)$. In addition, as $Y_{i,j}$ is not a white noise, also we have that $v_{i,j}$ has unbounded distribution because: suppose that $v_{i,j}$ has bounded distribution, then $\exists\delta > 0$ such that

$$0 = P(|v_{i,j}| > \delta) \geq P(|Y_{i,j} - \varepsilon_{i,j}| > \delta + |\mu_0|) \geq P(|\varepsilon_{i,j}| - |Y_{i,j}| > \delta + |\mu_0|).$$

Since $\exists M > 0$ such that $|Y_{i,j}| < M$ (because $E(|Y_{i,j}|^2) = \sigma^2_Y < \infty$), then

$$0 = P(|\varepsilon_{i,j}| - |Y_{i,j}| > \delta + |\mu_0|) \geq P(|\varepsilon_{i,j}| > \delta + |\mu_0| + M).$$

Which is absurd since $\varepsilon_{i,j}$ has an unbounded distribution.

Let

$$u_{i,j}(\beta,\sigma) = \mu_0 + v_{i,j} + (\varepsilon^b_{i,j}(\beta,\sigma) - Y_{i,j}), \quad \forall(i,j) \in (W_M \sim T). \tag{7.65}$$

We can write

$$\varepsilon^b_{i,j}(\beta,\sigma) = Y_{i,j} + (\varepsilon^b_{i,j}(\beta,\sigma) - Y_{i,j}) = Y_{i,j} + u_{i,j}(\beta,\sigma) - \mu_0 - v_{i,j} = \varepsilon_{i,j} + u_{i,j}(\beta,\sigma). \tag{7.66}$$

Note that (7.64) and (7.65) imply that $\forall\ \beta \in \mathcal{B}_0 \times [-d,d]$, $\sigma \leq \sigma_Y$ and $\forall N > N_0$ we have

$$\begin{aligned}
|u_{i,j}(\beta,\sigma)| &= |v_{i,j} - (-\mu_0 - (\varepsilon^b_{i,j}(\beta,\sigma) - Y_{i,j}))| \\
&\geq |v_{i,j}| - |\mu_0 + (\varepsilon^b_{i,j}(\beta,\sigma) - Y_{i,j})| \\
&\geq |v_{i,j}| - |\mu_0| - |\varepsilon^b_{i,j}(\beta,\sigma) - Y_{i,j}| \\
&\geq |v_{i,j}| - |\mu_0| - D.
\end{aligned}$$

Therefore, $\forall(i,j) \in (W_M \sim T)$ and $\forall N > N_0$ we have

$$\{|v_{i,j}| > D + |\mu_0| + 1\} \subset \left\{\inf_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq\sigma_Y} |u_{i,j}(\beta,\sigma)| \geq 1\right\}.$$

Since $v_{i,j}$ is stationary and its distribution is unbounded (all they have the same distribution), we have

$$\gamma = P(|v_{i,j}| > D + |\mu_0| + 1) > 0.$$

Let us call $A_{i,j} = \{|v_{i,j}| > D + |\mu_0| + 1\}$.

According to the definition of s_0, we have $E_{\beta_0}(\rho_1(\varepsilon_{i,j}/s_0)) = b$.

As we saw in Lemma 2, if $S(u,q) = E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}+u}{q}\right)\right)$, for $q \neq 0$ and $u \neq 0$, one gets that $S(0,q) < S(u,q)$, i.e., $E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}}{q}\right)\right) < E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}+u}{q}\right)\right)$.

In particular, if $q = s_0 \neq 0$, we obtain

$$b = E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}}{s_0}\right)\right) < E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}+u}{s_0}\right)\right) \quad \forall u \neq 0.$$

This implies that

$$\inf_{|u|\geq 1} E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}+u}{s_0}\right)\right) > b.$$

Later,

$$(1-\gamma)E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}}{s_0}\right)\right) + \gamma \inf_{|u|\geq 1} E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}+u}{s_0}\right)\right) > (1-\gamma)b + \gamma b = b. \tag{7.67}$$

Let F be a function defined by $F(q) = (1-\gamma)S(0,q) + \gamma \inf_{|u|\geq 1} S(u,q)$.

F is a decreasing function in $|q|$ (because $S(p,q)$ is decreasing in $|q|$) and continuous (by the Lebesgue's Dominated Convergence Theorem). By (7.67), $F(s_0) > b$. Later, $\exists \delta > 0$ such that

$$F(s_0+\delta) \geq b + \delta,$$

that is,

$$(1-\gamma)E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}}{s_0+\delta}\right)\right) + \gamma \inf_{|u|\geq 1} E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}+u}{s_0+\delta}\right)\right) \geq b + \delta. \tag{7.68}$$

Let

$$h(u) = E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}+u}{s_0+\delta}\right)\right)$$

and

$$\gamma_N = \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} I(A_{i,j}).$$

For each $(i,j)\in(W_M\sim T)$ and $N > N_0$ we have

$$\inf_{|u|\geq 1} h(u) \leq \inf_{|u_{i,j}(\beta,\sigma)|\geq 1, \beta\in\mathcal{B}_0\times[-d,d], \sigma\leq\sigma_Y} h(u_{i,j}(\beta,\sigma)),$$

then

$$\begin{aligned}
I(A_{i,j})\inf_{|u|\geq 1} h(u) &\leq I(A_{i,j}) \inf_{|u_{i,j}(\beta,\sigma)|\geq 1, \beta\in\mathcal{B}_0\times[-d,d], \sigma\leq\sigma_Y} h(u_{i,j}(\beta,\sigma)) \\
&= \inf_{|u_{i,j}(\beta,\sigma)|\geq 1, \beta\in\mathcal{B}_0\times[-d,d], \sigma\leq\sigma_Y} I(A_{i,j})h(u_{i,j}(\beta,\sigma)) \\
&\leq \inf_{|v_{i,j}|\geq D+|\mu_0|+1, \beta\in\mathcal{B}_0\times[-d,d], \sigma\leq\sigma_Y} I(A_{i,j})h(u_{i,j}(\beta,\sigma)) \\
&= \inf_{\beta\in\mathcal{B}_0\times[-d,d], \sigma\leq\sigma_Y} I(A_{i,j})h(u_{i,j}(\beta,\sigma)).
\end{aligned}$$

Later, adding over $(W_M\sim T)$, one gets that

$$\begin{aligned}
\gamma_N \inf_{|u|\geq 1} h(u) &\leq \frac{1}{N}\sum_{(i,j)\in W_M\sim T} \inf_{\beta\in\mathcal{B}_0\times[-d,d], \sigma\leq\sigma_Y} I(A_{i,j})h(u_{i,j}(\beta,\sigma)) \\
&\leq \inf_{\beta\in\mathcal{B}_0\times[-d,d], \sigma\leq\sigma_Y} \frac{1}{N}\sum_{(i,j)\in W_M\sim T} I(A_{i,j})h(u_{i,j}(\beta,\sigma)) \\
&\leq \frac{1}{N}\sum_{(i,j)\in W_M\sim T} I(A_{i,j})h(u_{i,j}(\beta,\sigma)) \quad \forall \beta\in\mathcal{B}_0\times[-d,d], \sigma\leq\sigma_Y, N > N_0.
\end{aligned} \tag{7.69}$$

Further, as

$$\begin{aligned} h(u) &= E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}+u}{s_0+\delta}\right)\right) \\ &= S(u, s_0+\delta) \\ &\geq S(0, s_0+\delta) \\ &= E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}}{s_0+\delta}\right)\right) \\ &= h(0) \quad \forall u, \end{aligned} \tag{7.70}$$

then by (7.69) and (7.70), we obtain

$$\begin{aligned} &\gamma_N \inf_{|u|\geq 1} h(u) + (1-\gamma_N)h(0) \leq \\ &\qquad \leq \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} I(A_{i,j})h(u_{i,j}(\beta,\sigma)) + \left(1 - \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} I(A_{i,j})\right) h(u_{i,j}(\beta,\sigma)) \\ &\qquad = \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} I(A_{i,j})h(u_{i,j}(\beta,\sigma)) + \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} I(A^c_{i,j})h(u_{i,j}(\beta,\sigma)) \\ &\qquad = \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} \left(I(A_{i,j})h(u_{i,j}(\beta,\sigma)) + I(A^c_{i,j})h(u_{i,j}(\beta,\sigma))\right) \\ &\qquad = \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} h(u_{i,j}(\beta,\sigma)) \\ &\qquad \leq \sup_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq\sigma_Y} \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} h(u_{i,j}(\beta,\sigma)). \end{aligned}$$

Hence,

$$\gamma_N \inf_{|u|\geq 1} h(u) + (1-\gamma_N)h(0) \leq \sup_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq\sigma_Y} \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} h(u_{i,j}(\beta,\sigma)),$$

and as $\gamma_N \to \gamma$ *a.e.* (by Law of Large Numbers for Ergodic Processes (Guyon (1995))), and by Eq. (7.68) we have

$$\begin{aligned} \liminf_{N\to\infty} \sup_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq\sigma_Y} \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} h(u_{i,j}(\beta,\sigma)) &\geq \gamma \inf_{|u|\geq 1} h(u) + (1-\gamma)h(0) \\ &\geq b+\delta \quad a.e.. \end{aligned}$$

Then 1) is proven.

Now let us show 2).

Let

$$\begin{aligned} R_{i,j}(\beta,\sigma) \qquad &= \rho_1\left(\tfrac{\varepsilon^b_{i,j}(\beta,\sigma)}{s_0+\delta}\right) - h(u_{i,j}(\beta,\sigma)) \\ &= \rho_1\left(\tfrac{\varepsilon_{i,j}+u_{i,j}(\beta,\sigma)}{s_0+\delta}\right) - h(u_{i,j}(\beta,\sigma)). \end{aligned}$$

Let us consider the following order relationship in I: given $(k,l),(i,j)\in I$ we say that (k,l) is related to (i,j) under the relationship $\preccurlyeq$ (denoted $(k,l)\preccurlyeq(i,j)$) if and only if $k\leq i$ and $l\leq j$. Later, we define for each

$(i,j) \in I$, the σ-algebra $\mathcal{F}_{i,j}$ generated by the set of random variables $\{R_{k,l} : (k,l) \preccurlyeq (i,j)\}$. We can prove that, given $(i,j) \in I$, $\mathcal{F}_{k,l} \subseteq \mathcal{F}_{i,j}$ for each $(k,l) \preccurlyeq (i,j)$; that is, $\{\mathcal{F}_{k,l}\}$ is a non-decreasing succession of sub σ-algebra of $\mathcal{A}$.

In the same way as for time series (Muler et al. (2009)), it results that $\{R_{i,j}(\beta,\sigma), \mathcal{F}_{i,j}\}$ is a martingale difference succession.

Let us get in the conditions of Law of Large Numbers for Martingale Differences (Quang & Van Huan (2010)): let $\{b_{(i,j)}\}$ be a succession given by $b_{(i,j)} = (i-L+1)(j-L+1)$. This succession satisfies that $\triangle b_{(i,j)} = 0$, $\forall (i,j) \in \mathbb{Z}^2$ and $b_{(i,j)} \to \infty$ when $(i,j) \to \infty$ (with both orders: $\preceq$ and $\preccurlyeq$). In addition,

$$\sum_{(0,0)\preccurlyeq(i,j)} \frac{E(|R_{i,j}(\beta,\sigma)|^2)}{b^2_{(i,j)}} = \sum_{(i,j)\in I} \frac{E(|R_{i,j}(\beta,\sigma)|^2)}{b^2_{(i,j)}} \leq M \sum_{(i,j)\in I} \frac{1}{b^2_{(i,j)}} < \infty.$$

Later, by this theorem, we have

$$\frac{1}{b_{(M,M)}} \sum_{(0,0)\preccurlyeq(i,j)\preccurlyeq(M,M)} R_{i,j}(\beta,\sigma) = \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} R_{i,j}(\beta,\sigma) \to 0 \quad a.e.. \tag{7.71}$$

and 2) is proven.

Since $R_{i,j}(\beta,\sigma)$ is continuous and using compactness arguments, $\forall \epsilon > 0$, we can find $(\beta_l, \sigma_l, \delta_l)$, $1 \leq l \leq m_o$ with $\beta_l \in \mathcal{B}_0 \times [-d,d]$, $\sigma_l \leq \sigma_Y$, such that, if we define

$$V_l = \{(\beta,\sigma) : ||\beta - \beta_l|| + |\sigma - \sigma_l| \leq \delta_l\},$$

one obtains that $\mathcal{B}_0 \times [-d,d] \times [0,\sigma_Y] \subset \cup_{l=1}^{m_0} V_l$ (finite covering of $\mathcal{B}_0 \times [-d,d] \times [0,\sigma_Y]$) and

$$\sup_{(\beta,\sigma)\in V_l} \left| \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} [R_{i,j}(\beta,\sigma) - R_{i,j}(\beta_l,\sigma_l)] \right| \leq \epsilon, \quad \forall l = 1,...,m_0.$$

Later,

$$\sup_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq\sigma_Y} \left| \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} R_{i,j}(\beta,\sigma) \right| \leq \sum_{l=1}^{m_0} \sup_{(\beta,\sigma)\in V_l} \left| \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} [R_{i,j}(\beta,\sigma) - R_{i,j}(\beta_l,\sigma_l)] \right| + \left| \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} R_{i,j}(\beta_l,\sigma_l) \right|.$$

Taking upper limit on this last inequality and by (7.71) we have

$$\limsup_{N\to\infty} \sup_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq\sigma_Y} \left| \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} R_{i,j}(\beta,\sigma) \right| \leq m_0.\epsilon \quad a.e.,$$

and as this applies for all $\epsilon > 0$, we obtain

$$\limsup_{N\to\infty} \sup_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq\sigma_Y} \left| \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} R_{i,j}(\beta,\sigma) \right| = 0 \quad a.e.. \tag{7.72}$$

Finally, by (7.63) and (7.72), we almost everywhere have that,

$$
\begin{aligned}
b+\delta &\le \sup_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\le\sigma_Y} \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\left(\rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s_0+\delta}\right)-R_{i,j}(\beta,\sigma)\right)\\
&\le \sup_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\le\sigma_Y} \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s_0+\delta}\right)+\sup_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\le\sigma_Y}\left|\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}R_{i,j}(\beta,\sigma)\right|.
\end{aligned}
$$

Later, taking lower limit:

$$
\begin{aligned}
b<b+\delta &\le \liminf_{N\to\infty}\sup_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\le\sigma_Y} \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s_0+\delta}\right)+\\
&\limsup_{N\to\infty}\sup_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\le\sigma_Y}\left|\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}R_{i,j}(\beta,\sigma)\right|\\
&=\liminf_{N\to\infty}\sup_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\le\sigma_Y} \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s_0+\delta}\right)\quad a.e.,
\end{aligned}
$$

which implies that $\forall\beta\in\mathcal{B}_0\times[-d,d]$ and $\sigma\le\sigma_Y$:

$$
\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s_0+\delta}\right)>\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\hat\sigma(\boldsymbol{\phi}))}{S_N(\varepsilon^b_N(\beta,\hat\sigma(\boldsymbol{\phi})))}\right)\quad a.e.
$$

and the lemma is proven. □

Lemma 8. *Under the assumptions of Theorem 2, there exits $d>0$ such that*

$$
\liminf_{N\to\infty}\inf_{|\mu|>d,\boldsymbol{\phi}\in\mathcal{B}_0} S_N(\boldsymbol{\varepsilon}^b_N(\beta,\hat\sigma(\boldsymbol{\phi})))\ge s_0+1\quad a.e..
$$

Proof of Lemma 8:

As we saw in Lemma 6,

$$
\begin{aligned}
|v_{i,j}(\beta,\sigma)| &= |\varepsilon^b_{i,j}(\beta,\hat\sigma(\boldsymbol{\phi}))-Y_{i,j}|\\
&=\left|\mu f_1^{i,j}(\boldsymbol{\phi})+\hat\sigma(\boldsymbol{\phi})f_2^{i,j}(\beta,\hat\sigma(\boldsymbol{\phi}))-\sum_{(m,n)\in W_M\setminus(W_M\sim T)}f^{i,j}_{m,n}(\boldsymbol{\phi})Y_{m,n}\right|\\
&\ge|\mu||f_1^{i,j}(\boldsymbol{\phi})|-\left|-\hat\sigma(\boldsymbol{\phi})f_2^{i,j}(\beta,\hat\sigma(\boldsymbol{\phi}))+\sum_{(m,n)\in W_M\setminus(W_M\sim T)}f^{i,j}_{m,n}(\boldsymbol{\phi})Y_{m,n}\right|\\
&\ge|\mu||f_1^{i,j}(\boldsymbol{\phi})|-\hat\sigma(\boldsymbol{\phi})|f_2^{i,j}(\beta,\hat\sigma(\boldsymbol{\phi}))|-\left|\sum_{(m,n)\in W_M\setminus(W_M\sim T)}f^{i,j}_{m,n}(\boldsymbol{\phi})Y_{m,n}\right|.
\end{aligned}
$$

Later,

$$
|\varepsilon^b_{i,j}(\beta,\hat\sigma(\boldsymbol{\phi}))|\ge|\mu||f_1^{i,j}(\boldsymbol{\phi})|-\hat\sigma(\boldsymbol{\phi})|f_2^{i,j}(\beta,\hat\sigma(\boldsymbol{\phi}))|-\left|\sum_{(m,n)\in W_M\setminus(W_M\sim T)}f^{i,j}_{m,n}(\boldsymbol{\phi})Y_{m,n}\right|-|Y_{i,j}|. \tag{7.73}
$$

Due to $f^{i,j}_{m,n}$ are polynomials on the compact $\mathcal{B}_0$ $\forall(m,n) \in W_M \setminus (W_M \sim T)$ and $\forall(i,j) \in (W_M \sim T)$, we get that

$$\sup_{\boldsymbol{\phi}\in\mathcal{B}_0} \left| \sum_{(m,n)\in W_M\setminus(W_M\sim T)} f^{i,j}_{m,n}(\boldsymbol{\phi}) Y_{m,n} \right| \leq D_1.$$

Besides, as $0 < \hat{\sigma}(\boldsymbol{\phi}) \leq \hat{\sigma}_Y$ and $f^{i,j}_2$ are bounded functions on compact sets we have

$$\sup_{\boldsymbol{\phi}\in\mathcal{B}_0} \hat{\sigma}(\boldsymbol{\phi})|f^{i,j}_2(\beta, \hat{\sigma}(\boldsymbol{\phi}))| \leq \hat{\sigma}_Y.C,$$

and then

$$\inf_{\boldsymbol{\phi}\in\mathcal{B}_0} \left(- \left| \sum_{(m,n)\in W_M\setminus(W_M\sim T)} f^{i,j}_{m,n}(\boldsymbol{\phi}) Y_{m,n} \right| - \hat{\sigma}(\boldsymbol{\phi})|f^{i,j}_2(\beta, \hat{\sigma}(\boldsymbol{\phi}))| \right) \geq -D_1 - \hat{\sigma}_Y.C. \tag{7.74}$$

Since $\mathcal{B}_0$ is compact, $\forall\boldsymbol{\phi} \in \mathcal{B}_0$ there exists $\epsilon > 0$ such that $\epsilon \leq f^{i,j}_1(\boldsymbol{\phi})$. Later,

$$\inf_{\boldsymbol{\phi}\in\mathcal{B}_0} |\mu||f^{i,j}_1(\boldsymbol{\phi})| \geq \frac{\epsilon}{2}|\mu|. \tag{7.75}$$

Therefore, taking the lowest in (7.73) and using the values found in (7.75) and (7.74), we have

$$\begin{aligned}
\inf_{\boldsymbol{\phi}\in\mathcal{B}_0} |\varepsilon^b_{i,j}(\beta, \hat{\sigma}(\boldsymbol{\phi}))| &\geq \inf_{\boldsymbol{\phi}\in\mathcal{B}_0} |\mu||f^{i,j}_1(\boldsymbol{\phi})| - |Y_{i,j}| \\
&\quad + \inf_{\boldsymbol{\phi}\in\mathcal{B}_0} \left(- \left| \sum_{(m,n)\in W_M\setminus(W_M\sim T)} f^{i,j}_{m,n}(\boldsymbol{\phi}) Y_{m,n} \right| - \hat{\sigma}(\boldsymbol{\phi})|f^{i,j}_2(\beta, \hat{\sigma}(\boldsymbol{\phi}))| \right) \\
&\geq \frac{\epsilon}{2}|\mu| - |Y_{i,j}| - D_1 - \hat{\sigma}_Y.C.
\end{aligned} \tag{7.76}$$

Due to the fact that $\sup \rho_1 > b$ (by hypothesis of Theorem 2) and $\lim_{n\to\infty} \rho_1(|x|) = \sup \rho_1$, there exist k_0 and $\lambda > 1$ such that $\forall|x| \geq k_0$,

$$\rho_1(x) \geq \lambda b. \tag{7.77}$$

In addition, as $\lim_{n\to\infty} \hat{\sigma}_Y = \sigma_Y$ *a.e.*, there exist $\tilde{D}_1$ and $\tilde{C}$ such that $D_1 + \hat{\sigma}_Y.C \leq \tilde{D}_1 + \sigma_Y.\tilde{C}$.

Let k_1 be a constant such that the set defined as $C_{i,j} = \{|Y_{i,j}| \leq k_1 - \tilde{D}_1 - \tilde{C}\sigma_Y\}$ satisfies $P(C_{i,j}) \geq \frac{1}{\lambda}$.
Let $k = max(k_1/(s_0 + 1), k_0)$ and d a constant such that $d > \frac{4k(s_0+1)}{\epsilon}$.
Then, by the definition of k and (7.76), on $C_{i,j}$ one gets that

$$\begin{aligned}
\inf_{\boldsymbol{\phi}\in\mathcal{B}_0, |\mu|>d} |\varepsilon^b_{i,j}(\beta, \hat{\sigma}(\boldsymbol{\phi}))| &\geq \frac{\epsilon}{2}d - k_1 - \tilde{D}_1 - \tilde{C}\sigma_Y + \tilde{D}_1 + \tilde{C}\sigma_Y \\
&= \frac{\epsilon}{2}d - k_1 \\
&> k(s_0 + 1) > k.
\end{aligned} \tag{7.78}$$

For all $\beta = (\boldsymbol{\phi}, \mu)$ such that $|\mu| > d$, $\boldsymbol{\phi} \in \mathcal{B}_0$ we have that

$$\left| \frac{\varepsilon^b_{i,j}(\beta, \hat{\sigma}(\boldsymbol{\phi}))}{s_0 + 1} \right| \geq \inf_{|\mu|>d, \boldsymbol{\phi}\in\mathcal{B}_0} \left| \frac{\varepsilon^b_{i,j}(\beta, \hat{\sigma}(\boldsymbol{\phi}))}{s_0 + 1} \right|.$$

As ρ_1 satisfies **P1**, $\forall|\mu|>d, \boldsymbol{\phi}\in\mathcal{B}_0$ and $\forall(i,j)\in(W_M\sim T)$ we obtain

$$\begin{aligned}\rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))}{s_0+1}\right) &\geq \rho_1\left(\inf_{|\mu|>d,\boldsymbol{\phi}\in\mathcal{B}_0}\left|\frac{\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))}{s_0+1}\right|\right)\\ &\geq \rho_1\left(\inf_{|\mu|>d,\boldsymbol{\phi}\in\mathcal{B}_0}\left|\frac{\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))}{s_0+1}\right|\right)I(C_{i,j}).\end{aligned}$$

Then, $\forall|\mu|>d, \boldsymbol{\phi}\in\mathcal{B}_0$

$$\frac{1}{N}\sum_{(i,j)\in W_M\sim T}\rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))}{s_0+1}\right)\geq\frac{1}{N}\sum_{(i,j)\in W_M\sim T}\rho_1\left(\inf_{|\mu|>d,\boldsymbol{\phi}\in\mathcal{B}_0}\left|\frac{\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))}{s_0+1}\right|\right)I(C_{i,j}).$$

Later, taking the infimum one has:

$$\inf_{|\mu|>d,\boldsymbol{\phi}\in\mathcal{B}_0}\frac{1}{N}\sum_{(i,j)\in W_M\sim T}\rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))}{s_0+1}\right)\geq\frac{1}{N}\sum_{(i,j)\in W_M\sim T}\rho_1\left(\inf_{|\mu|>d,\boldsymbol{\phi}\in\mathcal{B}_0}\left|\frac{\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))}{s_0+1}\right|\right)I(C_{i,j}). \tag{7.79}$$

Furthermore, by (7.78) and due to ρ_1 satisfies **P1**, we obtain

$$\rho_1\left(\inf_{\boldsymbol{\phi}\in\mathcal{B}_0,|\mu|>d}|\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))|\right)\quad>\rho_1(k),\quad\forall(i,j)\in(W_M\sim T),$$

then, adding over $(W_M\sim T)$:

$$\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_1\left(\inf_{\boldsymbol{\phi}\in\mathcal{B}_0,|\mu|>d}|\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))|\right)I(C_{i,j})>\frac{\rho_1(k)}{N}\sum_{(i,j)\in(W_M\sim T)}I(C_{i,j});$$

and taking lower limit, we have

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_1\left(\inf_{\boldsymbol{\phi}\in\mathcal{B}_0,|\mu|>d}|\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))|\right)I(C_{i,j})\geq\liminf_{N\to\infty}\frac{\rho_1(k)}{N}\sum_{(i,j)\in(W_M\sim T)}I(C_{i,j}). \tag{7.80}$$

By the equation (7.77), the fact that $\{I(C_{i,j})\}$ is stationary and ergodic and $E(I(C_{i,j}))=P(C_{i,j})\geq 1/\lambda$, we have by the Ergodic Theorem (Guyon (1995)) that

$$\liminf_{N\to\infty}\frac{\rho_1(k)}{N}\sum_{(i,j)\in(W_M\sim T)}I(C_{i,j})=\rho_1(k)P(C_{i,j})\geq\lambda b\frac{1}{\lambda}=b\quad a.e..$$

Then, by (7.79) and (7.80) we have

$$\liminf_{N\to\infty}\inf_{|\mu|>d,\boldsymbol{\phi}\in\mathcal{B}_0}\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))}{s_0+1}\right)\geq b\quad a.e..$$

Later, for the last one, for N large enough and $\forall|\mu|>d, \boldsymbol{\phi}\in\mathcal{B}_0$,

$$\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))}{s_0+1}\right)\geq b=\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_1\left(\frac{\varepsilon^b_{i,j}(\beta,\hat{\sigma}(\boldsymbol{\phi}))}{S_N(\boldsymbol{\varepsilon}^b_N(\beta,\hat{\sigma}(\boldsymbol{\phi})))}\right)\quad a.e.,$$

and, by similar arguments to those used in Lemma 3, there exist $(i_0,j_0)\in(W_M\sim T)$ such that $\frac{|\varepsilon^b_{i_0,j_0}(\beta,\hat{\sigma}(\boldsymbol{\phi}))|}{s_0+1}>\frac{|\varepsilon^b_{i_0,j_0}(\beta,\hat{\sigma}(\boldsymbol{\phi}))|}{S_N(\boldsymbol{\varepsilon}^b_N(\beta,\hat{\sigma}(\boldsymbol{\phi})))}$ $a.e.$. Thus, for $N\gg 0$ and $\forall|\mu|>d, \boldsymbol{\phi}\in\mathcal{B}_0$ we have

$$S_N(\boldsymbol{\varepsilon}^b_N(\beta,\hat{\sigma}(\boldsymbol{\phi})))>s_0+1\quad a.e..$$

Therefore,

$$\liminf_{N\to\infty} \inf_{|\mu|>d, \phi\in\mathcal{B}_0} S_N(\boldsymbol{\varepsilon}^b_N(\beta, \hat{\sigma}(\boldsymbol{\phi}))) \geq s_0 + 1 \quad a.e..$$

and the lemma is proven.

□

Next we demonstrate Theorem 2.

Proof of Theorem 2:

Let $\tilde{\delta} > 0$ as in Lemma 7.

From Lemmas 7 and 8 we define $\delta = \min(\tilde{\delta}, 1)$ such that

$$\liminf_{N\to\infty} \inf_{\beta\in\mathcal{B}} S_N(\boldsymbol{\varepsilon}^b_N(\beta, \hat{\sigma}(\boldsymbol{\phi}))) \geq s_0 + \delta \quad a.e..$$

Then, $\exists N_1$ such that

$$\inf_{\beta\in\mathcal{B}} S_N(\boldsymbol{\varepsilon}^b_N(\beta, \hat{\sigma}(\boldsymbol{\phi}))) \geq s_0 + \delta \quad a.e., \forall N > N_1$$

or equivalently,

$$S_N(\boldsymbol{\varepsilon}^b_N(\beta, \hat{\sigma}(\boldsymbol{\phi}))) \geq s_0 + \delta \quad a.e., \quad \forall \beta \in \mathcal{B}, \quad \forall N > N_1.$$

In particular,

$$S_N(\boldsymbol{\varepsilon}^b_N(\hat{\beta}^b_S, \hat{\sigma}(\hat{\phi}^b_S))) \geq s_0 + \delta \quad a.e., \quad \forall N > N_1,$$

that is,

$$s^b_N \geq s_0 + \delta \quad a.e., \quad \forall N > N_1.$$

By Theorem 1 (ii), $s_N \longrightarrow s_0 \quad a.e.$. Thus, there exists N_2 such that $|s_N - s_0| < \delta$ $a.e.$ $\forall N > N_2$. Later,

$$s_N < \delta + s_0 \quad a.e., \quad \forall N > N_2.$$

Then, if $N > max(N_1, N_2)$ we obtain that $s_N < \delta + s_0 \leq s^b_N \quad a.e.$.

Therefore,

$$s^*_N = min(s_N, s^b_N) = s_N, \quad a.e. \quad \forall N > max(N_1, N_2)$$

and on account that $s_N \longrightarrow s_0$ $a.e.$, then $s^*_N \longrightarrow s_0$ $a.e.$ and the theorem is proven.

□

The next three lemmas intervene directly in the proof of Theorem 3. Lemma 9 presents properties about the expected function of the residuals of the AR-2D model on the function ρ_2 $(m(\beta))$. Lemmas 10 and 11 establish relationships between the function $m(\beta)$ and the objective function that determines the M-estimation of the parameters of the AR-2D model using the residual functions of that model.

Lemma 9. *Let Y be a process that satisfies* **P2** *with an innovation process ε satisfying* **P3***. Suppose that ρ_2 is a function satisfying* **P1***. Let $m : \mathcal{B} \to \mathbb{R}$ be a function defined by:*

$$m(\beta) := E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right).$$

Then,

i)

$$\beta_0 = arg \min_{\beta\in\mathcal{B}} m(\beta).$$

ii) m is a continuous function.

Proof of Lemma 9:

First, let us show (i).

Let $\beta = (\boldsymbol{\phi}, \mu) \neq \beta_0 = (\boldsymbol{\phi}_0, \mu_0)$. As in Lemma 2, we can express: $\varepsilon_{i,j}(\beta) = \varepsilon_{i,j} + \triangle_{i,j}(\beta)$ where $\triangle_{i,j}(\beta) = \sum_{(k,l)\in I\setminus\{(0,0)\}} w_{k,l}\varepsilon_{i-k,j-l} + \left(1 - \sum_{(k,l)\in T} \phi_{k,l}\right).(\mu_0 - \mu)$.

Then,

$$m(\beta) = E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right)$$
$$= E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j} + \triangle_{i,j}(\beta)}{s_0}\right)\right).$$

In addition, as $\tilde{S}(p,q) := E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}+p}{q}\right)\right)$ is decreasing in $|q|$ and $\tilde{S}(0,q) < \tilde{S}(p,q)$ for all $p \neq 0, q \neq 0$, then

$$m(\beta) = \tilde{S}(\triangle_{i,j}(\beta), s_0)$$
$$\geq \tilde{S}(0, s_0)$$
$$= E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}}{s_0}\right)\right)$$
$$= m(\beta_0).$$

Equality holds if and only if $\triangle_{i,j}(\beta) = 0$ *a.e.*. Due to the identifiability of the AR model this happens if and only if $\beta = \beta_0$. Later, if $\beta \neq \beta_0$ then $m(\beta) > m(\beta_0)$. Therefore, $\beta_0 = arg \min_{\beta\in\mathcal{B}} m(\beta)$.

Now let us see (ii):

The continuity of $m(\beta)$ is immediate since $\varepsilon_{i,j}(\beta)$ and ρ_2 are continuous and, in addition, ρ_2 is bounded. Then by Lebesgue's Dominated Convergence Theorem, m is continuous.

□

Lemma 10. *Let Y be a process that satisfies* **P2**. *Suppose that ρ_2 is a function satisfying* **P1**. *We define*

$$M_N(\beta) = \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*}\right),$$

as in (4.3). Then,

$$\lim_{N\to\infty} \sup_{\beta\in\mathcal{B}_0\times[-d,d]} \left| M_N(\beta) - E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right)\right| = 0 \quad a.e., \quad \forall d > 0.$$

Proof of Lemma 10:

By the Dominated Convergence Theorem, as ρ_2 is a continuous and bounded function and $\varepsilon_{i,j}(\beta)$ is continuous, we have that

$$M(\beta, v) = E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{v}\right)\right)$$

is a continuous function respect to the two variables.

Then, given $\epsilon > 0$ and $\beta \in \mathcal{B}_0 \times [-d,d]$, by the continuity of $M(\beta, v)$ in $v = s_0$ we have there exists $0 < \delta(\beta) < s_0$ such that if $|v - s_0| < \delta(\beta)$ then $|M(\beta, v) - M(\beta, s_0)| < \epsilon/2$ for each $\beta \in \mathcal{B}_0 \times [-d,d]$. By the compactness of $\mathcal{B}_0 \times [-d,d]$, we obtain that $\exists\delta > 0$ such that $|M(\beta, v) - M(\beta, s_0)| \leq \epsilon/2$, $\forall\beta \in \mathcal{B}_0 \times [-d,d]$ and $\forall v \in [s_0 - \delta, s_0 + \delta]$, and then

$$\sup_{\beta\in\mathcal{B}_0\times[-d,d], v\in[s_0-\delta, s_0+\delta]} |M(\beta, v) - M(\beta, s_0)| \leq \epsilon/2. \quad (7.81)$$

Let us consider the continuous function $f(y,\beta,v) = \left|\rho_2\left(\frac{\Phi(B_1,B_2)(y-\mu)}{v}\right) - E_{\beta_0}\left(\rho_2\left(\frac{\Phi(B_1,B_2)(y-\mu)}{v}\right)\right)\right|$ defined in $\mathbb{R} \times C_0$ with $C_0 = \{(\beta, v) : \beta \in \mathcal{B}_0 \times [-d,d], v \in [s_0 - \delta, s_0 + \delta]\}$ compact. Since $\{Y_{i,j}\}$ is an ergodic process, $E_{\beta_0}(f(Y,\beta,v)) = 0$ and $\sup_{(\beta,v)\in C_0} |f(Y,\beta,v)| \leq K$ with K a constant, by Lemma 3 of Muler & Yohai (2002), we have that

$$\lim_{N\longrightarrow\infty} \sup_{(\beta,v)\in C_0} \left| \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{v}\right) - E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{v}\right)\right)\right| = 0 \quad a.e.. \tag{7.82}$$

By Theorem 2, $\lim_{N\to\infty} s_N^* = s_0$ $a.e.$. Then with probability 1, there exists N_0 such that $\forall N > N_0$, $s_N^* \in [s_0 - \delta, s_0 + \delta]$ and

$$\sup_{\beta\in\mathcal{B}_0\times[-d,d]} \left| \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*}\right) - E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*}\right)\right)\right| < \epsilon/2 \quad a.e.. \tag{7.83}$$

By (7.81) and (7.83) we have that $\forall N > N_0$,

$$\begin{aligned}
&\left| \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*}\right) - E\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right)\right| \leq \\
&\leq \left| \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*}\right) - E\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*}\right)\right)\right| + \left|E\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*}\right)\right) - E\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right)\right| \\
&< \epsilon/2 + \epsilon/2, \quad \forall\beta \in \mathcal{B}_0 \times [-d,d] \quad a.e..
\end{aligned}$$

Then, taking the supreme:

$$\sup_{\beta\in\mathcal{B}_0\times[-d,d]} \left| \frac{1}{N} \sum_{(i,j)\in W_M\sim T} \rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*}\right) - E\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right)\right| < \epsilon \ a.e., \quad \forall N > N_0$$

and the lemma is proven.

□

Lemma 11. *Under the assumptions of Theorem 3, there exist $d > 0$ and $\delta > 0$ such that*

$$\liminf_{N\to\infty} \inf_{|\mu|>d,\boldsymbol{\phi}\in\mathcal{B}_0} M_N(\beta) > m(\beta_0) + \delta \quad a.e.,$$

where $m(\beta_0)$ is defined as in Lemma 9 and $M_N(\beta)$ as in Lemma 10.

Proof of Lemma 11:

Given $\beta = (\boldsymbol{\phi}, \mu)$ with $\boldsymbol{\phi} \in \mathcal{B}_0$. Let $\vartheta_{i,j}(\beta) = \varepsilon_{i,j}(\beta) - \varepsilon_{i,j}(\boldsymbol{\phi}, 0)$. By the definition of $\varepsilon_{i,j}(\beta)$ (7.20) we have that

$$\vartheta_{i,j}(\beta) = \mu\xi = -\mu\left(1 - \sum_{(k,l)\in T} \phi_{k,l}\right), \qquad \forall(i,j) \in (W_M \sim T). \tag{7.84}$$

Furthermore, it is easy to see that

$$\vartheta_{i,j}(\beta) = \mu.\vartheta_{i,j}(\boldsymbol{\phi}, 1).$$

Using the compactness of $\mathcal{B}_0$, there exist $\tilde{\delta} > 0$ and $K_1 > 0$ such that for all $\boldsymbol{\phi} \in \mathcal{B}_0$,

$$\tilde{\delta} \leq 1 - \sum_{(k,l)\in T} \phi_{k,l} \leq K_1. \tag{7.85}$$

Later, by (7.84) and using (7.85) we obtain that

$$\inf_{\phi\in\mathcal{B}_0}|\vartheta_{i,j}(\beta)| = \inf_{\phi\in\mathcal{B}_0}\left|\mu\left(1-\sum_{(k,l)\in T}\phi_{k,l}\right)\right| \geq \frac{\tilde{\delta}}{2}|\mu| \tag{7.86}$$

and by Lemma 1, we have

$$\sup_{\phi\in\mathcal{B}_0}|\varepsilon_{i,j}(\boldsymbol{\phi},0)| \leq W^0_{i,j} \tag{7.87}$$

where $W^0 = \{W^0_{i,j}\}$ is a stationary process.

Since the process ε satisfies **P3** and ρ_2 satisfies **P1**, then $m(\beta_0) = E(\rho_2(\varepsilon_{i,j}/s_0)) < \sup\rho_2$. Later, $\exists\delta > 0$ such that $\sup\rho_2 > m(\beta_0) + \delta$. In addition, as $lim_{x\longrightarrow\infty}\rho_2(|x|) = \sup\rho_2$, there exist $k_0 > 0$ and $\lambda > 1$ such that $\forall|x| \geq k_0$ we get that

$$\rho_2(x) \geq \lambda(m(\beta_0) + \delta). \tag{7.88}$$

As $\{W^0_{i,j}\}$ is a strictly stationary process, for each (i,j), the variables $W^0_{i,j}$'s have the same distribution, so there exists m such that

$$P(W^0_{i,j} < m/2) > \frac{1}{\lambda}. \tag{7.89}$$

We define k as

$$k = max\left(\frac{m}{s^*_N}, k_0\right) \tag{7.90}$$

and let d be a constant such that $d \geq \max(4s^*_N k/\tilde{\delta}, |\mu_0|)$. Then, using (7.86) we obtain that

$$\inf_{\phi\in\mathcal{B}_0,|\mu|>d}|\vartheta_{i,j}(\beta)| \geq \frac{\tilde{\delta}}{2}d \geq 2s^*_N k. \tag{7.91}$$

Due to the fact that ρ_2 satisfies **P1**, one gets that

$$\inf_{\phi\in\mathcal{B}_0,|\mu|>d}\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s^*_N}\right) \geq \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_2\left(\inf_{\phi\in\mathcal{B}_0,|\mu|>d}\left|\frac{\varepsilon_{i,j}(\beta)}{s^*_N}\right|\right)I(A_{i,j}) \tag{7.92}$$

where $A_{i,j} = \{W^0_{i,j} < m/2\}$ and $I(A_{i,j})$ denotes the indicator function of $A_{i,j}$. By the equation (7.87) and the definition of $\vartheta_{i,j}(\beta)$, we can write

$$|\varepsilon_{i,j}(\beta)| \geq |\vartheta_{i,j}(\beta)| - |\varepsilon_{i,j}(\boldsymbol{\phi},0)| \geq |\vartheta_{i,j}(\beta)| - W^0_{i,j}. \tag{7.93}$$

Then, Eqs. (7.90), (7.93) and (7.91) imply

$$A_{i,j} \subset \{W^0_{i,j} < k.s^*_N\} \subset \left\{\inf_{|\mu|>d,\phi\in\mathcal{B}_0}|\varepsilon_{i,j}(\beta)| > k.s^*_N\right\}. \tag{7.94}$$

As $\rho_2 \geq 0$, and $\rho_2(|u|)$ is non-decreasing, by (7.94) we have

$$\begin{aligned}\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_2\left(\inf_{\phi\in\mathcal{B}_0,|\mu|>d}\left|\frac{\varepsilon_{i,j}(\beta)}{s^*_N}\right|\right)I(A_{i,j}) &\geq \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_2(k)I(A_{i,j}) \\ &= \frac{\rho_2(k)}{N}\sum_{(i,j)\in(W_M\sim T)}I(A_{i,j}).\end{aligned} \tag{7.95}$$

Since $W^0_{i,j}$ is a stationary and ergodic process, by Ergodic Theorem (Guyon (1995)) and (7.89), we obtain that

$$\lim_{N\longrightarrow\infty}\frac{1}{N}\sum_{(i,j)\in W_M\sim T} I(A_{i,j}) = E(I(A_{i,j})) = P(A_{i,j}) > \frac{1}{\lambda} \tag{7.96}$$

in $\mathcal{L}^2$ and, therefore, it converges almost everywhere. Then, Eqs. (7.92) and (7.95), imply that

$$\inf_{\phi\in\mathcal{B}_0,|\mu|>d}\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s^*_N}\right) \geq \frac{\rho_2(k)}{N}\sum_{(i,j)\in(W_M\sim T)} I(A_{i,j}).$$

Taking lower limit and by expression (7.96), we have

$$\liminf_{N\longrightarrow\infty}\inf_{\phi\in\mathcal{B}_0,|\mu|>d}\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s^*_N}\right) > \frac{\rho_2(k)}{\lambda} \quad a.e.. \tag{7.97}$$

In addition, by (7.88) and (7.90) we obtain

$$\frac{\rho_2(k)}{\lambda} \geq m(\beta_0) + \delta. \tag{7.98}$$

Thus, from (7.97) and (7.98):

$$\liminf_{N\longrightarrow\infty}\inf_{\phi\in\mathcal{B}_0,|\mu|>d}\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s^*_N}\right) > m(\beta_0) + \delta \quad a.e..$$

That is,

$$\liminf_{N\to\infty}\inf_{|\mu|>d,\phi\in\mathcal{B}_0} M_N(\beta) > m(\beta_0) + \delta \quad a.e.$$

and the lemma is proven.

□

Next we demonstrate Theorem 3.

Proof of Theorem 3:

Given $\epsilon > 0$ and let d and δ be as in Lemma 11. By Lebesgue's Dominated Convergence Theorem, the function $m(\beta)$ defined in (9) is continuous since ρ_2 and $\varepsilon_{i,j}(\beta)$ are continuous functions and ρ_2 is bounded.

By Lemma 9, $m(\beta)$ reaches an absolute minimum at β_0. In addition, $m(\beta)$ is continuous at β_0 because $m(\beta)$ is continuous $\forall\beta \in \mathcal{B}$. Thus, $\exists 0 < \gamma < \delta$ such that if $||\beta-\beta_0|| \geq \epsilon$ then $m(\beta)-m(\beta_0) = |m(\beta)-m(\beta_0)| > \gamma$. Later,

$$m(\beta) > \gamma + m(\beta_0), \quad \forall\beta \text{ such that } ||\beta-\beta_0|| \geq \epsilon,$$

that is,

$$\min_{||\beta-\beta_0||\geq\epsilon} m(\beta) > \gamma + m(\beta_0).$$

Consequently,

$$\min_{\beta\in\mathcal{B}_0\times[-d,d],||\beta-\beta_0||\geq\epsilon} m(\beta) \geq \min_{||\beta-\beta_0||\geq\epsilon} m(\beta) > \gamma + m(\beta_0).$$

Hence,

$$\min_{\beta\in\mathcal{B}_0\times[-d,d],||\beta-\beta_0||\geq\epsilon} m(\beta) > \gamma + m(\beta_0). \tag{7.99}$$

On the other hand, by Lemma 10, $\exists N_1$ such that $\sup_{\beta\in\mathcal{B}_0\times[-d,d]}|M_N(\beta) - m(\beta)| < \frac{\gamma}{4}$ *a.e.* $\forall N > N_1$. Then,

$$-M_N(\beta)+m(\beta)<\frac{\gamma}{4}\quad a.e.,\quad \forall\beta\in\mathcal{B}_0\times[-d,d],$$

further,

$$m(\beta)-\frac{\gamma}{4}<M_N(\beta)\quad a.e.,\quad \forall\beta\in\mathcal{B}_0\times[-d,d].$$

Then,

$$m(\beta)-\frac{\gamma}{4}<M_N(\beta)\quad a.e.,\quad \forall\beta\in\mathcal{B}_0\times[-d,d] \text{ and } ||\beta-\beta_0||\geq\epsilon.$$

Therefore, from the previous inequality and using (7.99), we have

$$\begin{aligned}\min_{\beta\in\mathcal{B}_0\times[-d,d],||\beta-\beta_0||\geq\epsilon} M_N(\beta) &> \min_{\beta\in\mathcal{B}_0\times[-d,d],||\beta-\beta_0||\geq\epsilon} m(\beta)-\frac{\gamma}{4}\\ &>\gamma+m(\beta_0)-\frac{\gamma}{4}=m(\beta_0)+\frac{3}{4}\gamma\\ &>m(\beta_0)+\frac{\gamma}{2}\quad a.e..\end{aligned}$$

Later,

$$\min_{\beta\in\mathcal{B}_0\times[-d,d],||\beta-\beta_0||\geq\epsilon} M_N(\beta)>m(\beta_0)+\frac{\gamma}{2}\quad a.e.,\quad \forall N>N_1. \tag{7.100}$$

In addition, since $\beta_0\in\mathcal{B}_0\times[-d,d]$ and by Lemma 10 we can write

$$|M_N(\beta_0)-m(\beta_0)|<\frac{\gamma}{4}\quad a.e.,\quad \forall N>N_1,$$

that is,

$$M_N(\beta_0)<\frac{\gamma}{4}+m(\beta_0)\quad a.e.,\quad \forall N>N_1. \tag{7.101}$$

Then, by Lemma 11,

$$\liminf_{N\to\infty}\inf_{|\mu|>d,\phi\in\mathcal{B}_0} M_N(\beta)>m(\beta_0)+\delta\quad \text{a.e.},$$

thus,

$$\sup_{N\geq 0}\left(\inf_{k\geq N}\left(\inf_{|\mu|>d,\phi\in\mathcal{B}_0} M_k(\beta)\right)\right)\geq m(\beta_0)+\delta\quad a.e..$$

Due to the fact that $\inf_{k\geq N}\left(\inf_{|\mu|>d,\phi\in\mathcal{B}_0} M_k(\beta)\right)$ is an increasing succession, $\exists N_2$ such that

$$\inf_{k\geq N}\left(\inf_{|\mu|>d,\phi\in\mathcal{B}_0} M_k(\beta)\right)\geq m(\beta_0)+\delta\quad a.e.,\ \forall N\geq N_2,$$

so,

$$\inf_{|\mu|>d,\phi\in\mathcal{B}_0} M_k(\beta)\geq m(\beta_0)+\delta\quad a.e.,\ \forall k\geq N\quad \forall N\geq N_2.$$

In particular:

$$\inf_{|\mu|>d,\phi\in\mathcal{B}_0} M_N(\beta)\geq m(\beta_0)+\delta>m(\beta_0)+\frac{\gamma}{2}\quad a.e.,\ \forall N\geq N_2. \tag{7.102}$$

Now let us prove that $\hat{\beta}_M\longrightarrow\beta_0$ *a.e.*. Given $\epsilon>0$, let $N_0:=max(N_1,N_2)$. If $N>N_0$, (7.99), (7.100), (7.101) and (7.102) are satisfied.
From (7.100) and (7.102) one obtains

$$\inf_{\beta\in\mathcal{B},||\beta-\beta_0||\geq\epsilon} M_N(\beta)\geq m(\beta_0)+\frac{\gamma}{2}\quad a.e.. \tag{7.103}$$

By the definition of $\hat{\beta}_M$,

$$M_N(\beta)\geq M_N(\hat{\beta}_M),\quad \forall\beta\in\mathcal{B}.$$

In particular,

$$M_N(\beta_0) \geq M_N(\hat{\beta}_M). \tag{7.104}$$

If $||\hat{\beta}_M - \beta_0|| \geq \epsilon$ then by (7.103) it would have that

$$M_N(\hat{\beta}_M) \geq m(\beta_0) + \frac{\gamma}{2}. \tag{7.105}$$

Then, by (7.101), (7.104) and (7.105) it would have that

$$m(\beta_0) + \frac{\gamma}{4} > M_N(\beta_0) \geq M_N(\hat{\beta}_M) \geq m(\beta_0) + \frac{\gamma}{2} \quad a.e.,$$

which is absurd. Therefore, must be that $||\hat{\beta}_M - \beta_0|| < \epsilon$ *a.e.* $\forall N > N_0 = max(N_1, N_2)$, that is, $\hat{\beta}_M \longrightarrow \beta_0$ *a.e.* and the theorem is proven.

□

The next two lemmas establish relationships between the function $m(\beta)$ and the objective function that determines the M-estimation of the parameters of the AR-2D model using the residual functions of the BIP-AR 2D model. These lemmas make it possible to prove the final theorem: Theorem 4.

Lemma 12. *Under the assumptions of Theorem 3, for all* $d > 0$*, there exists* $\delta > 0$ *such that*

$$\liminf_{N\to\infty} \inf_{\beta\in\mathcal{B}_0\times[-d,d]} M_N^b(\beta) \geq m(\beta_0) + \delta \quad a.e..$$

Proof of Lemma 12:

The demonstration of this lemma is similar to Lemma 7.

By Lemma 6 we can find constants $C_1 > 0$ and $C_2 > 0$ such that $\forall d > 0$ and $\forall \tilde{\sigma} > 0$,

$$\sup_{\beta\in\mathcal{B}_0\times[-d,d]} \sup_{0\leq\sigma\leq\tilde{\sigma}} |\varepsilon_{i,j}^b(\beta,\sigma) - Y_{i,j}| \leq C_1\tilde{\sigma} + C_2.$$

Given $\alpha > 0$ and calling $D = C_1(s_0 + \alpha) + C_2$, $\forall \sigma \in [0, s_0 + \alpha]$, we have that

$$\sup_{\beta\in\mathcal{B}_0\times[-d,d]} |\varepsilon_{i,j}^b(\beta,\sigma) - Y_{i,j}| \leq D. \tag{7.106}$$

In addition. we can write the process $\{Y_{i,j}\}$ as $Y_{i,j} = \mu_0 + \varepsilon_{i,j} + v_{i,j}$, where $v_{i,j}$ is a stationary process that depend of $\varepsilon_{k,l}$ when $(k,l) \prec (i,j)$ ($v_{i,j} = \sum_{(k,l)\in T} \phi_{k,l}\varepsilon_{i-k,j-l}$).

Due to the fact that $Y_{i,j}$ is not a white noise and the distribution of $\varepsilon_{i,j}$ is unbounded, we have also that $v_{i,j}$ has unbounded distribution.

Let

$$u_{i,j}(\beta,\sigma) = \mu_0 + v_{i,j} + (\varepsilon_{i,j}^b(\beta,\sigma) - Y_{i,j}), \quad \forall (i,j) \in (W_M \sim T). \tag{7.107}$$

We can write

$$\varepsilon_{i,j}^b(\beta,\sigma) = Y_{i,j} + (\varepsilon_{i,j}^b(\beta,\sigma) - Y_{i,j}) = Y_{i,j} + u_{i,j}(\beta,\sigma) - \mu_0 - v_{i,j} = \varepsilon_{i,j} + u_{i,j}(\beta,\sigma). \tag{7.108}$$

Therefore, by (7.106) and (7.107), $\forall (i,j) \in (W_M \sim T)$ we have

$$\{|v_{i,j}| > D + |\mu_0| + 1\} \subset \left\{ \inf_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq\sigma_Y} |u_{i,j}(\beta,\sigma)| \geq 1 \right\}.$$

Since $v_{i,j}$ is stationary and its distribution is unbounded (all they have the same distribution), we obtain

$$\gamma = P(|v_{i,j}| > D + |\mu_0| + 1) > 0.$$

Let us call $A_{i,j} = \{|v_{i,j}| > D + |\mu_0| + 1\}$.

According to the definition of $m(\beta)$, we have $m(\beta_0) = E_{\beta_0}\left(\rho_2(\varepsilon_{i,j}/s_0)\right)$. As we saw in Lemma 2, for $q \neq 0$ and $u \neq 0$, $E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}}{q}\right)\right) < E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}+u}{q}\right)\right)$ is satisfied.

In particular, if $q = s_0 \neq 0$ we have

$$m(\beta_0) = E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}}{s_0}\right)\right) < E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}+u}{s_0}\right)\right), \quad \forall u \neq 0.$$

This implies that,

$$\inf_{|u|\geq 1} E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}+u}{s_0}\right)\right) > m(\beta_0).$$

Later,

$$(1-\gamma)E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}}{s_0}\right)\right) + \gamma \inf_{|u|\geq 1} E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}+u}{s_0}\right)\right) > (1-\gamma)m(\beta_0) + \gamma m(\beta_0) = m(\beta_0).$$

With arguments similar to those used in Lemma 7, we can find a $\delta > 0$ such that

$$(1-\gamma)E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}}{s_0+\delta}\right)\right) + \gamma \inf_{|u|\geq 1} E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}+u}{s_0+\delta}\right)\right) \geq m(\beta_0) + \delta. \tag{7.109}$$

Let

$$h(u) = E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}+u}{s_0+\delta}\right)\right)$$

and

$$\gamma_N = \frac{1}{N} \sum_{(i,j)\in(W_M \sim T)} I(A_{i,j}).$$

In a similar way to that done in Lemma 7, for all $\beta \in \mathcal{B}_0 \times [-d,d]$, $0 \leq \sigma \leq s_0 + \alpha$,

$$\gamma_N \inf_{|u|\geq 1} h(u) + (1-\gamma_N)h(0) \leq \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} h(u_{i,j}(\beta,\sigma)).$$

Therefore,

$$\gamma_N \inf_{|u|\geq 1} h(u) + (1-\gamma_N)h(0) \leq \inf_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq s_0+\alpha} \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} h(u_{i,j}(\beta,\sigma)). \tag{7.110}$$

Since $\gamma_N \to \gamma$ $a.e.$ (by Law of Large Numbers for Ergodic Processes (Guyon (1995))), and by (7.110) and (7.109), we have that

$$\begin{aligned}\liminf_{N\to\infty} \inf_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq s_0+\alpha} \frac{1}{N} \sum_{(i,j)\in W_M\sim T} h(u_{i,j}(\beta,\sigma)) &\geq \gamma \inf_{|u|\geq 1} h(u) + (1-\gamma)h(0) \\ &\geq m(\beta_0) + \delta \quad a.e.. \end{aligned} \tag{7.111}$$

In the other hand, let

$$\begin{aligned} R_{i,j}(\beta,\sigma) &= \rho_2\left(\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s_0+\delta}\right) - h(u_{i,j}(\beta,\sigma)) \\ &= \rho_2\left(\frac{\varepsilon_{i,j}+u_{i,j}(\beta,\sigma)}{s_0+\delta}\right) - h(u_{i,j}(\beta,\sigma)). \end{aligned}$$

Similarly to Lemma 7, taking the σ-algebra $\mathcal{F}_{i,j}$ generated by the set of random variables $\{R_{k,l} : (k,l) \preccurlyeq (i,j)\}$ under the relationship $\preccurlyeq$ (see Lemma 7), it results that $\{R_{i,j}(\beta,\sigma), \mathcal{F}_{i,j}\}$ is a martingale difference succession. Later, by Large Numbers Law for Martingale Differences (Quang & Van Huan (2010)), one gets that

$$\frac{1}{N} \sum_{(i,j)\in(W_M \sim T)} R_{i,j}(\beta,\sigma) = 0 \quad a.e..$$

By the compactness of $\mathcal{B}_0 \times [-d,d] \times [0, s_0+\alpha]$, we obtain

$$\limsup_{N\to\infty} \sup_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq s_0+\alpha} \left| \frac{1}{N} \sum_{(i,j)\in W_M\sim T} R_{i,j}(\beta,\sigma) \right| = 0 \quad a.e.. \tag{7.112}$$

Finally, by (7.111) and (7.112) we have that for N large enough, almost everywhere

$$\begin{aligned} m(\beta_0)+\delta &\leq \inf_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq s_0+\alpha} \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \left(\rho_2\left(\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s_0+\delta} \right) - R_{i,j}(\beta,\sigma) \right) \\ &\leq \inf_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq s_0+\alpha} \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_2\left(\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s_0+\delta} \right) \\ &+ \sup_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq s_0+\alpha} \left| \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} R_{i,j}(\beta,\sigma) \right|. \end{aligned}$$

Later, taking lower limit:

$$m(\beta_0)+\delta \leq \liminf_{N\to\infty} \inf_{\beta\in\mathcal{B}_0\times[-d,d],\sigma\leq s_0+\alpha} \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_2\left(\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s_0+\delta} \right) \quad a.e.,$$

that is, for $N > N_1$, $\forall \sigma \leq s_0+\alpha$ we have

$$m(\beta_0)+\delta \leq \inf_{\beta\in\mathcal{B}_0\times[-d,d]} \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_2\left(\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s_0+\delta} \right) \quad a.e..$$

By Theorem 2, $\exists N_2$ such that $s^*_N < s_0 + min(\alpha,\delta)$, $\forall N > N_2$. Later, $0 \leq s^*_N \leq s_0+\alpha$ and since ρ_2 satisfies **P1**, we get that

$$\begin{aligned} m(\beta_0)+\delta &\leq \inf_{\beta\in\mathcal{B}_0\times[-d,d]} \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_2\left(\frac{\varepsilon^b_{i,j}(\beta,s^*_N)}{s_0+\delta} \right) \\ &\leq \inf_{\beta\in\mathcal{B}_0\times[-d,d]} \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \rho_2\left(\frac{\varepsilon^b_{i,j}(\beta,s^*_N)}{s^*_N} \right) \quad a.e.. \end{aligned}$$

Therefore,

$$m(\beta_0)+\delta \leq \liminf_{N\to\infty} \inf_{\beta\in\mathcal{B}_0\times[-d,d]} M^b_N(\beta) \quad a.e..$$

Later, the lemma is proven.

□

Lemma 13. *Under the assumptions of Theorem 3, there exist $d>0$ and $\delta>0$ such that*

$$\liminf_{N\to\infty} \inf_{|\mu|>d,\phi\in\mathcal{B}_0} M^b_N(\beta) \geq m(\beta_0)+\delta \quad a.e..$$

Proof of Lemma 13:

The demonstration of this lemma is similar to that of Lemma 8.

In the same way that Lemma 8, there exist positive constants ϵ, D_1 and C such that for any $\alpha > 0$, if $0 < \sigma \leq s_0 + \alpha$, we have that

$$\inf_{\phi \in \mathcal{B}_0} |\varepsilon^b_{i,j}(\beta, \sigma)| \geq \frac{\epsilon}{2}|\mu| - D_1 - (s_0 + \alpha).C - |Y_{i,j}|. \tag{7.113}$$

Since $\sup \rho_2 > m(\beta_0)$, there exists $\delta > 0$ such that $\sup \rho_2 > m(\beta_0) + \delta$. Later, as $\lim_{n\to\infty} \rho_2(|x|) = \sup \rho_2$, there exist k_0 and $\lambda > 1$ such that $\forall |x| \geq k_0$ we obtain

$$\rho_2(x) \geq \lambda(m(\beta_0) + \delta). \tag{7.114}$$

Let k_1 be a constant such that the set $C_{i,j} = \{|Y_{i,j}| \leq k_1 - D_1 - C(s_0 + \alpha)\}$ satisfies $P(C_{i,j}) \geq \frac{1}{\lambda}$.

Let

$$k = max(k_1/s^*_N, k_0)$$

and let d constant such that

$$d > \max(\frac{4ks^*_N}{\epsilon}, |\mu_0|).$$

Then, by the definition of k and (7.113), on $C_{i,j}$ it happens that

$$\begin{aligned}\inf_{\phi \in \mathcal{B}_0, |\mu|>d} |\varepsilon^b_{i,j}(\beta, \sigma)| &\geq \frac{\epsilon}{2}d - D_1 - C(s_0 + \alpha) - k_1 + D_1 + C(s_0 + \alpha) \\ &= \frac{\epsilon}{2}d - k_1 \\ &> k.s^*_N. \end{aligned} \tag{7.115}$$

For all $\beta = (\phi, \mu)$ such that $|\mu| > d$, $\phi \in \mathcal{B}_0$, $0 \leq \sigma \leq s_0 + \alpha$, we have

$$\left|\frac{\varepsilon^b_{i,j}(\beta, \sigma)}{s^*_N}\right| \geq \inf_{|\mu|>d, \phi \in \mathcal{B}_0} \left|\frac{\varepsilon^b_{i,j}(\beta, \sigma)}{s^*_N}\right|.$$

Due to the fact that ρ_2 satisfies **P1**, we have that $\forall |\mu| > d, \phi \in \mathcal{B}_0$:

$$\begin{aligned}\rho_2\left(\frac{\varepsilon^b_{i,j}(\beta, \sigma)}{s^*_N}\right) &\geq \rho_2\left(\inf_{|\mu|>d, \phi \in \mathcal{B}_0} \left|\frac{\varepsilon^b_{i,j}(\beta, \sigma)}{s^*_N}\right|\right) \\ &\geq \rho_2\left(\inf_{|\mu|>d, \phi \in \mathcal{B}_0} \left|\frac{\varepsilon^b_{i,j}(\beta, \sigma)}{s^*_N}\right|\right) I(C_{i,j}) \quad \forall (i,j) \in (W_M \sim T).\end{aligned}$$

Then $\forall |\mu| > d$ and $\phi \in \mathcal{B}_0$, adding over $(W_M \sim T)$:

$$\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} \rho_2\left(\frac{\varepsilon^b_{i,j}(\beta, \sigma)}{s^*_N}\right) \geq \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} \rho_2\left(\inf_{|\mu|>d, \phi \in \mathcal{B}_0} \left|\frac{\varepsilon^b_{i,j}(\beta, \sigma)}{s^*_N}\right|\right) I(C_{i,j}).$$

Later, taking lower limit we get

$$\inf_{|\mu|>d, \phi \in \mathcal{B}_0} \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} \rho_2\left(\frac{\varepsilon^b_{i,j}(\beta, \sigma)}{s^*_N}\right) \geq \frac{1}{N}\sum_{(i,j)\in(W_M\sim T)} \rho_2\left(\inf_{|\mu|>d, \phi \in \mathcal{B}_0} \left|\frac{\varepsilon^b_{i,j}(\beta, \sigma)}{s^*_N}\right|\right) I(C_{i,j}) \tag{7.116}$$

for $\sigma \leq s_0 + \alpha$. In addition, by the equation (7.115) and due to the fact that ρ_2 satisfies **P1** we have

$$\rho_2\left(\inf_{\phi\in\mathcal{B}_0,|\mu|>d}\left|\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s^*_N}\right|\right) \geq \rho_2(k) \quad \forall(i,j)\in(W_M\sim T).$$

Then, adding over $(W_M \sim T)$:

$$\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_2\left(\inf_{\phi\in\mathcal{B}_0,|\mu|>d}\left|\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s^*_N}\right|\right) I(C_{i,j}) > \frac{\rho_2(k)}{N}\sum_{(i,j)\in(W_M\sim T)} I(C_{i,j}),$$

and taking lower limit:

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_2\left(\inf_{\phi\in\mathcal{B}_0,|\mu|>d}\left|\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s^*_N}\right|\right) I(C_{i,j}) \geq \liminf_{N\to\infty}\frac{\rho_2(k)}{N}\sum_{(i,j)\in(W_M\sim T)} I(C_{i,j}). \tag{7.117}$$

Using the definition of k and by (7.114) and $\{I(C_{i,j})\}$ is a stationary and ergodic process with $E(I(C_{i,j})) = P(C_{i,j}) \geq 1/\lambda$, we have by Ergodic Theorem (Guyon (1995)) that

$$\begin{aligned}\liminf_{N\to\infty}\frac{\rho_2(k)}{N}\sum_{(i,j)\in(W_M\sim T)} I(C_{i,j}) &= \rho_2(k)P(C_{i,j})\\ &\geq \lambda(m(\beta_0)+\delta)\frac{1}{\lambda}\\ &= m(\beta_0)+\delta \quad a.e..\end{aligned} \tag{7.118}$$

Then, Eqs. (7.116), (7.117) and (7.118) imply for $0 \leq \sigma \leq s_0 + \alpha$ that

$$\liminf_{N\to\infty}\inf_{|\mu|>d,\phi\in\mathcal{B}_0}\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_2\left(\frac{\varepsilon^b_{i,j}(\beta,\sigma)}{s^*_N}\right) \geq m(\beta_0)+\delta \quad a.e..$$

In addition, as $s^*_N \to s_0$ *a.e.*, for $N \gg 0$, $s^*_N \leq s_0 + \alpha$ is satified and

$$\inf_{|\mu|>d,\phi\in\mathcal{B}_0}\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\rho_2\left(\frac{\varepsilon^b_{i,j}(\beta,s^*_N)}{s^*_N}\right) \geq m(\beta_0)+\delta \quad a.e.,$$

that is,

$$\liminf_{N\to\infty}\inf_{|\mu|>d,\phi\in\mathcal{B}_0} M^b_N(\beta) \geq m(\beta_0)+\delta \quad a.e..$$

Later, the lemma is proven.

□

Finally, we will prove Theorem 4.

Proof of Theorem 4:

Given $d > 0$ and $\delta_1 > 0$ as in Lemma 13, let $\delta_2 > 0$ be as in Lemma 12. Then, there exists $\delta = min(\delta_1, \delta_2) > 0$ such that

$$\liminf_{N\to\infty}\inf_{\beta\in\mathcal{B}} M^b_N(\beta) \geq m(\beta_0)+\delta. \tag{7.119}$$

First let us show that

$$M_N(\hat{\beta}_M) \longrightarrow m(\beta_0) \ a.e.. \tag{7.120}$$

For the continuity of $m(\beta)$ and as $\hat{\beta}_M \longrightarrow \beta_0$ *a.e.*, then $m(\hat{\beta}_M) \longrightarrow m(\beta_0)$ *a.e.* and hence $|m(\hat{\beta}_M) - m(\beta_0)| \longrightarrow 0$ *a.e.*.

In addition, by Lemma 10, $\lim_{N\to\infty} \sup_{\beta\in\mathcal{B}_0\times[-d,d]} |M_N(\beta) - m(\beta)| = 0 \quad a.e.$. Due to the fact that $\hat{\beta}_M \longrightarrow \beta_0$ *a.e.* and $\beta_0 \in \mathcal{B}\times[-d,d]$, we have $\hat{\beta}_M \in \mathcal{B}\times[-d,d]$ for a N large enough. Later, $|M_N(\hat{\beta}_M) - m(\hat{\beta}_M)| \longrightarrow 0$ and (7.120) is proven.

Hence, on account of $|M_N(\hat{\beta}_M) - m(\beta_0)| \leq |M_N(\hat{\beta}_M) - m(\hat{\beta}_M)| + |m(\hat{\beta}_M) - m(\beta_0)|$, we have that $|M_N(\hat{\beta}_M) - m(\beta_0)| \longrightarrow 0 \quad a.e.$.

By (7.119), $\exists N_1$ such that $M_N^b(\beta) \geq \inf_{\beta\in\mathcal{B}} M_N^b(\beta) \geq m(\beta_0) + \delta$, $\forall N > N_1$. In particular, $m(\beta_0) + \delta \leq M_N^b(\hat{\beta}_M^b)$.

By (7.120), $\exists N_2$ such that $|M_N(\hat{\beta}_M) - m(\beta_0)| < \delta$, $\forall N > N_2$. Then $M_N(\hat{\beta}_M) < m(\beta_0) + \delta$, $\forall N > N_2$.

Therefore, $M_N(\hat{\beta}_M) < m(\beta_0) + \delta \leq M_N^b(\hat{\beta}_M^b)$.

Finally, by the definition of $\hat{\beta}_M^*$, one obtains that $\hat{\beta}_M^* = \hat{\beta}_M$, $\forall N > max(N_1, N_2)$ and since $\hat{\beta}_M \longrightarrow \beta_0$ *a.e.*, then we have

$$\hat{\beta}_M^* \longrightarrow \beta_0 \ a.e.$$

as we wanted to demonstrate.

$\square$

The following lemmas allow us to test the asymptotic normality of the $\hat{\beta}_M$ estimator.

Lemma 14. *Under the assumptions of Theorem 5, we obtain that*

$$\frac{1}{\sqrt{N}} \sum_{(i,j)\in(W_M\sim T)} \bigtriangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right) \xrightarrow{D} \mathcal{N}(0, V_0),$$

where

$$V_0 = E\left[\bigtriangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right) . \bigtriangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)^t\right].$$

Proof of Lemma 14:

As we saw in (7.17):

$$V_0 = E\left(\frac{1}{s_0^2}\psi_2^2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right) .E\left[\bigtriangledown(\varepsilon_{i,j}(\beta_0))\bigtriangledown(\varepsilon_{i,j}(\beta_0))^t\right].$$

Since $E\left(\frac{1}{s_0^2}\psi_2^2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right) < \infty$, ψ_2 is bounded and $E\left[\bigtriangledown(\varepsilon_{i,j}(\beta))\bigtriangledown(\varepsilon_{i,j}(\beta))^t\right] < \infty$ (because $E(Y^2) < \infty$) then $V_0 < \infty$.

In the following, we will prove that given a vector column $\boldsymbol{c} \neq 0$ in $\mathbb{R}^{(L+1)^2}$, we have that

$$Z_{i,j} := \boldsymbol{c}' \bigtriangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)$$

together with $\mathcal{F}_{i,j}$ (σ-algebra generated by $\{Z_{s,t} : (s,t)\prec(i,j)\}$) is a stationary martingale difference succession:

1) Due to ψ_2 is bounded, $E(Y^2) < \infty$ and (7.7), we have:

$$\begin{aligned} E(|Z_{i,j}|) &= E\left(\left|\boldsymbol{c}' \bigtriangledown \left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\right|\right) \\ &= E\left(\left|\sum_{k=1}^{(L+1)^2} c_k \bigtriangledown_k \left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\right|\right) \\ &\le E\left(\sum_{k=1}^{(L+1)^2} |c_k|.\left|\bigtriangledown_k \left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\right|\right) \\ &= \sum_{k=1}^{(L+1)^2} |c_k|.E\left(\left|\bigtriangledown_k \left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\right|\right) \\ &= \sum_{k=1}^{(L+1)^2} \frac{|c_k|}{s_0} E\left(\left|\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right|\right).E\left(\left|\bigtriangledown_k\left(\varepsilon_{i,j}(\beta_0)\right)\right|\right) < \infty. \end{aligned}$$

2) It remains to be seen that $E(Z_{i,j}|Z_{s,t}) = 0$ if $(s,t) \prec (i,j)$. Because (7.7) holds, $\bigtriangledown(\varepsilon_{i,j}(\beta_0))$ is a function of $Z_{s,t}$, $\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)$ is independent of $Z_{s,t}$ and by (7.14) we obtain

$$\begin{aligned} E\left(\boldsymbol{c}' \bigtriangledown \left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\middle| Z_{s,t}\right) &= \sum_{k=1}^{(L+1)^2} c_k E\left(\bigtriangledown_k \left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\middle| Z_{s,t}\right) \\ &= \sum_{k=1}^{(L+1)^2} c_k E\left(\frac{1}{s_0}\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right) \bigtriangledown_k\left(\varepsilon_{i,j}(\beta_0)\right)\middle| Z_{s,t}\right) \\ &= \sum_{k=1}^{(L+1)^2} c_k E\left(\frac{1}{s_0}\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\middle| Z_{s,t}\right) \bigtriangledown_k\left(\varepsilon_{i,j}(\beta_0)\right) \\ &= \sum_{k=1}^{(L+1)^2} c_k E\left(\frac{1}{s_0}\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right) \bigtriangledown_k\left(\varepsilon_{i,j}(\beta_0)\right) = 0. \end{aligned}$$

Therefore, $\{Z_{i,j}, \mathcal{F}_{i,j}\}$ is a martingale difference succession. Let us show that it is stationary.
a) By (7.15):

$$\begin{aligned} E(Z_{i,j}) &= E\left(\boldsymbol{c}' \bigtriangledown \left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\right) \\ &= \sum_{k=1}^{(L+1)^2} c_k E\left(\bigtriangledown_k \left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\right) = 0. \end{aligned}$$

b)

$$
\begin{aligned}
Cov(Z_{i,j}, Z_{i+l,j+m}) &= E(Z_{i,j}.Z_{i+l,j+m}) \\
&= E\left(\sum_{k=1}^{(L+1)^2} c_k \bigtriangledown_k \left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right). \sum_{r=1}^{(L+1)^2} c_r \bigtriangledown_r \left(\rho_2\left(\frac{\varepsilon_{i+l,j+m}(\beta_0)}{s_0}\right)\right)\right) \\
&= \sum_{k=1}^{(L+1)^2}\sum_{r=1}^{(L+1)^2} c_k c_r E\left(\bigtriangledown_k \left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right). \bigtriangledown_r \left(\rho_2\left(\frac{\varepsilon_{i+l,j+m}(\beta_0)}{s_0}\right)\right)\right) \\
&= \sum_{k=1}^{(L+1)^2}\sum_{r=1}^{(L+1)^2} \frac{c_k c_r}{s_0^2} E\left(\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\psi_2\left(\frac{\varepsilon_{i+h,j+l}(\beta_0)}{s_0}\right)\bigtriangledown_k\left(\varepsilon_{i,j}(\beta_0)\right)\bigtriangledown_r\left(\varepsilon_{i+l,j+m}(\beta_0)\right)\right) \\
&= \sum_{k=1}^{(L+1)^2} A_k. \qquad (7.121)
\end{aligned}
$$

- If $l \neq 0$ or $m \neq 0 \Rightarrow A_k = 0\ \forall k = 1, ..., (L+1)^2$ by (7.14).
- If $l = 0$ and $m = 0$, (7.121) is equal to

$$
\begin{aligned}
&= E\left(\psi_2^2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\sum_{k=1}^{(L+1)^2}\sum_{r=1}^{(L+1)^2}\frac{c_k c_r}{s_0^2}E(\bigtriangledown_k\left(\varepsilon_{i,j}(\beta_0)\right).\bigtriangledown_r\left(\varepsilon_{i,j}(\beta_0)\right)) \\
&= E\left(\psi_2^2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\left[\sum_{k=1}^{(L+1)^2-1}\sum_{r=1}^{(L+1)^2-1}\frac{c_k c_r}{s_0^2}E(\bigtriangledown_k\left(\varepsilon_{i,j}(\beta_0)\right).\bigtriangledown_r\left(\varepsilon_{i,j}(\beta_0)\right))\right. \\
&+ \left. 2\frac{c_{(L+1)^2}}{s_0^2}\xi_0\sum_{k=1}^{(L+1)^2-1}c_k E(\bigtriangledown_k\left(\varepsilon_{i,j}(\beta_0)\right)) + \frac{c_{(L+1)^2}^2}{s_0^2}\xi_0^2\right] \\
&= E\left(\psi_2^2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\left[\sum_{k=1}^{(L+1)^2-1}\sum_{r=1}^{(L+1)^2-1}\frac{c_k c_r}{s_0^2}(Cov((\bigtriangledown_k\left(\varepsilon_{i,j}(\beta_0)\right), \bigtriangledown_r\left(\varepsilon_{i,j}(\beta_0)\right)) + \mu^2)\right. \\
&+ \left. 2\frac{c_{(L+1)^2}}{s_0^2}\xi_0\sum_{k=1}^{(L+1)^2-1}c_k\mu + \frac{c_{(L+1)^2}^2}{s_0^2}\xi_0^2\right],
\end{aligned}
$$

which is independent of i, j since $Y_{i,j}$ is stationary.

Furthermore

$$
\begin{aligned}
E(|Z_{i,j}|^2) &= E\left(\psi_2^2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\sum_{k=1}^{(L+1)^2}\sum_{r=1}^{(L+1)^2}\frac{c_k c_r}{s_0^2}E(\bigtriangledown_k\left(\varepsilon_{i,j}(\beta_0)\right).\bigtriangledown_r\left(\varepsilon_{i,j}(\beta_0)\right)) \\
&= E\left(\frac{1}{s_0^2}\psi_2^2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\sum_{k=1}^{(L+1)^2}c_k\sum_{r=1}^{(L+1)^2}c_r E(\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right).\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)^t)_{k,r} \\
&= E\left(\frac{1}{s_0^2}\psi_2^2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\boldsymbol{c}^t.E(\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right).\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)^t).\boldsymbol{c} \\
&= \boldsymbol{c}^t.V_0.\boldsymbol{c}.
\end{aligned}
$$

Later, by Central Limit Theorem for Martingales (see Cohen (2016)) we have:

$$\frac{1}{\sqrt{N}} \sum_{(i,j)\in(W_M\sim T)} Z_{i,j} \xrightarrow{D} \mathcal{N}(0, \boldsymbol{c}'.V_0.\boldsymbol{c}),$$

that is,

$$\frac{1}{\sqrt{N}} \sum_{(i,j)\in(W_M\sim T)} \boldsymbol{c}'.\bigtriangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right) \xrightarrow{D} \mathcal{N}(0, \boldsymbol{c}'.V_0.\boldsymbol{c})$$

which, by the Theorem 29.4 of Billingsley (2013), implies

$$\frac{1}{\sqrt{N}} \sum_{(i,j)\in(W_M\sim T)} \bigtriangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right) \xrightarrow{D} \mathcal{N}(0, V_0).$$

□

Lemma 15. *Under the assumptions of Theorem 5, we obtain*

$$\frac{1}{\sqrt{N}}\left\|\sum_{(i,j)\in(W_M\sim T)}\left[\bigtriangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_N^*}\right)\right) - \bigtriangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\right]\right\| \to 0 \quad \textit{in probability.}$$

Proof of Lemma 15:
By (7.7), we can write

$$\begin{aligned}&\frac{1}{\sqrt{N}} \sum_{(i,j)\in(W_M\sim T)}\left[\bigtriangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_N^*}\right)\right) - \bigtriangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\right]\\ &= \frac{1}{\sqrt{N}} \sum_{(i,j)\in(W_M\sim T)}\left[\frac{1}{s_N^*}\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_N^*}\right) - \frac{1}{s_0}\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right].\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right).\end{aligned}$$

We define the functions $A_{N,k}(v)$ for $0 \le v \le 1$ and $1 \le k \le (L+1)^2$ as

$$A_{N,k}(v) = \frac{1}{\sqrt{N}} \sum_{(i,j)\in(W_M\sim T)} \psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{(0.5+v)s_0}\right).\bigtriangledown_k\left(\varepsilon_{i,j}(\beta_0)\right).$$

Since Theorem 2 establishes $\lim_{N\to\infty} s_N^* = s_0$ *a.e.*, then the convergence is in probability too. To prove this lemma is sufficient to prove that $A_{N,k}(v)$ are tight for $1 \le k \le (L+1)^2$.

Using Theorem 12.3 of Billingsley (2013), it is sufficient to prove the following two conditions,

(i) $A_{N,k}(0)$ is tight.

(ii) For any $0 \le v_1 \le v_2$ and any $\lambda > 0$, we have that there exists a constant c_1 such that

$$P(|A_{N,k}(v_2) - A_{N,k}(v_1)| \ge \lambda) \le \frac{c_1}{\lambda^2}(v_2 - v_1)^2.$$

Let us prove (ii). We define for $1 \le k \le (L+1)^2$:

$$G(a, v) := \psi_2\left(\frac{a}{(0.5+v)s_0}\right).$$

Then,

$$
\begin{aligned}
&E((A_{N,k}(v_2) - A_{N,k}(v_1))^2) \qquad (7.122)\\
&= \frac{1}{N} E\left(\left(\sum_{(i,j)\in(W_M\sim T)} (G(\varepsilon_{i,j}, v_2) - G(\varepsilon_{i,j}, v_1)) \bigtriangledown_k (\varepsilon_{i,j}(\beta_0))\right)^2\right)\\
&= \frac{1}{N}\left(\sum_{(i,j)\in(W_M\sim T)} E(B_{i,j}^2 C_{i,j}^2) + \sum_{(i,j)\in(W_M\sim T)} \sum_{(l,m)\in(W_M\sim T)|(l,m)\neq(i,j)} E(B_{i,j}C_{l,m}B_{l,m}C_{i,j})\right)
\end{aligned}
$$

where

$$
B_{i,j} = \psi_2\left(\frac{\varepsilon_{i,j}}{(0.5+v_2)s_0}\right) - \psi_2\left(\frac{\varepsilon_{i,j}}{(0.5+v_1)s_0}\right) \qquad (7.123)
$$

and

$$
C_{i,j} = \bigtriangledown_k (\varepsilon_{i,j}(\beta_0)). \qquad (7.124)
$$

Let $\tilde{\boldsymbol{Y}}_{i,j} = (Y_{i-1,j}, Y_{i-1,j-1}, Y_{i,j-1}, Y_{i-2,j}, Y_{i-2,j-1}, Y_{i-2,j-2}, Y_{i-1,j-2}, Y_{i,j-2}, ...)$ be the vector with the pasts of $Y_{i,j}$. If $l \leq i$ and $m \leq j$ but $(l,m) \neq (i,j)$, it is true that

$$
E(B_{i,j}C_{l,m}B_{l,m}C_{i,j}) = E(E(B_{i,j}C_{l,m}B_{l,m}C_{i,j}|\tilde{\boldsymbol{Y}}_{i,j})).
$$

Due to $B_{l,m}$ depends of $Y_{l,m} \in \tilde{\boldsymbol{Y}}_{i,j}$:

$$
E(E(B_{i,j}C_{l,m}B_{l,m}C_{i,j}|\tilde{\boldsymbol{Y}}_{i,j})) = E(E(B_{i,j}C_{l,m}C_{i,j}|\tilde{\boldsymbol{Y}}_{i,j})B_{l,m}).
$$

As $C_{i,j} \in \tilde{\boldsymbol{Y}}_{i,j}$:

$$
E(E(B_{i,j}C_{l,m}C_{i,j}|\tilde{\boldsymbol{Y}}_{i,j})B_{l,m}) = E(E(B_{i,j}C_{l,m}|\tilde{\boldsymbol{Y}}_{i,j})B_{l,m}C_{i,j}).
$$

Since $C_{l,m} \in \tilde{\boldsymbol{Y}}_{l,m} \subset \tilde{\boldsymbol{Y}}_{i,j}$:

$$
E(E(B_{i,j}C_{l,m}|\tilde{\boldsymbol{Y}}_{i,j})B_{l,m}C_{i,j}) = E(E(B_{i,j}|\tilde{\boldsymbol{Y}}_{i,j})B_{l,m}C_{i,j}C_{l,m}).
$$

In addition, as $B_{i,j}$ depends of $\varepsilon_{i,j}$, it results that $B_{i,j}$ does not depend of $\tilde{\boldsymbol{Y}}_{i,j}$ and then

$$
E(B_{i,j}|\tilde{\boldsymbol{Y}}_{i,j}) = E(B_{i,j}) = 0. \qquad (7.125)
$$

On the other hand, as $B_{i,j}$ is independent of $C_{i,j}$, we have that

$$
E(B_{i,j}^2 C_{i,j}^2) = E(B_{i,j}^2)E(C_{i,j}^2). \qquad (7.126)
$$

Eqs. (7.122), (7.123), (7.124), (7.125) and (7.126) imply that

$$
\begin{aligned}
&E((A_{N,k}(v_2) - A_{N,k}(v_1))^2) \qquad (7.127)\\
&\quad = \frac{1}{N}\left(\sum_{(i,j)\in(W_n\sim T)} E(B_{i,j}^2)E(C_{i,j}^2)\right)\\
&\quad = E\left(\psi_2\left(\frac{\varepsilon_{i,j}}{(0.5+v_2)s_0}\right) - \psi_2\left(\frac{\varepsilon_{i,j}}{(0.5+v_1)s_0}\right)\right)^2 .E(\bigtriangledown_k (\varepsilon_{i,j}(\beta_0)))^2.
\end{aligned}
$$

Let $v_1 < v < v_2$. Then, using Mean Value Theorem we have that

$$
\psi_2\left(\frac{\varepsilon_{i,j}}{(0.5+v_2)s_0}\right) - \psi_2\left(\frac{\varepsilon_{i,j}}{(0.5+v_1)s_0}\right) = \frac{(v_2-v_1)}{s_0(0.5+v)^2}\varepsilon_{i,j}\psi_2'\left(\frac{\varepsilon_{i,j}}{(0.5+v)s_0}\right),
$$

then,

$$E\left(\psi_2\left(\frac{\varepsilon_{i,j}}{(0.5+v_2)s_0}\right)-\psi_2\left(\frac{\varepsilon_{i,j}}{(0.5+v_1)s_0}\right)\right)^2=\frac{(v_2-v_1)^2}{s_0^2(0.5+v)^4}E\left(\varepsilon_{i,j}\psi_2'\left(\frac{\varepsilon_{i,j}}{(0.5+v)s_0}\right)\right)^2.$$

Later, since ψ_2' is bounded, $\varepsilon_{i,j}$ has finite moment second and $s_0>0$, we can conclude that there exists $k_0>0$ such that

$$E\left(\psi_2\left(\frac{\varepsilon_{i,j}}{(0.5+v_2)s_0}\right)-\psi_2\left(\frac{\varepsilon_{i,j}}{(0.5+v_1)s_0}\right)\right)^2\le k_0(v_2-v_1)^2. \tag{7.128}$$

In addition, as $E(Y_{i,j}^2)<\infty$ (since Y has finite moment second) it results that $E(\bigtriangledown_k(\varepsilon_{i,j}(\beta_0)))^2<\infty$ (see equations (7.1) and (7.2) to know the entries of the vector $\bigtriangledown_k(\varepsilon_{i,j}(\beta_0))$).

Then, by (7.127) and (7.128), there exists $c_1>0$ such that

$$E((A_{N,k}(v_2)-A_{N,k}(v_1))^2)\le c_1(v_2-v_1)^2.$$

Thus, (ii) follows by Chebyshev's inequality.

Let us demonstrate (i). Let us calculate $E(A_{N,k}(0)^2)$.

$$\begin{aligned}
&E((A_{N,k}(0))^2)\\
&=\frac{1}{N}E\left(\left(\sum_{(i,j)\in(W_M\sim T)}\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0/2}\right)\bigtriangledown_k(\varepsilon_{i,j}(\beta_0))\right)^2\right)\\
&=\frac{1}{N}\left(\sum_{(i,j)\in(W_M\sim T)}E(\tilde{B}_{i,j}^2\tilde{C}_{i,j}^2)+\sum_{(i,j)\in(W_M\sim T)}\sum_{(l,m)\in(W_M\sim T)|(l,m)\neq(i,j)}E(\tilde{B}_{i,j}\tilde{C}_{l,m}\tilde{B}_{l,m}\tilde{C}_{i,j})\right)
\end{aligned}$$

where

$$\tilde{B}_{i,j}=\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0/2}\right)$$

and

$$\tilde{C}_{i,j}=\bigtriangledown_k(\varepsilon_{i,j}(\beta_0)).$$

Let us see what happens with the second adding:

In the same way like in part (ii), we can take

$$\tilde{\boldsymbol{Y}}_{i,j}=(Y_{i-1,j},Y_{i-1,j-1},Y_{i,j-1},Y_{i-2,j},Y_{i-2,j-1},Y_{i-2,j-2},Y_{i-1,j-2},Y_{i,j-2},...),$$

the vector with the past of $Y_{i,j}$. If $l\le i$ and $m\le j$ but $(l,m)\neq(i,j)$,

$$E(\tilde{B}_{i,j}\tilde{C}_{l,m}\tilde{B}_{l,m}\tilde{C}_{i,j})=E(E(\tilde{B}_{i,j}|\tilde{\boldsymbol{Y}}_{i,j})\tilde{B}_{l,m}\tilde{C}_{i,j}\tilde{C}_{l,m})$$

and as $\tilde{B}_{i,j}$ depends of $\varepsilon_{i,j}$, it results that $\tilde{B}_{i,j}$ does not depend of $\tilde{\boldsymbol{Y}}_{i,j}$, then

$$E(\tilde{B}_{i,j}|\tilde{\boldsymbol{Y}}_{i,j})=E(\tilde{B}_{i,j}).$$

In addition, as ψ_2 is odd and the distribution of $\varepsilon_{i,j}$ is symmetric, we have that $E(\tilde{B}_{i,j})=0$. Later, the second summing is zero.

Now let us see the first summing:

Due to $\tilde{B}_{i,j}$ is independent of $\tilde{C}_{i,j}$,

$$E(\tilde{B}^2_{i,j}\tilde{C}^2_{i,j}) = E(\tilde{B}^2_{i,j})E(\tilde{C}^2_{i,j}). \tag{7.129}$$

Since the function ψ_2 is bounded, there exists a constant $M > 0$ such that $E(\tilde{B}^2_{i,j}) < M$, $\forall(i,j)$. In addition, from the expression of $\bigtriangledown_k (\varepsilon_{i,j}(\beta_0))^2$, we have that $E(\tilde{C}^2_{i,j}) \leq \kappa$ where κ is such that $\kappa = max(\sigma^2_Y + \mu^2_0, \xi^2_0) = max(\sigma^2_Y + \mu^2_0, (-1 + \sum_{(k,l)\in T} \phi^0_{k,l})^2)$. Later, by the equation (7.129) we obtain that

$$E(\tilde{B}^2_{i,j}\tilde{C}^2_{i,j}) < M.\kappa.$$

Thus,

$$\begin{aligned}
E(A_{N,k}(0)^2) &= \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} E(\tilde{B}^2_{i,j}\tilde{C}^2_{i,j}) \\
&< \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} M.\kappa \\
&= M.\kappa.
\end{aligned}$$

Let $\epsilon > 0$ and $\delta = \left(\frac{M.\kappa}{\epsilon}\right)^{1/2}$. By Chebyshev we get that

$$\begin{aligned}
P(|A_{N,k}(0)| > \delta) &\leq \frac{1}{\delta^2} E(A^2_{N,k}(0)) \\
&= \frac{E(A^2_{N,k}(0))}{M.\kappa/\epsilon} \\
&< \frac{M\kappa}{M.\kappa}\epsilon \\
&= \epsilon.
\end{aligned}$$

Therefore, $A_{N,k}(0)$ is tight and the lemma is proven.

□

Lemma 16. *Under the assumptions of Theorem 5, we obtain that $\forall d > 0$ that*

i)

$$\lim_{N\to\infty} \sup_{\beta\in\mathcal{B}_0\times[-d,d]} \left\| \frac{1}{N} \sum_{(i,j)\in(W_M\sim T)} \bigtriangledown^2 \left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s^*_N}\right)\right) - E\left[\bigtriangledown^2\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right)\right] \right\| = 0 \quad a.e.,$$

where $||A||$ denotes the norm l_2 of the A matrix.

ii)

$$E\left[\bigtriangledown^2\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\right] = \frac{1}{s^2_0} E\left(\psi'_2\left(\frac{\varepsilon_{i,j}}{s_0}\right)\right) .E\left(\bigtriangledown(\varepsilon_{i,j}(\beta_0)). \bigtriangledown (\varepsilon_{i,j}(\beta_0))^t\right).$$

Proof of Lemma 16:
The proof of (i) is similar to that of Lemma 10:

By the Theorem of Dominated Convergence, as $\psi'_2 = \rho''_2$ is continuous and bounded and $\varepsilon_{i,j}(\beta)$ is continuous, we have

$$\tilde{M}(\beta, v) = E_{\beta_0}\left(\bigtriangledown^2\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{v}\right)\right)\right)$$

is a continuous function respect to the two variables.

Then, given $\epsilon > 0$ and $\beta \in \mathcal{B}_0 \times [-d, d]$, by the continuity of $\tilde{M}(\beta, v)$ in $v = s_0$ we have that there exists $0 < \delta(\beta) < s_0$ such that if $|v - s_0| < \delta(\beta)$ then $|\tilde{M}(\beta, v) - \tilde{M}(\beta, s_0)| < \epsilon/2$ for each $\beta \in \mathcal{B}_0 \times [-d, d]$. By compactness of $\mathcal{B}_0 \times [-d, d]$ we obtain that $\exists \delta > 0$ such that $|\tilde{M}(\beta, v) - \tilde{M}(\beta, s_0)| \leq \epsilon/2$, $\forall \beta \in \mathcal{B}_0 \times [-d, d]$ and $\forall v \in [s_0 - \delta, s_0 + \delta]$, and so

$$\sup_{\beta \in \mathcal{B}_0 \times [-d,d], v \in [s_0-\delta, s_0+\delta]} |\tilde{M}(\beta, v) - \tilde{M}(\beta, s_0)| \leq \epsilon/2. \tag{7.130}$$

We consider the function

$$f(y, \beta, v) = \sum_{k=1}^{(L+1)^2} \sum_{l=1}^{(L+1)^2} \left| \triangledown^2_{k,l} \left(\rho_2 \left(\frac{\Phi(B_1, B_2)(y - \mu)}{v} \right) \right) - E_{\beta_0} \left(\triangledown^2_{k,l} \left(\rho_2 \left(\frac{\Phi(B_1, B_2)(y - \mu)}{v} \right) \right) \right) \right|.$$

This function is continuous, defined in $\mathbb{R} \times C_0$ with $C_0 = \{(\beta, v) : \beta \in \mathcal{B}_0 \times [-d, d], v \in [s_0 - \delta, s_0 + \delta]\}$ compact. As $\{Y_{i,j}\}$ is an ergodic process, $E_{\beta_0}(f(Y, \beta, v)) = 0$ and $\sup_{(\beta,v) \in C_0} |f(Y, \beta, v)| \leq K$ with K constant, by Lemma 3 of Muler & Yohai (2002) we have that

$$\lim_{N \longrightarrow \infty} \sup_{(\beta,v) \in C_0} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} f(Y_{i,j}, \beta, v) = 0 \quad a.e..$$

Later,

$$\begin{aligned}
&\lim_{N \longrightarrow \infty} \sup_{(\beta,v) \in C_0} \left\| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{v} \right) \right) - E_{\beta_0} \left(\triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{v} \right) \right) \right) \right\| \\
&\leq \lim_{N \longrightarrow \infty} \sup_{(\beta,v) \in C_0} \left(\sum_{k=1}^{(L+1)^2} \sum_{l=1}^{(L+1)^2} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \left| \triangledown^2_{k,l} \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{v} \right) \right) - E_{\beta_0} \left(\triangledown^2_{k,l} \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{v} \right) \right) \right) \right|^2 \right)^{1/2} \\
&= \lim_{N \longrightarrow \infty} \sup_{(\beta,v) \in C_0} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} f(Y_{i,j}, \beta, v) = 0 \quad a.e.
\end{aligned}$$

Since Theorem 2 establishes that $\lim_{N \to \infty} s^*_N = s_0$ *a.e.*, then, with probability 1, there exists N_0 such that $\forall N > N_0$, $s^*_N \in [s_0 - \delta, s_0 + \delta]$ and

$$\sup_{\beta \in \mathcal{B}_0 \times [-d,d]} \left\| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{v} \right) \right) - E_{\beta_0} \left(\triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{v} \right) \right) \right) \right\| < \epsilon/2 \quad a.e.. \tag{7.131}$$

Eqs. (7.130) and (7.131) imply that, $\forall N > N_0$,

$$\begin{aligned}
&\left\| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s^*_N} \right) \right) - E_{\beta_0} \left(\triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \right) \right) \right\| \\
&\leq \left\| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s^*_N} \right) \right) - E_{\beta_0} \left(\triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s^*_N} \right) \right) \right) \right\| + \\
&+ \left\| E_{\beta_0} \left(\triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s^*_N} \right) \right) \right) - E_{\beta_0} \left(\triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \right) \right) \right\| \\
&< \epsilon/2 + \epsilon/2, \quad \forall \beta \in \mathcal{B}_0 \times [-d, d] \quad a.e..
\end{aligned}$$

Then, taking supreme one gets

$$\sup_{\beta\in\mathcal{B}_0\times[-d,d]}\left\|\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\triangledown^2\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*}\right)\right)-E\left[\triangledown^2\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_0}\right)\right)\right]\right\|<\epsilon\ a.e.,\quad \forall N>N_0$$

and the part (i) of the lemma is proven.

Now we will demonstrate (ii). By (7.10) and from the fact that $\triangledown(\varepsilon_{i,j}(\beta_0))$ and $\triangledown^2(\varepsilon_{i,j}(\beta_0))$ depend of $\tilde{\boldsymbol{Y}}_{i,j}$, we obtain that

$$\begin{aligned}E\left(\triangledown^2\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\right)&=\frac{1}{s_0^2}E\left(\psi_2'\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right).E\left(\triangledown\left(\varepsilon_{i,j}(\beta_0)\right)\triangledown\left(\varepsilon_{i,j}(\beta_0)\right)^t\right)\\&+\frac{1}{s_0}E\left(\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right).E\left(\triangledown^2\left(\varepsilon_{i,j}(\beta_0)\right)\right).\end{aligned}$$

By (7.14), we have that $E(\psi_2(\varepsilon_{i,j}(\beta_0)/s_0))=0$. Therefore,

$$E\left(\triangledown^2\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\right)=\frac{1}{s_0^2}E\left(\psi_2'\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right).E\left(\triangledown\left(\varepsilon_{i,j}(\beta_0)\right)\triangledown\left(\varepsilon_{i,j}(\beta_0)\right)^t\right)\tag{7.132}$$

and (ii) is proven.

□

Proof of Theorem 5:

By the definition of $\hat{\beta}_M$ we get that

$$M_N(\hat{\beta}_M)\leq M_N(\beta),\quad \forall\beta\in\mathcal{B}$$

Thus, $\hat{\beta}_M$ satisfies:

$$\sum_{(i,j)\in(W_M\sim T)}\triangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\hat{\beta}_M)}{s_N^*}\right)\right)=0.$$

Due to the fact that β_0 and $\hat{\beta}_M$ are in $\mathcal{B}$, then by the Mean Value Theorem we obtain that

$$\begin{aligned}0&=\sum_{(i,j)\in(W_M\sim T)}\triangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\hat{\beta}_M)}{s_N^*}\right)\right)\\&=\sum_{(i,j)\in(W_M\sim T)}\triangledown\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_N^*}\right)\right)+\sum_{(i,j)\in(W_M\sim T)}\triangledown^2\left(\rho_2\left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_N^*}\right)\right)(\hat{\beta}_M-\beta_0)\end{aligned}\tag{7.133}$$

where $\tilde{\beta}$ is an intermediate point between β_0 and $\hat{\beta}_M$, that is, $\tilde{\beta}=\beta_0+\theta(\hat{\beta}_M-\beta_0)$ with $0<\theta<1$.

By Theorem 3, it is has that $\hat{\beta}_M\to\beta_0$ $a.e.$ and, hence, $\tilde{\beta}\to\beta_0\quad a.e.$.
Let $d>0$ be such that $d>|\mu_0|$. Later, with probability 1, there exits N_0 such that $\hat{\beta}_M\in\mathcal{B}_0\times[-d,d]$, $\forall N\geq N_0$.
Let

$$A_N=\frac{1}{N}\sum_{(i,j)\in(W_M\sim T)}\triangledown^2\left(\rho_2\left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_N^*}\right)\right).$$

Let us prove that

$$\lim_{N\to\infty}A_N=E\left[\triangledown^2\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)\right]\quad a.e..\tag{7.134}$$

Given $\epsilon > 0$. As $\tilde{\beta} \rightarrow \beta_0$ *a.e.*, one gets that $\tilde{\beta} \in \mathcal{B} \times [-d, d]$ for an N large enough. Then, by Lemma 16-(i), $\exists N_0$ such that

$$\left\| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_N^*} \right) \right) - E \left[\triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_0} \right) \right) \right] \right\| < \epsilon/2 \quad a.e. \quad \forall N > N_0. \tag{7.135}$$

Due to the function $\rho_2'' = \psi_2'$ is continuous and bounded and the residual functions $\varepsilon_{i,j}(\beta)$'s are continuous, by Lebesgue's Dominated Convergence Theorem, it turns out that $E\left[\triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \right) \right]$ is a continuous function. Then, since $\tilde{\beta} \rightarrow \beta_0$ *a.e.*, we can obtain N_1 such that $\forall N > N_1$:

$$\left\| E \left[\triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \right] - E \left[\triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_0} \right) \right) \right] \right\| < \epsilon/2 \quad a.e.. \tag{7.136}$$

Hence, as

$$\begin{aligned} &\left\| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_N^*} \right) \right) - E \left[\triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \right] \right\| \leq \\ &\left\| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_N^*} \right) \right) - E \left[\triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_0} \right) \right) \right] \right\| \\ &+ \left\| E \left[\triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_0} \right) \right) \right] - E \left[\triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \right] \right\|, \end{aligned} \tag{7.137}$$

Eqs. (7.135) and (7.136), imply for all $N > max(N_0, N_1)$ that (7.137) is less than ϵ. Then, (7.134) is satisfied.

By Lemma 16-(ii), $A := E\left[\triangledown^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \right]$ is not singular and since $A_N \rightarrow A \quad a.e.$ then, for an N large enough, it is obtained that A_N is not singular.

Otherwise, dividing (7.133) by N and calling

$$C_N := \frac{1}{\sqrt{N}} \sum_{(i,j) \in (W_M \sim T)} \triangledown \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_N^*} \right) \right),$$

it is obtained that the equation (7.133) is equivalent to:

$$\begin{aligned} \frac{\sqrt{N}}{N} C_N + A_N(\hat{\beta}_M - \beta_0) &= 0 \\ \frac{1}{\sqrt{N}} C_N + A_N(\hat{\beta}_M - \beta_0) &= 0 \\ C_N + \sqrt{N} A_N(\hat{\beta}_M - \beta_0) &= 0 \\ A_N \sqrt{N}(\hat{\beta}_M - \beta_0) &= -C_N. \end{aligned}$$

Then, for N large enough such that A_N is not singular, the previous equality becomes

$$\sqrt{N}(\hat{\beta}_M - \beta_0) = -A_N^{-1} C_N.$$

In order to prove the theorem we will see that $A_N^{-1} C_N \xrightarrow{D} \mathcal{N}(0, D)$.

First let us show that $C_N \xrightarrow{D} \mathcal{N}(0, V_0)$ with V_0 as in (7.16).

Let $Z_N := \frac{1}{\sqrt{N}} \sum_{(i,j)\in(W_M \sim T)} \bigtriangledown \left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)$. Then, by Lemma 15, it turns out

$$||C_N - Z_N|| \to 0 \quad \text{in probability,}$$

that is,

$$C_N - Z_N \to 0 \quad \text{in probability.}$$

In addition, Lemma 14 establishes that $Z_N \xrightarrow{D} Z$ with $Z \sim \mathcal{N}(0, V_0)$. Then, by Slutsky,

$$Z_N + (C_N - Z_N) \xrightarrow{D} Z + 0 = Z,$$

that is,

$$C_N \xrightarrow{D} Z \text{ with } Z \sim \mathcal{N}(0, V_0). \tag{7.138}$$

Now let us prove that $A_N^{-1} C_N \xrightarrow{D} \mathcal{N}(0, A^{-1} V_0 (A^{-1})^t)$.

Let $h : \mathbb{R}^{(L+1)^2} \to \mathbb{R}^{(L+1)^2}$ be a function defined by $h(X) = A^{-1}X$. h is measurable and continuous function in $\mathbb{R}^{(L+1)^2}$ and by (7.138) and Theorem 29.2 of Billingsley (2013) we obtain that

$$h(C_N) = A^{-1} C_N \xrightarrow{D} h(Z) = A^{-1} Z \text{ with } A^{-1} Z \sim \mathcal{N}(0, A^{-1} V_0 (A^{-1})^t).$$

On the other hand, as C_N is bounded (ρ_2' is bounded) and $A_N^{-1} \to A^{-1}$ $a.e.$ (invert a matrix is a continuous function), one gets that

$$||A^{-1} C_N - A_N^{-1} C_N|| \to 0 \text{ in probability.}$$

Then,

$$A^{-1} C_N - A_N^{-1} C_N \to 0 \text{ en probability.} \tag{7.139}$$

Later, since $A_N^{-1} C_N = (A_N^{-1} C_N - A^{-1} C_N) + A^{-1} C_N$, by Theorem 5.1.5 of Lehmann (2004) and using (7.139) and (7.138) we have that

$$A_N^{-1} C_N \xrightarrow{D} \mathcal{N}(0, A^{-1} V_0 (A^{-1})^t)$$

Therefore,

$$\sqrt{N}(\hat{\beta}_M - \beta_0) \xrightarrow{D} \mathcal{N}(0, A^{-1} V_0 (A^{-1})^t).$$

It remains to be seen that $D = A^{-1} V_0 (A^{-1})^t$.

By (7.17),

$$V_0 = \frac{1}{s_0^2} E\left[\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)^2\right] . E\left[\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right) . \bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)^t\right];$$

and by (7.10), we have that

$$\bigtriangledown^2\left(\rho_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right) = \frac{1}{s_0^2}\psi_2'\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right) . \bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right) + \frac{1}{s_0}\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right) . \bigtriangledown^2\left(\varepsilon_{i,j}(\beta_0)\right).$$

Then,

$$\begin{aligned} A &= E\left[\frac{1}{s_0^2}\psi_2'\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right) . \bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right) . \bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)^t + \frac{1}{s_0}\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right) . \bigtriangledown^2\left(\varepsilon_{i,j}(\beta_0)\right)\right] \\ &= E\left(\frac{1}{s_0^2}\psi_2'\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right) . E\left[\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right) \bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)^t\right] + E\left(\frac{1}{s_0}\psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right) . E\left[\bigtriangledown^2\left(\varepsilon_{i,j}(\beta_0)\right)\right]. \end{aligned}$$

By (7.14), the last equation becomes

$$A = E\left(\frac{1}{s_0^2}\psi_2'\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right).E\left[\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)^t\right].$$

Later,

$$A^{-1} = \frac{1}{E\left(\frac{1}{s_0^2}\psi_2'\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)}.E\left[\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)^t\right]^{-1}.$$

In addition, since $E\left[\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)^t\right]$ is symmetric, then $E\left[\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)^t\right]^{-1}$ is symmetric. Hence,

$$\begin{aligned} A^{-1}V_0(A^{-1})' &= A^{-1}V_0A^{-1} = \frac{E\left(\frac{1}{s_0^2}\psi_2^2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)}{E\left(\frac{1}{s_0^2}\psi_2'\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)^2}.E\left[\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)'\right]^{-1} \\ &= \frac{s_0^2E\left(\psi_2^2\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)}{E\left(\psi_2'\left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0}\right)\right)^2}.E\left[\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)'\right]^{-1}. \end{aligned}$$

Finally, by (7.18), one has that

$$E\left[\bigtriangledown\left(\varepsilon_{i,j}(\beta)\right)\bigtriangledown\left(\varepsilon_{i,j}(\beta)\right)^t\right] = \begin{pmatrix} \sigma_\varepsilon^2 C & 0_{L\times 1} \\ 0_{1\times L} & \xi^2 \end{pmatrix}.$$

Then,

$$E\left[\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)\bigtriangledown\left(\varepsilon_{i,j}(\beta_0)\right)^t\right]^{-1} = \begin{pmatrix} \sigma_\varepsilon^{-2} C^{-1} & 0 \\ 0 & \xi_0^{-2} \end{pmatrix}$$

and, therefore, the theorem is proven.

□

Proof of Theorem 6:

By the demonstrate of Theorem 4, $\exists N_0$ such that if $N > N_0$ then $\hat{\beta}_M^* = \hat{\beta}_M$.

Let $Z_N = \sqrt{N}(\hat{\beta}_M^* - \beta_0)$ with cumulative distribution function F_{Z_N} and $Y_N = \sqrt{N}(\hat{\beta}_M - \beta_0)$ with cumulative distribution function F_{Y_N}.

Then, since $Z_N = Y_N$ for $N > N_0$, result that $F_{Z_N} = F_{Y_N}$ for $N > N_0$.

By Theorem 5, $\exists N_1$ such that if $N > N_1$, $|F_{Y_N}(x) - F_Z(x)| < \varepsilon\ \forall x$ where F_Z is the cumulative distribution function corresponding to $\mathcal{N}(0, D)$.

Then, for $N > max(N_0, N_1)$, we obtain that $|F_{Z_N}(x) - F(x)| < \varepsilon\ \forall x$. Therefore, $Z_N \xrightarrow{D} Z$ and the theorem is proven.

□

References

Alata, O., & Olivier, C. (2003). Choice of a 2-d causal autoregressive texture model using information criteria. *Pattern Recognition Letters*, *24*, 1191–1201.

Allende, H., Galbiati, J., & Vallejos, R. (1998). Digital image restoration using autoregressive time series type models. *Bulletin European Spatial Agency*, *434*, 53–59.

Allende, H., Galbiati, J., & Vallejos, R. (2001). Robust image modeling on image processing. *Pattern Recognition Letters*, *22*, 1219–1231.

Basu, S., & Reinsel, G. (1993). Properties of the spatial unilateral first-order arma model. *Advances in applied Probability*, *25*, 631–648.

Billingsley, P. (2013). *Convergence of probability measures*. John Wiley & Sons.

Britos, G. (2019). *Estimación Robusta en Modelos ARMA Bidimensionales. Aplicación al Procesamiento de Imágenes Digitales*. Tesis doctoral Universidad Nacional de Córdoba. Facultad de Matemática, Astronomía, Física y Computación.

Britos, G., & Ojeda, S. (2018). Robust estimation for spatial autoregressive processes based on bounded innovation propagation representations. *Computational Statistics*, (pp. 1–21).

Bustos, O., Ojeda, S., & Vallejos, R. (2009). Spatial arma models and its applications to image filtering. *Brazilian Journal of Probability and Statistics*, *23*, 141–165.

Bustos, O. H., & Yohai, V. J. (1986). Robust estimates for arma models. *Journal of the American Statistical Association*, *81*, 155–168.

Cohen, G. (2016). A clt for multi-dimensional martingale differences in a lexicographic order. *Stochastic Processes and their Applications*, *126*, 1503–1510.

Dormann, C., McPherson, J., Araújo, M., Bivand, R., Bolliger, J., Carl, G., Davies, R., Hirzel, A., Jetz, W., Kissling, W. et al. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, *30*, 609–628.

Guyon, X. (1995). *Random fields on a network: modeling, statistics, and applications*. Springer Science & Business Media.

Huber, P. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, *35*, 73–101.

Kashyap, R., & Eom, K. (1988). Robust image modeling techniques with an image restoration application. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *36*, 1313–1325.

Latha, M., Poojith, A., Reddy, B., & Kumar, G. (2014). Image processing in agriculture. *International Journal of Innovative Research In Electrical, Electronics, Instrumentation And Control Engineering*, *2*.

Lehmann, E. (2004). *Elements of large-sample theory*. Springer Science & Business Media.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, *11*, 431–441.

Muler, N., Peña, D., & Yohai, V. (2009). Robust estimation for arma models. *The Annals of Statistics*, *37*, 816–840.

Muler, N., & Yohai, V. (2002). Robust estimates for arch processes. *J. Time Ser. Anal.*, *23*, 341–375.

Ojeda, S., Vallejos, R., & Bustos, O. (2010). A new image segmentation algorithm with applications to image inpainting. *Computational Statistics and Data Analysis*, *54*, 2082–2093.

Ojeda, S., Vallejos, R., & Lucini, M. (2002). Performance of robust ra estimator for bidimensional autoregressive models. *Journal of Statistical Computation and Simulation*, *72*, 47–62.

Quang, N., & Van Huan, N. (2010). A hájek-rényi-type maximal inequality and strong laws of large numbers for multidimensional arrays. *Journal of Inequalities and Applications*, *2010*, 569759.

Quintana, C., Ojeda, S., Tirao, G., & Valente, M. (2011). Mammography image detection processing for automatic micro-calcification recognition. *Chilean Journal of Statistics*, *2*.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. URL: https://www.R-project.org/.

Sahu, O., Anand, V., Kanhangad, V., & Pachori, R. (2015). Classification of magnetic resonance brain images using bidimensional empirical mode decomposition and autoregressive model. *Biomedical Engineering Letters*, *5*, 311–320.

Sain, S., & Cressie, N. (2007). A spatial model for multivariate lattice data. *Journal of Econometrics*, *140*, 226–259.

Smith, M., Nichols, S., Henkelman, R., & Wood, M. (1986). Application of autoregressive moving average parametric modeling in magnetic resonance image reconstruction. *IEEE Transactions on Medical Imaging*, *5*, 132–139.

Vallejos, R., & Mardesic, T. (2004). A recursive algorithm to restore images based on robust estimation of nshp autoregressive models. *Journal of Computational and Graphical Statistics*, *13*, 674–682.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, (pp. 434–449).

Yohai, V. (1985). *High breakdown-point and high efficiency robust estimates for regression*. Technical Report 66 Department of Statistics, University of Washington.

Zielinski, J., Bouaynaya, N., & Schonfeld, D. (2010). Two-dimensional arma modeling for breast cancer detection and classification. *Signal Processing and Communications (SPCOM)*, (pp. 1–4).