

Estimación Robusta en Modelos ARMA Bidimensionales. Aplicación al Procesamiento de Imágenes Digitales

por Grisel Maribel Britos

Presentado ante la Facultad de Matemática, Astronomía, Física y Computación como parte
de los requerimientos para la obtención del grado de Doctora en Matemática de la

UNIVERSIDAD NACIONAL DE CÓRDOBA

Marzo de 2019

©FAMAF-UNC 2019

Directora: Dra. Silvia María Ojeda



Titulo por Grisel Maribel Britos se distribuye bajo una Licencia Creative Commons
Atribución-NoComercial-CompartirIgual 4.0 Internacional

Resumen

Este trabajo se focalizó en el problema de la estimación robusta de los parámetros en modelos autorregresivos bidimensionales con contaminación. Se propone un nuevo método de estimación robusta de los parámetros de estos modelos, denominado BMM 2D, que se basa en la representación de un proceso autoregresivo bidimensional con un modelo auxiliar, como una estrategia para atenuar el efecto de la contaminación.

Hasta el momento y desde la definición de los modelos autorregresivos bidimensionales, tratados inicialmente por Whittle (1954), diversos autores han desarrollado propuestas para abordar el problema de estimación de los parámetros, generando alternativas de estimación cuando la contaminación del modelo es aditiva o innovativa. Estas propuestas, si bien han mostrado buen desempeño en aplicaciones, no vienen acompañadas (en general) por estudios que den cuenta de sus propiedades teóricas, tales como consistencia y normalidad asintótica. En esta tesis, se presentó un nuevo estimador para estimar los parámetros del modelo en condiciones más generales de contaminación y se demostró la consistencia y la normalidad asintótica del estimador.

El trabajo incluyó un análisis comparativo entre el método propuesto, los estimadores robustos existentes hasta el momento y el estimador de mínimos cuadrados, a través de un estudio de simulación de Monte Carlo, variando el tamaño de la ventana de observación del proceso, y el tipo y nivel de contaminación. Los resultados evidenciaron que el nuevo estimador constituye una propuesta competitiva, tanto en exactitud como en precisión, con relación a otros estimadores clásicos y robustos utilizados hasta ahora.

Finalmente, se presentó una aplicación al filtrado de imágenes, que ilustra cómo funciona el estimador BMM 2D en situaciones prácticas.

El procedimiento intenta generalizar a dos dimensiones la iniciativa presentada por Muler et al (2009), desarrollada para modelos ARMA de series de tiempo.

Palabras claves: Modelos AR-2D; Estimadores Robustos; Procesamiento de Imágenes; Consistencia; Normalidad Asintótica

Abstract

This work focused on the problem of robust estimation of parameters in two-dimensional autoregressive models with contamination. A new method of robust estimation of the parameters of these models is proposed, called BMM 2D, which is based on the representation of a two-dimensional autoregressive process with an auxiliary model, as a strategy to attenuate the effect of contamination.

Up to the moment and from the definition of the two-dimensional autoregressive models, initially treated by Whittle (1954), diverse authors have developed proposals to approach the problem of estimation of parameters, generating estimation alternatives when the contamination of the model is additive or innovative. These proposals, although they have shown good performance in applications, are not accompanied (in general) by studies that account for their theoretical properties, such as consistency and asymptotic normality. In this thesis, a new estimator to estimate the parameters of the model in more general conditions of contamination was presented and the consistency and asymptotic normality of the estimator were proved.

The work included a comparative analysis between the proposed method, the existing robust estimators and the least squares estimator, through a Monte Carlo simulation study, varying the size of the process observation window, and the type and level of contamination. The results showed that the new estimator is a competitive proposal, both in accuracy and precision, in relation to other classical and robust estimators used so far.

Finally, an application to image filtering was presented, which illustrates how the BMM 2D estimator works in practical situations.

The procedure attempts to generalize to two dimensions the initiative presented by Muler et al (2009), developed for ARMA models of time series.

Key words: AR-2D Models; Robust Estimators; Image Processing; Consistency; Asymptotic Normality

Agradecimientos

Quiero agradecer a mi directora, Silvia Ojeda, que desde el primer momento me recibió con los brazos abiertos, por enseñarme tanto, pero por sobre todo por la calidez, la paciencia y la confianza que me ha demostrado todos estos años.

A mi familia por apoyarme a lo largo de mi vida. A mis padres por el sacrificio diario y el amor que me han brindado en el día a día. A mi esposo, Martín Ocampo, por el amor y el apoyo que me ha brindado siempre; por acompañarme en este camino con tanta alegría y dulzura, y por haber sido un pilar muy importante para mí. A mis hermanas por estar presentes físicamente y a la distancia toda mi vida.

A mis amigos del secundario y los amigos que encontré en la facultad, por darme fuerzas en los momentos que necesitaba y por la comprensión y apoyo que tuvieron conmigo siempre. En especial a mis compañeros y amigos de oficina, por hacer valer las tantas horas compartidas.

A la FaMAF y al CIEM por haberme brindado el lugar de trabajo. A CONICET y SeCyT por el apoyo económico. En especial, a mi comisión asesora de doctorado y al tribunal de esta tesis.

Al Grupo de Probabilidad y Estadística que me acogieron y siempre estuvieron dispuestos a ayudarme en lo que necesitara. En especial a Oscar Bustos por ser fuente inagotable de consultas, brindando su tiempo y conocimiento tan generosamente. A todos los docentes de la FaMAF por la formación y por el impacto que tuvieron en mi camino.

Índice general

Resumen	I
Abstract	III
Agradecimientos	v
Introducción	IX
1. Procesos Aleatorios Bidimensionales	1
1.1. Preliminares	1
1.2. Ergodicidad	3
1.3. Orden en \mathbb{Z}^2	4
1.4. Martingalas	7
1.5. Procesos AR Bidimensionales Reales	9
1.5.1. Representaciones AR 2D Fuertemente Causales	10
1.6. Procesamiento de Imágenes	11
1.6.1. Interpretación matemática de una imagen	11
1.6.2. Modelos para imágenes	12
1.6.3. Contaminación en procesos AR-2D	13
1.6.4. Expresividad del modelo AR-2D	14
2. Estimadores BMM para modelos AR Bidimensionales	17
2.1. Modelo BIP-AR 2D	19
2.2. Estimador de Mínimos Cuadrados en el modelo AR 2D	20
2.3. Estimadores M, GM y RA en el modelo AR 2D	21
2.4. Estimador BMM 2D para procesos AR-2D	22

3. Comportamiento del Estimador BMM	27
3.1. El modelo	27
3.2. Estimación de los parámetros	28
3.3. Experimentos	29
3.3.1. Evaluación del Tiempo Computacional	34
3.4. Aplicación a imágenes reales	34
4. Propiedades Teóricas del Estimador BMM Bidimensional	43
4.1. Preliminares	43
4.2. Resultados que se desprenden del estimador BMM 2D	46
4.3. Consistencia	49
4.4. Normalidad Asintótica	60
Conclusiones y Trabajos Futuros	67
Apéndice	69
A.1. Demostraciones de lemas	69
A.2. Rutinas en R	113
A.2.1. Experimentos	113
A.2.2. Aplicación	126
Bibliografía	136

Introducción

Los métodos estadísticos para el análisis de datos indexados en el espacio han experimentado en las últimas décadas avances significativos. Este progreso, se ha visto acompañado y potenciado, por el vertiginoso desarrollo de la tecnología y de las ciencias de la computación, permitiendo introducir al espacio como eje fundamental de análisis, tal como ocurre por ejemplo, en estudios de planificación territorial, en trabajos de epidemiología panorámica y en modelos econométricos espaciales (Martori and Hoberg (2008)). Muchos de estos métodos se derivan de enfoques presentados en el contexto de series de tiempo y surgen como adaptaciones específicas de éstos últimos para hacer frente a los desafíos propios que presenta el análisis de datos indexados en el espacio. Estos datos no solo contienen información sobre un conjunto de atributos de interés, sino que también proporcionan la ubicación geográfica o espacial donde estos atributos fueron observados. Una de sus características clave es la autocorrelación espacial: Las observaciones próximas espacialmente tienden a ser más similares de lo que se espera que lo sean las observaciones que están más separadas en el espacio.

Un ejemplo claro de datos espaciales corresponde a observaciones medidas sobre grillas o redes, llamados datos regionalizados. En este caso la información se registra en todos los puntos de un dominio $D \subset \mathbb{Z}^2$ fijo, finito o infinito. Las ubicaciones espaciales de estos datos se denominan sitios y estos sitios generalmente representan regiones con una cierta área, tales como ocurre con los datos de imágenes. Desde el punto de vista de un observador común, una imagen x es una escena provista de ciertos atributos que le son propios y la caracterizan, tales como bordes, color, brillo, textura, etc.; mientras que desde una perspectiva formal, x constituye una realización de un proceso aleatorio $X : \Omega \rightarrow E^W$, donde E es un conjunto acompañado por una estructura de σ -álgebra, y W es una grilla rectangular finita, llamada “ventana de observación de la imagen x ”. En el contexto previo, $D = W$ y cada sitio representa la posición espacial precisa de un píxel de la imagen.

Uno de los enfoques para estudiar la autocorrelación espacial en datos regionalizados consiste en proponer modelos que especifican de manera directa la matriz de varianza-covarianza de los datos, la cual limita la medida de proximidad espacial a las distancias entre ubicaciones de puntos que se supone que representan diferentes regiones. Otro enfoque se basa en proponer modelos de autorregresión espacial, que permiten incorporar las estructuras de vecindario. Estos patrones, llamados modelos autorregresivos bidimensionales (modelos AR-2D), representan los datos en la ubicación s de la grilla D como una combinación lineal de valores registrados en posiciones espaciales de D vecinas a s y esta autoregresión es la encargada de inducir la dependencia espacial entre los datos. Whittle (1954) estudió por primera vez los modelos AR-2D y a partir de entonces éstos han sido exitosamente usados en diversas áreas del modelado y procesamiento de imágenes como filtrado y reconstrucción de imágenes satelitales (Bustos et al (2009), Ojeda et al (2010)) y detección de bordes en imágenes médicas (Quintana et al (2011), Zielinski et al (2010)).

En general, un modelo matemático es una simplificación de la realidad y requiere, por parte de

los datos, del cumplimiento de ciertos supuestos. Los procedimientos clásicos de estimación de los parámetros de un modelo son óptimos en situaciones en las que los supuestos del modelo se cumplen de manera rigurosa, pero proporcionan estimaciones desacertadas cuando las exigencias del modelo no se observan exactamente. Una alternativa a esta problemática es trabajar con procedimientos robustos, los cuales permiten relajar los requisitos que teóricamente deben satisfacer los datos y proporcionan la flexibilidad necesaria a los efectos de obtener estimaciones razonablemente aceptables de los parámetros del modelo, en circunstancias de no cumplimiento estricto de los supuestos.

La estimación robusta de los parámetros en modelos AR-2D ha sido tratada para datos afectados con dos tipos de contaminación: innovativa y aditiva. Los autores Kashyap and Eom (1988) proponen la clase de los M-estimadores 2D y en Allende et al (1998) y Allende et al (2001) se introducen los estimadores GM 2D. Los primeros tienen un buen desempeño bajo modelos contaminados innovativamente pero son inadecuados bajo contaminación aditiva, mientras que los estimadores GM 2D tienen un buen desempeño en ambos casos. Posteriormente, Ojeda (1999) presentó los estimadores RA 2D que además de tener un buen comportamiento bajo estos dos tipos de contaminación, poseen propiedades teóricas estadísticas interesantes, tales como la consistencia y la normalidad asintótica. Estas propiedades son de gran interés ya que permiten realizar estudios de inferencia estadística, a partir de test de hipótesis e intervalos de confianza. Sin embargo, el estimador RA 2D exhibe como desventaja un alto costo computacional.

La propuesta de este trabajo de tesis consiste en definir un nuevo estimador robusto de los parámetros del modelo AR-2D bajo otros escenarios de contaminación que incluyen a la contaminación aditiva, capaz de competir con los estimadores M, GM y RA 2D mencionados. Para evitar la propagación del efecto de un dato atípico cuando se calculan los residuos del modelo, se propone un nuevo enfoque que consiste en definir estos residuos usando un modelo auxiliar. Con la ayuda del modelo auxiliar, se define un estimador para los parámetros del modelo AR-2D resistente a diferentes tipos de outliers. La propuesta intenta generalizar a dos dimensiones los estimadores BMM desarrollados por Muler et al (2009) para modelos ARMA de series de tiempo. El nuevo estimador, BMM 2D, además de mejorar en cuanto a precisión y exactitud las propuestas ya conocidas, resulta ser consistente y normalmente asintótico.

El trabajo se organiza de la siguiente manera: en el Capítulo 1 se introducen conceptos y definiciones básicas sobre procesos espaciales, en particular, se definen los modelos AR-2D y se los analiza en el contexto del procesamiento de imágenes; en el Capítulo 2 se presenta la nueva propuesta; en el Capítulo 3 se analiza el comportamiento del nuevo estimador comparándolo mediante estudios de Monte Carlo con los estimadores LS, M, GM y RA 2D; en el Capítulo 4 se demuestran las propiedades teóricas del estimador BMM 2D. Finalmente, se establecen algunas conclusiones y posibles lineamientos para trabajos futuros.

Capítulo 1

Procesos Aleatorios Bidimensionales

En este capítulo se introducen los conceptos básicos necesarios para trabajar a lo largo de la tesis. Los mismos pueden encontrarse en Guyon (1995), Ojeda (1999) y Chuvieco (2008).

1.1. Preliminares

Sea \mathbb{R} el conjunto de los números reales y \mathcal{B} la σ -álgebra de Borel sobre \mathbb{R} y sea $\mathbb{R}^{\mathbb{Z}^2} = \{f : \mathbb{Z}^2 \rightarrow \mathbb{R} / f \text{ es función}\}$. Consideremos el espacio medible $(\mathbb{R}, \mathcal{B})$. Para cada $1 \leq n < \infty$ natural sea:

$$(\mathbb{Z}^2)^{(n)} = \{(k_1, l_1), \dots, (k_n, l_n)\} : (k_i, l_i) \in \mathbb{Z}^2, (k_i, l_i) \neq (k_j, l_j) \text{ si } i \neq j, 1 \leq i, j \leq n\}.$$

Además, definamos $\mathcal{O}(\mathbb{Z}^2)$ y $S(\mathbb{Z}^2)$ como:

$$\mathcal{O}(\mathbb{Z}^2) = \cup\{(\mathbb{Z}^2)^{(n)} : 1 \leq n < \infty, \text{ natural}\},$$

$$S(\mathbb{Z}^2) = \{S \subset \mathbb{Z}^2 : 0 < \#(S) < \infty\}.$$

Entonces, para cada $1 \leq n < \infty$, natural y para cada $\lambda = ((k_1, l_1), \dots, (k_n, l_n))$ en $(\mathbb{Z}^2)^{(n)}$, definimos la proyección sobre \mathbb{R}^n como la función $\sigma_\lambda : \mathbb{R}^{\mathbb{Z}^2} \rightarrow \mathbb{R}^n$ dada por:

$$\sigma_\lambda(X) = (X(k_1, l_1), \dots, X(k_n, l_n)), \quad \forall X \in \mathbb{R}^{\mathbb{Z}^2}.$$

Para cada $m \in \mathbb{N}$, sea \mathcal{B}^m la σ -álgebra producto sobre \mathbb{R}^m y sea $\mathcal{B}^{\mathbb{Z}^2}$ la mínima σ -álgebra sobre $\mathbb{R}^{\mathbb{Z}^2}$ que hace medibles las proyecciones σ_λ , $\forall \lambda \in \mathcal{O}(\mathbb{Z}^2)$.

Definición 1.1.1. *Sea (Ω, \mathcal{A}, P) un espacio de probabilidad. Entonces $X : \Omega \rightarrow \mathbb{R}^{\mathbb{Z}^2}$ es un proceso aleatorio bidimensional real, si $\forall A \in \mathcal{B}^{\mathbb{Z}^2}$, $X^{-1}(A) \in \mathcal{A}$.*

Notemos que de acuerdo a la definición anterior, $X : \Omega \rightarrow \mathbb{R}^{\mathbb{Z}^2}$ es un proceso aleatorio bidimensional real si y solo si, X es una función \mathcal{A} -medible con respecto a la σ -álgebra $\mathcal{B}^{\mathbb{Z}^2}$. Estos procesos son también llamados procesos aleatorios 2D reales o procesos aleatorios sobre \mathbb{Z}^2 reales. Una manera frecuente de denotar al proceso X es mediante la sucesión de variables aleatorias:

$$X = \{X_{i,j}\}_{(i,j) \in \mathbb{Z}^2},$$

donde para todo $(i, j) \in \mathbb{Z}^2$, $X_{i,j} : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria definida sobre (Ω, \mathcal{A}, P) tal que $X_{i,j}(\omega) = (X(\omega))(i, j) \forall \omega \in \Omega$.

Definición 1.1.2. Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y sea $X : \Omega \rightarrow \mathbb{R}^{\mathbb{Z}^2}$ un proceso aleatorio 2D real. Definimos como P_X a la probabilidad sobre $(\mathbb{R}^{\mathbb{Z}^2}, \mathcal{B}^{\mathbb{Z}^2})$ dada por $P_X(A) = P(X^{-1}(A)) \forall A \in \mathcal{B}^{\mathbb{Z}^2}$.

Definición 1.1.3. Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y sea X un proceso aleatorio 2D real definido sobre (Ω, \mathcal{A}, P) . Para cada $n \in \mathbb{N}$ se definen las funciones de distribución n -dimensionales como:

$$F_{\underline{(k,l)}_n}(x_1, \dots, x_n) = P(X_{k_1, l_1} < x_1, \dots, X_{k_n, l_n} < x_n),$$

donde $\underline{(k,l)}_n = ((k_1, l_1), \dots, (k_n, l_n)) \in (\mathbb{Z}^2)^n$ y $(x_1, \dots, x_n) \in \mathbb{R}^n$.

Observación 1.1.1. Las funciones de distribución n -dimensionales son funciones de distribución de probabilidad; más precisamente $F_{\underline{(k,l)}_n}$ es la función de distribución de probabilidad conjunta de las variables aleatorias $X_{k_1, l_1}, \dots, X_{k_n, l_n}$.

Definición 1.1.4. Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y sea $X = \{X_{i,j}\}_{(i,j) \in \mathbb{Z}^2}$ un proceso aleatorio 2D real definido sobre (Ω, \mathcal{A}, P) . Decimos que X es un proceso con momento de segundo orden finito si $\forall (i,j) \in \mathbb{Z}^2, X_{i,j} \in \mathcal{L}^2(\Omega, \mathcal{A}, P, \mathbb{R})$.

Definición 1.1.5. Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y sea $X = \{X_{i,j}\}_{(i,j) \in \mathbb{Z}^2}$ un proceso aleatorio 2D real definido sobre (Ω, \mathcal{A}, P) , con momento de segundo orden finito.

a) Se llama función de media del proceso X a $\mu_X : \mathbb{Z}^2 \rightarrow \mathbb{R}$ dada por:

$$\mu_X(i, j) = E_P(X_{i,j}), \quad \forall (i, j) \in \mathbb{Z}^2.$$

b) Se llama función de autocovarianza del proceso X a $C_X : \mathbb{Z}^2 \times \mathbb{Z}^2 \rightarrow \mathbb{R}$ dada por:

$$C_X((i, j), (s, t)) = E_P(X_{i,j}X_{s,t}) - E_P(X_{i,j})E_P(X_{s,t}), \quad \forall (i, j), (s, t) \in \mathbb{Z}^2.$$

Observación 1.1.2. Si X es un proceso aleatorio 2D real definido sobre (Ω, \mathcal{A}, P) , con momento de segundo orden finito, y función de autocovarianza C_X , $\forall (i, j) \in \mathbb{Z}^2$ la variable $X_{i,j}$ tiene varianza finita dada por $C_X((i, j), (i, j))$. Así, la función $\sigma_X^2 : \mathbb{Z}^2 \rightarrow \mathbb{R}$ dada por:

$$\sigma_X^2(i, j) = C_X((i, j), (i, j)) \quad \forall (i, j) \in \mathbb{Z}^2,$$

define la función de varianza del proceso X .

NOTA: De aquí en adelante, al referirnos al proceso X , asumiremos (salvo expresa mención en contra) los siguientes dos supuestos:

- i) Que X es un proceso aleatorio 2D real, definido sobre el espacio de probabilidad (Ω, \mathcal{A}, P) .
- ii) Que el proceso X tiene momento de segundo orden finito, con funciones de media, autocovarianza y varianza, denotadas respectivamente como μ_X , C_X y σ_X^2 .

Definición 1.1.6. Decimos que el proceso X es P -débilmente estacionario si:

$$i) \mu_X(i, j) = \mu_X(s, t) \quad \forall (i, j), (s, t) \in \mathbb{Z}^2,$$

$$ii) C_X((i, j), (s, t)) = C_X((i-s, j-t), (0, 0)) \quad \forall (i, j), (s, t) \in \mathbb{Z}^2.$$

Definición 1.1.7. Decimos que el proceso X es P -estrictamente estacionario si para cada $n \in \mathbb{N}$ y para cada $(i_1, j_1), (i_2, j_2), \dots, (i_n, j_n), (s, t) \in \mathbb{Z}^2$ la función de distribución de probabilidad conjunta de las variables aleatorias

$$X_{i_1, j_1}, X_{i_2, j_2}, \dots, X_{i_n, j_n},$$

es igual a la función de distribución de probabilidad conjunta de las variables aleatorias

$$X_{i_1+s, j_1+t}, X_{i_2+s, j_2+t}, \dots, X_{i_n+s, j_n+t}.$$

Observación 1.1.3. Si X es un proceso P -estrictamente estacionario, entonces X es P -débilmente estacionario (Ojeda (1999)).

1.2. Ergodicidad

Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y sea $X = \{X_{i,j}\}_{(i,j) \in \mathbb{Z}^2}$ un proceso aleatorio P -estrictamente estacionario y consideremos el espacio medible $(\mathbb{R}^{\mathbb{Z}^2}, \mathcal{B}^{\mathbb{Z}^2})$.

Definición 1.2.1. Para cada $(m, n) \in \mathbb{Z}^2$ se define la traslación por (m, n) a la función $\tau_{(m,n)} : \mathbb{R}^{\mathbb{Z}^2} \rightarrow \mathbb{R}^{\mathbb{Z}^2}$ dada por $\tau_{(m,n)}(X)(i, j) = X_{i-m, j-n}$.

Observación 1.2.1. $\forall (m, n) \in \mathbb{Z}^2$, $\tau_{(m,n)}$ es biyectiva y medible.

Definición 1.2.2. Sea $\tau : \mathbb{R}^{\mathbb{Z}^2} \rightarrow \mathbb{R}^{\mathbb{Z}^2}$ biyectiva y medible. Se dice que $A \in \mathcal{B}^{\mathbb{Z}^2}$ es τ -invariante si $\tau(A) = A$.

Sea

$$\mathcal{I} = \{A \in \mathcal{B}^{\mathbb{Z}^2} : A \text{ es } \tau_{(m,n)}\text{-invariante } \forall (m, n) \in \mathbb{Z}^2\}.$$

Definición 1.2.3. Decimos que X es P -ergódico si $\forall A \in \mathcal{I}$, $P_X(A) = 0$ ó $P_X(A) = 1$.

Observación 1.2.2. 1) \mathcal{I} es una σ -álgebra. 2) X es P -estrictamente estacionario si $P_X \equiv P_{\tau_{(m,n)}(X)}$, $\forall (m, n) \in \mathbb{Z}^2$.

Definición 1.2.4. Sea $D \subset \mathbb{Z}^2$, convexo. Se llama diámetro interior de D a

$$d(D) = \sup\{r > 0 : \exists (i, j) \in D \text{ con } B((i, j), r) \subset D\},$$

donde $B((i, j), r) = \{(m, n) \in \mathbb{Z}^2 : (i-m)^2 + (j-n)^2 < r^2\}$.

Observación 1.2.3. Nos referimos a conjuntos convexos en \mathbb{Z}^2 como la intersección de convexos de \mathbb{R}^2 con \mathbb{Z}^2 .

Teorema 1.2.1. (Teorema Ergódico). Sea X P -estrictamente estacionario y P -ergódico, $g \in \mathcal{L}^1(\mathbb{R}^{\mathbb{Z}^2}, \mathcal{B}^{\mathbb{Z}^2}, P_X, \mathbb{R})$ y $\{D_n\}_{n \in \mathbb{N}}$ sucesión acotada de convexos acotados en \mathbb{Z}^2 tal que $d(D_n) \xrightarrow{n \rightarrow \infty} \infty$. Entonces:

$$i) \frac{1}{\#D_n} \sum_{(i,j) \in D_n} g(\tau_{(i,j)}(X)) \rightarrow E_P(g(X)) \text{ en } \mathcal{L}^1(\mathbb{R}^{\mathbb{Z}^2}, \mathcal{B}^{\mathbb{Z}^2}, P_X, \mathbb{R}).$$

ii) Si además $\{D_n\}_{n \in \mathbb{N}}$ es creciente, entonces la convergencia en (i) ocurre excepto sobre conjuntos de medida P -nula, lo que se denotará es a.e..

Dem.: Ver Guyon (1995). □

1.3. Orden en \mathbb{Z}^2

En el contexto de las series de tiempo indexadas en $T \subset \mathbb{Z}$ el concepto de “pasado” de una posición $t \in T$ se define de manera evidente a partir del orden total en \mathbb{Z} (el pasado de t es el conjunto $\{t' \in \mathbb{Z} : t' < t\}$). Por el contrario, en \mathbb{Z}^2 no se dispone de un orden consustancial que determine de manera natural cuál es el pasado de una posición $(i, j) \in \mathbb{Z}^2$. Dicho orden será de gran importancia por dos motivos: en primer lugar, en las próximas secciones y capítulos hará falta realizar sumas infinitas sobre \mathbb{Z}^2 y tomar límites que convergan cuando un punto $(m, n) \in \mathbb{Z}^2$ tiende a infinito por lo que tendremos que ordenar los elementos de alguna manera; en segundo lugar, el orden del vector de parámetros a estimar en el modelo de estudio determinará como es la matriz de varianza-covarianza del estimador que se propondrá más adelante en esta tesis.

Dado un elemento (i, j) , puede definirse el pasado de (i, j) de diferentes maneras. En este trabajo, necesitamos algún orden total en \mathbb{Z}^2 . Antes de introducir el orden total de \mathbb{Z}^2 con el que trabajaremos, introducimos algunas definiciones que se relacionan con los “vecinos pasados” de una posición $(i, j) \in \mathbb{Z}^2$.

Definición 1.3.1. Para cada $(i, j) \in \mathbb{Z}^2$,

a) Se llama región de predicción no causal de (i, j) a:

$$S_1(i, j) = \{(k, l) \in \mathbb{Z}^2 : (k, l) \neq (i, j)\}.$$

b) Se llama región de predicción semicausal de (i, j) a:

$$S_2(i, j) = \{(k, l) \in \mathbb{Z}^2 : k < i\} \cup \{(i, l) \in \mathbb{Z}^2 : l \neq j\}.$$

c) Se llama región de predicción causal de (i, j) a:

$$S_3(i, j) = \{(k, l) \in \mathbb{Z}^2 : k < i\} \cup \{(i, l) \in \mathbb{Z}^2 : l < j\}.$$

d) Se llama región de predicción fuertemente causal de (i, j) a:

$$S_4(i, j) = \{(k, l) \in \mathbb{Z}^2 : k < i, l \leq j\} \cup \{(i, l) \in \mathbb{Z}^2 : l < j\}.$$

Estas regiones de causalidad pueden apreciarse gráficamente en las imágenes de la Fig. 1.1.

Definición 1.3.2. Dado $(i, j) \in \mathbb{Z}^2$ y $M \in \mathbb{N}$ tal que $i, j < M$, se llama ventana fuertemente causal cuadrada de orden M en (i, j) a:

$$W_M(i, j) = \{(k, l) \in S_4(i, j) : i - M \leq k \leq i, j - M \leq l \leq j\} \cup \{(i, j)\}.$$

En esta tesis trabajaremos siempre sobre regiones de predicción fuertemente causales y con ventanas fuertemente causales cuadradas. Cuando $(i, j) = (M, M)$, $W_M(M, M)$ se indicará W_M ; esta notación se usará también en aquellos casos en que la mención específica del punto (i, j) no sea relevante.

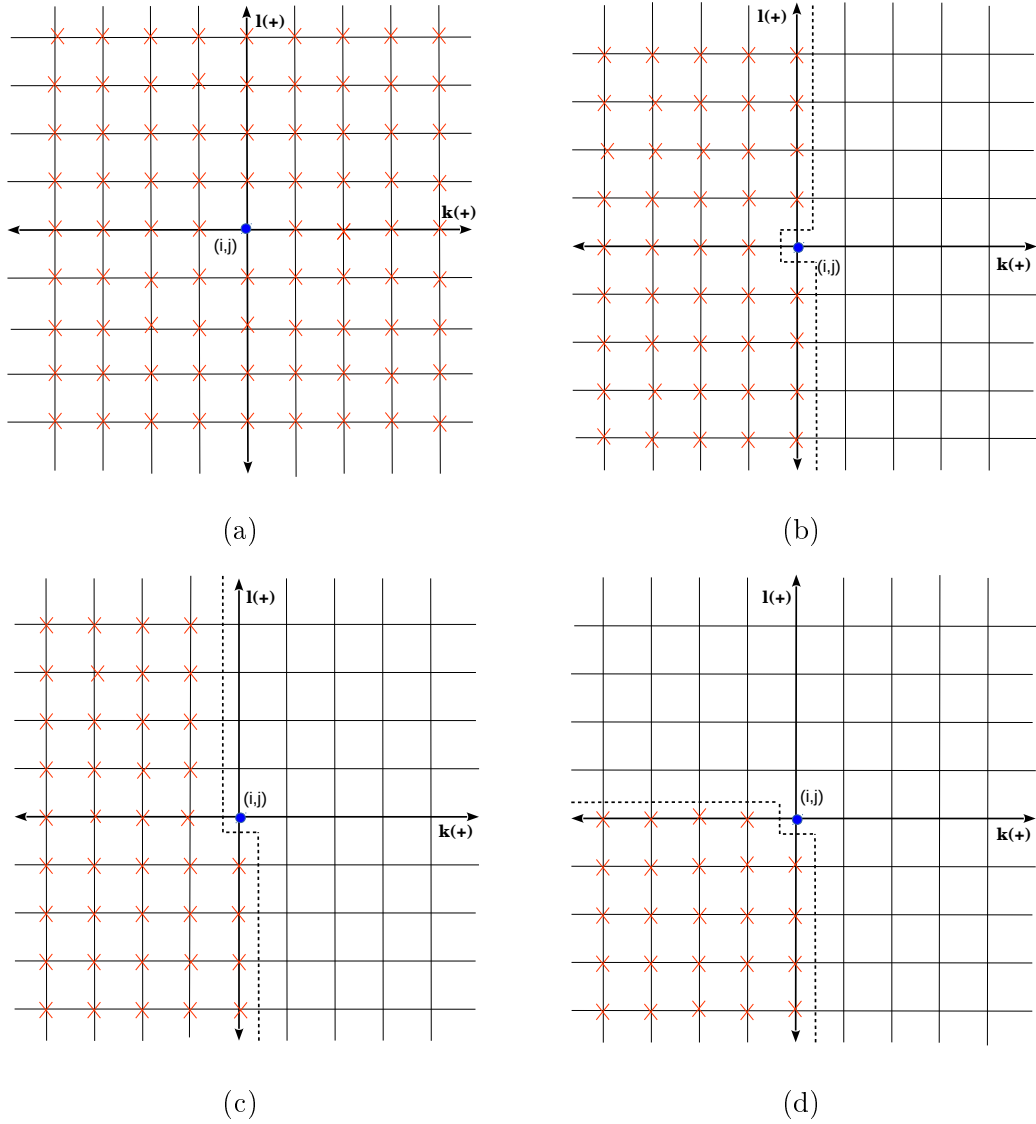


Figura 1.1: (a) $S_1(i, j)$; (b) $S_2(i, j)$; (c) $S_3(i, j)$; (d) $S_4(i, j)$

A continuación, daremos a \mathbb{Z}^2 un orden total al que llamamos “orden espiral”.

Sea $(i, j) \in \mathbb{Z}^2 \setminus \{(0, 0)\}$ y γ una función de (i, j) que puede asumir los valores $|i|$ ó $|j|$. Dado (i, j) fijo, se definen los conjuntos:

$$\begin{aligned}
 A_1(\gamma(i, j)) &= \{(k, l) : |k|, |l| < \gamma(i, j)\}, \\
 A_2(\gamma(i, j)) &= \{(k, l) : k = \gamma(i, j), -\gamma(i, j) < l < j\}, \\
 A_3(\gamma(i, j)) &= \{(k, l) : k = \gamma(i, j), -\gamma(i, j) < l < \gamma(i, j)\}, \\
 A_4(\gamma(i, j)) &= \{(k, l) : l = \gamma(i, j), -\gamma(i, j) < k \leq \gamma(i, j)\}.
 \end{aligned}$$

Ahora definimos cuatro conjuntos disjuntos, R_k , cuya unión es $\mathbb{Z}^2 \setminus \{(0,0)\}$ y el conjunto de “pasados” de cualquier punto $(i,j) \in \mathbb{Z}^2 \setminus \{(0,0)\}$, $P_{(i,j)}^k$:

I) Consideremos $i > 0$ y $\gamma(i,j) = |i| = i$, luego

$$R_1 = \{(i,j) \in \mathbb{Z}^2 : i > 0, -\gamma(i,j) < j \leq \gamma(i,j)\} \quad y$$

$$P_{(i,j)}^1 = A_1(\gamma(i,j)) \cup A_2(\gamma(i,j)).$$

II) Consideremos $j > 0$ y $\gamma(i,j) = |j| = j$ luego

$$R_2 = \{(i,j) \in \mathbb{Z}^2 : j > 0, -\gamma(i,j) \leq i < \gamma(i,j)\} \quad y$$

$$P_{(i,j)}^2 = A_1(\gamma(i,j)) \cup A_3(\gamma(i,j)) \cup \{(k,l) : l = \gamma(i,j), i < k \leq \gamma(i,j)\}.$$

III) Consideremos $i < 0$ y $\gamma(i,j) = |i| = -i$, luego

$$R_3 = \{(i,j) \in \mathbb{Z}^2 : i < 0, -\gamma(i,j) \leq j < \gamma(i,j)\} \quad y$$

$$P_{(i,j)}^3 = A_1(\gamma(i,j)) \cup A_3(\gamma(i,j)) \cup A_4(\gamma(i,j)) \cup \{(k,l) : k = -\gamma(i,j), j < l \leq \gamma(i,j)\}.$$

IV) Consideremos $j < 0$ y $\gamma(i,j) = |j| = -j$, luego

$$R_4 = \{(i,j) \in \mathbb{Z}^2 : j < 0, -\gamma(i,j) < i \leq \gamma(i,j)\} \quad y$$

$$P_{(i,j)}^4 = A_1(\gamma(i,j)) \cup A_3(\gamma(i,j)) \cup A_4(\gamma(i,j)) \cup \{(k,l) : k = -\gamma(i,j), -\gamma(i,j) < l \leq \gamma(i,j)\}$$

$$\cup \{(k,l) : l = -\gamma(i,j), -\gamma(i,j) \leq k < i\}.$$

Luego, $\mathbb{Z}^2 \setminus \{(0,0)\} = \cup_{i=1}^4 R_i$ donde la unión es disjunta; es decir, $\{R_i\}_{i=1}^4$ es una partición de $\mathbb{Z}^2 \setminus \{(0,0)\}$.

Se define el orden espiral en \mathbb{Z}^2 (orden total en \mathbb{Z}^2) a través de la relación \preceq de la siguiente manera:

Definición 1.3.3. Sean $(i,j), (k,l) \in \mathbb{Z}^2$. Asumamos que $(i,j) \in R_s$ con $s \in \{1,2,3,4\}$. Diremos que $(k,l) \preceq (i,j)$ si se satisface alguna de las siguientes condiciones;

- $(i,j) = (k,l)$, es decir, $i = k$ y $j = l$, o bien;
- $(k,l) \in P_{(i,j)}^s$.

Este ordenamiento espiral de elementos de \mathbb{Z}^2 se esquematiza en la Fig. 1.2 (a).

Consideremos ahora el conjunto

$$I = \{(i,j) \in \mathbb{Z}^2 : i \geq 0, j \geq 0\}.$$

Definición 1.3.4. Definimos sobre I un orden total, dado por la restricción a I del orden espiral \preceq definido previamente para \mathbb{Z}^2 .

El orden total sobre I se observa de manera esquemática en la Fig. 1.2 (b).

De ahora en más, a menos que se haga explícita mención de lo contrario, cualquier suma que se haga sobre elementos indexados en I se realizará siguiendo el orden total definido sobre I .

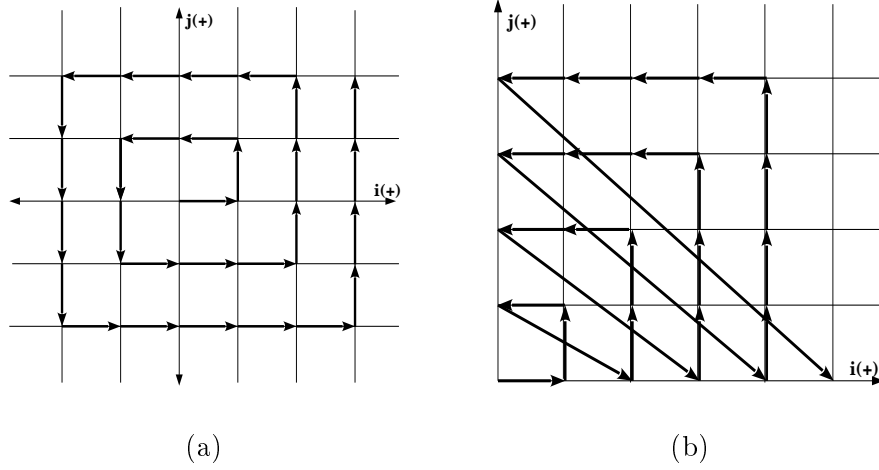


Figura 1.2: (a) Orden espiral en \mathbb{Z}^2 ; (b) Orden espiral restringido a I .

1.4. Martingalas

Dado $X = \{X_{i,j}\}_{(i,j) \in \mathbb{Z}^2}$ un proceso espacial definido sobre (Ω, \mathcal{A}, P) a valores en \mathbb{R} y \preceq alguna relación de orden parcial en I , podemos introducir la siguiente definición:

Definición 1.4.1. Dada una sucesión creciente de sub σ -álgebras de \mathcal{A} , $\mathcal{F}_{i,j}$, i.e. $\mathcal{F}_{k,l} \subseteq \mathcal{F}_{i,j}$ si $(k,l) \preceq (i,j)$, y $X_{i,j}$ que sea $\mathcal{F}_{i,j}$ -medible e integrable. Diremos que $(X_{i,j}, \mathcal{F}_{i,j})$ es un arreglo diferencia martingala si $E(X_{i,j} | \mathcal{F}_{k,l}) = 0$ cuando $(k,l) \preceq (i,j)$ y $(k,l) \neq (i,j)$.

En la sección anterior se definió un orden total sobre I . Sin embargo, en esta sección es necesario contar con un orden parcial que permita construir a partir de un proceso $X = \{X_{i,j}\}_{(i,j) \in \mathbb{Z}^2}$ definido sobre (Ω, \mathcal{A}, P) , una sucesión de sub σ -álgebras de \mathcal{A} crecientes como en la definición anterior. Es así que introduciremos la siguiente definición con el objetivo de poder definir este nuevo orden sobre I :

Definición 1.4.2. Para cada elemento $(i,j) \in I$, se define la “vecindad” de la posición (i,j) , $V_{i,j}$, como:

$$V_{i,j} = \{(k,l) \in I : 0 \leq k \leq i, 0 \leq l \leq j\} \setminus \{(i,j)\}.$$

Con este concepto de vecindad es ahora posible establecer la siguiente relación de orden parcial \preceq en I :

Definición 1.4.3. Dado $(k, l), (i, j) \in I$ decimos que (k, l) está relacionado con (i, j) bajo la relación \preceq (denotamos $(k, l) \preceq (i, j)$) si y solamente si $V_{k,l} \subseteq V_{i,j}$.

Observese que $(k, l) \preceq (i, j)$ es equivalente a decir $k \leq i$ y $l \leq j$.

Luego, definimos para cada $(i, j) \in I$, la σ -álgebra $\mathcal{F}_{i,j}$ generada por el conjunto de variables aleatorias $\{X_{k,l} : (k, l) \in V_{i,j}\}$, es decir,

$$\mathcal{F}_{i,j} = \sigma(\{X_{k,l} : (k, l) \in V_{i,j}\}).$$

Se puede probar que dado $(i, j) \in I$, $\mathcal{F}_{k,l} \subseteq \mathcal{F}_{i,j}$ para cada $(k, l) \preceq (i, j)$; es decir, $\{\mathcal{F}_{k,l}\}$ es una sucesión no decreciente de sub σ -álgebras de \mathcal{A} .

Observación 1.4.1. La condición $E(X_{i,j} | \mathcal{F}_{k,l}) = 0$ si $(k, l) \preceq (i, j)$ y $(k, l) \neq (i, j)$, de la definición 1.4.1 es equivalente a la condición $E(X_{i,j} | X_{k,l}) = 0$ para todo $(k, l) \preceq (i, j)$ y $(k, l) \neq (i, j)$.

Dada una sucesión $\{b_{(i,j)} : (i, j) \in I\}$ de números reales positivos tal que $b_{(0,0)} = b_{(i,0)} = b_{(0,j)} = 0$, se define $\Delta b_{(i,j)} = b_{(i,j)} - b_{(i,j-1)} - b_{(i-1,j)} + b_{(i-1,j-1)}$. Como corolario del Teorema 2.3 del trabajo Quang and Van Huan (2010) se tiene:

Teorema 1.4.1. Dado $X = \{X_{i,j}\}_{(i,j) \in \mathbb{Z}^2}$ un proceso espacial real definido sobre (Ω, \mathcal{A}, P) . Para todo arreglo diferencia martingala $(X_{i,j}, \mathcal{F}_{i,j})$, y para toda sucesión de números positivos $\{b_{(i,j)}\}$ que satisface que $\Delta b_{(i,j)} \geq 0 \forall (i, j) \in I$ y tal que $b_{(i,j)} \rightarrow \infty$ cuando $(i, j) \rightarrow \infty$ (si y solo si $i \rightarrow \infty$ y $j \rightarrow \infty$), la condición

$$\sum_{(0,0) \preceq (i,j)} \frac{E(|X_{i,j}|^2)}{b_{(i,j)}^2} < \infty$$

implica que

$$\frac{1}{b_{(M,M)}} \sum_{(0,0) \preceq (i,j) \preceq (M,M)} X_{i,j} \rightarrow 0 \quad \text{a.e. cuando } M \rightarrow \infty.$$

Dem.: Ver Quang and Van Huan (2010). □

Observación 1.4.2. La versión general del teorema anterior requiere que el espacio E sobre el cual está definido el proceso X sea un espacio 2-smoothable. Se dice que un espacio de Banach E es 2-smoothable si existe una norma equivalente bajo la cual E es 2-uniformly smooth, es decir, $\rho(t) := \sup \left\{ \frac{\|x+y\| + \|x-y\|}{2} - 1 : \forall x, y \in E, \|x\| = 1, \|y\| = t \right\} = \mathcal{O}(t^2)$ (orden t^2). Esto se puede ver en Quang and Van Huan (2010). En nuestro caso, el espacio E resulta ser \mathbb{R} que es un espacio de Hilbert y por lo visto en Chao and Woyczynski (2006), todo Hilbert es 2-smoothable.

Este teorema se conoce como Ley Fuerte de los Grandes Números para arreglos diferencia martingalas.

Sea $M \in \mathbb{N}$ y $D(M) = \{(k, l) \in I : 0 \leq k < M, 0 \leq l < M\}$. Como corolario del Teorema 2.1 de Cohen (2016) se tiene el siguiente teorema:

Teorema 1.4.2. *Dado un arreglo diferencia martingala $(X_{i,j}, \mathcal{F}_{i,j})$ con $\{X_{i,j}\}$ proceso aleatorio real definido sobre (Ω, \mathcal{A}, P) con momento de segundo orden finito, fuertemente estacionario y ergódico. Entonces $\frac{1}{\sqrt{M}} \sum_{(k,l) \in D(M)} X_{k,l}$ converge en distribución a $\mathcal{N}(0, E(|X_{0,0}|^2))$ cuando $M \rightarrow \infty$.*

Dem.: Ver Cohen (2016). □

Este teorema es conocido como el Teorema Central del Límite para arreglos diferencia martingalas.

A continuación se presentan algunos conceptos previos que permiten introducir la definición de un proceso AR bidimensional.

1.5. Procesos AR Bidimensionales Reales

Sea $I = \{(i, j) \in \mathbb{Z}^2 : i \geq 0, j \geq 0\}$ y sea $M \in \mathbb{N}$.

Definición 1.5.1. *Sea T el subconjunto finito de I dado por: $T = \{(k, l) \in I : k \leq L, l \leq L, (k, l) \neq (0, 0)\}$, con $L \in \mathbb{N}$, conocido y $L < M$. Y sea $\{\phi_{m,n}\}_{(m,n) \in T}$ una sucesión de números en $\mathbb{R} \setminus \{0\}$ indexada por T . Se dice que la función $\Phi : \mathbb{C}^2 \rightarrow \mathbb{C}$ dada por*

$$\Phi(z_1, z_2) = 1 - \sum_{(k,l) \in T} \phi_{k,l} z_1^k z_2^l,$$

es un polinomio unilateral con soporte T .

Observación 1.5.1. *Un polinomio unilateral es una función analítica en \mathbb{C}^2 .*

Definición 1.5.2. *Sea $Y : \Omega \rightarrow \mathbb{R}^{\mathbb{Z}^2}$ un proceso aleatorio bidimensional real definido sobre (Ω, \mathcal{A}, P) .*

- a) *Para cada $k \in \mathbb{Z}$ se define el operador de retroceso unidimensional de orden k , a la función $B_1^k : \mathbb{R}^{\mathbb{Z}^2} \rightarrow \mathbb{R}^{\mathbb{Z}^2}$ dada por: $B_1^k(Y)(i, j) = Y_{i-k, j}, \forall (i, j) \in \mathbb{Z}^2$.*
- b) *Para cada $l \in \mathbb{Z}$ se define el operador de retroceso unidimensional de orden l , a la función $B_2^l : \mathbb{R}^{\mathbb{Z}^2} \rightarrow \mathbb{R}^{\mathbb{Z}^2}$ dada por: $B_2^l(Y)(i, j) = Y_{i, j-l}, \forall (i, j) \in \mathbb{Z}^2$.*
- c) *Para cada $(k, l) \in \mathbb{Z}^2$ se define el operador de retroceso bidimensional de orden (k, l) , a la función $B^{(k,l)} : \mathbb{R}^{\mathbb{Z}^2} \rightarrow \mathbb{R}^{\mathbb{Z}^2}$ dada por: $B^{(k,l)}(Y)(i, j) = Y_{i-k, j-l}, \forall (i, j) \in \mathbb{Z}^2$.*

Observación 1.5.2. *El operador de retroceso bidimensional $B^{(k,l)}$ se puede escribir en término de los operadores de retroceso unidimensionales B_1^k y B_2^l . Esto es:*

$$B^{(k,l)} = B_1^k \circ B_2^l = B_2^l \circ B_1^k. \quad (1.5.1)$$

Definición 1.5.3. *Sea $\{a_{k,l}\}_{(k,l) \in I}$ sucesión de números reales y sea $f : D \subset \mathbb{C}^2 \rightarrow \mathbb{C}$, D abierto, tal que*

$$f(z_1, z_2) = \sum_{(k,l) \in I} a_{k,l} z_1^k z_2^l \quad \text{uniformemente en } D.$$

Se define el operador funcional de retroceso bidimensional asociado a f , $f(B) : \mathbb{R}^{\mathbb{Z}^2} \rightarrow \mathbb{R}^{\mathbb{Z}^2}$ dado por:

$$f(B)(Y)(i, j) = \sum_{(k,l) \in I} a_{k,l} B^{(k,l)}(Y)(i, j), \quad \forall Y \in \mathbb{R}^{\mathbb{Z}^2} \quad \forall (i, j) \in \mathbb{Z}^2.$$

Por (1.5.1), $\forall Y \in \mathbb{R}^{\mathbb{Z}^2}$, $f(B)(Y)(i, j) = \sum_{(k,l) \in I} a_{k,l}(B_1^k \circ B_2^l)(Y)(i, j) \quad \forall (i, j) \in \mathbb{Z}^2$. Por lo tanto, el operador funcional de retroceso bidimensional asociado a f se denotará indistintamente $f(B)$ ó $f(B_1, B_2)$.

Definición 1.5.4. Sea $\varepsilon = \{\varepsilon_{i,j}\}_{(i,j) \in \mathbb{Z}^2}$ un proceso aleatorio 2D, real con varianza finita σ_ε^2 , definido sobre el espacio probabilístico (Ω, \mathcal{A}, P) . Entonces se dice que ε es un proceso de ruido blanco si:

$$i) \quad E_P(\varepsilon_{i,j}) = 0, \quad \forall (i, j) \in \mathbb{Z}^2 \text{ y}$$

ii)

$$E_P(\varepsilon_{i,j} \varepsilon_{k,l}) = \begin{cases} \sigma_\varepsilon^2, & \text{si } (i, j) = (k, l); \\ 0, & \text{c.c..} \end{cases}$$

Definición 1.5.5. Sean (Ω, \mathcal{A}, P) espacio de probabilidad, $\varepsilon = \{\varepsilon_{i,j}\}_{(i,j) \in \mathbb{Z}^2}$ proceso de ruido blanco definido sobre (Ω, \mathcal{A}, P) , $Y = \{Y_{i,j}\}_{(i,j) \in \mathbb{Z}^2}$ proceso aleatorio bidimensional estacionario definido sobre (Ω, \mathcal{A}, P) con media μ y $\Phi : \mathbb{C}^2 \rightarrow \mathbb{C}$ polinomio unilateral con soporte T , como en la definición 1.5.1, que no se anula en $D^* = \{(z_1, z_2) \in \mathbb{C}^2 : |z_1| \leq 1, |z_2| \leq 1\}$. Entonces se dice que Y es un proceso AR-2D con polinomio unilateral Φ y proceso de ruido blanco ε , si $\forall (i, j) \in \mathbb{Z}^2$:

$$Y_{i,j} = \mu + \sum_{(k,l) \in T} \phi_{k,l}(Y_{i-k, j-l} - \mu) + \varepsilon_{i,j}. \quad (1.5.2)$$

O bien, en forma compacta:

$$\Phi(B_1, B_2)(Y - \mu) = \varepsilon.$$

Notación: El proceso Y se denota como $AR(\Phi, \varepsilon)$.

1.5.1. Representaciones AR 2D Fuertemente Causales

Sea Y un proceso aleatorio bidimensional real con momento de segundo orden finito, definido sobre (Ω, \mathcal{A}, P) . Para cada $(m, n) \in \mathbb{Z}^2$ se define el conjunto $\mathcal{H}_{m,n}(Y)$ como la clausura del espacio vectorial real generado por las realizaciones del proceso Y en la región $S_4(m, n) \cup \{(m, n)\}$, es decir:

$$\mathcal{H}_{m,n}(Y) = \overline{\text{span}}\{Y_{k,l} : (k, l) \in S_4(m, n) \cup \{(m, n)\}\}.$$

Definición 1.5.6. Sea Y un proceso aleatorio 2D real definido sobre (Ω, \mathcal{A}, P) con momento de segundo orden finito y sea ε un proceso de ruido blanco sobre (Ω, \mathcal{A}, P) . Decimos que ε es el proceso de innovación fuertemente causal de Y si $\mathcal{H}_{m,n}(Y) \subseteq \mathcal{H}_{m,n}(\varepsilon) \quad \forall (m, n) \in \mathbb{Z}^2$.

Observación 1.5.3. Si Y es un proceso $AR(\Phi, \varepsilon)$, entonces $\mathcal{H}_{m,n}(\varepsilon) \subseteq \mathcal{H}_{m,n}(Y) \quad \forall (m, n) \in \mathbb{Z}^2$.

Definición 1.5.7. Sea Y un proceso $AR(\Phi, \varepsilon)$. Decimos que (Φ, ε) es una representación fuertemente causal de Y si ε es el proceso de innovación fuertemente causal de Y , es decir, $\mathcal{H}_{m,n}(Y) = \mathcal{H}_{m,n}(\varepsilon) \quad \forall (m, n) \in \mathbb{Z}^2$.

Dado un proceso Y $AR(\Phi, \varepsilon)$, con polinomio unilateral Φ dado por:

$$\Phi(z_1, z_2) = 1 - \sum_{(k,l) \in T} \phi_{k,l} z_1^k z_2^l,$$

una condición suficiente para que (Φ, ε) sea una representación fuertemente causal de Y es que

$$\sum_{(k,l) \in T} |\phi_{k,l}| < 1 \quad (1.5.3)$$

(ver Guyon (1995)). En lo que sigue de esta tesis asumiremos que los procesos $AR(\Phi, \varepsilon)$ con los que trabajamos satisfacen siempre esta condición.

En este trabajo nos interesa la estimación de los parámetros de un proceso Y , $AR(\Phi, \varepsilon)$, y media μ tal que satisface la condición (1.5.3). En este caso, es posible expresar (1.5.2) como:

$$\dot{Y}_{i,j} = \sum_{(k,l) \in I} \lambda_{k,l} \varepsilon_{i-k,j-l}, \quad (1.5.4)$$

donde $\dot{Y}_{i,j} = Y_{i,j} - \mu$, $\lambda_{k,l} \in \mathbb{R} \forall (k,l) \in I$ y $\lambda_{0,0} = 1$. A esta representación se la llama, “Representación de Medias Móviles del proceso Y ”.

Observación 1.5.4. *La función $\Phi(z_1, z_2)^{-1}$ es una función analítica en $\{(z_1, z_2) \in \mathbb{C}^2 : \Phi(z_1, z_2) \neq 0\}$. Si el polinomio Φ satisface la condición (1.5.3), $\Phi(z_1, z_2)^{-1}$ puede representarse de una única manera como una serie de potencias convergente en $D^* \subset \{(z_1, z_2) \in \mathbb{C}^2 : \Phi(z_1, z_2) \neq 0\}$, con coeficientes reales:*

$$\Phi(z_1, z_2)^{-1} = \sum_{(k,l) \in I} a_{k,l} z_1^k z_2^l,$$

donde $a_{k,l} \in \mathbb{R}$, $\forall (k,l) \in I$ y $\sum_{(k,l) \in I} |a_{k,l}| < \infty$ (Dieudonné (2013)). Cabe destacar que como $\ell^1 \subsetneq \ell^2$ se tiene que $\sum_{(k,l) \in I} a_{k,l}^2 < \infty$.

Entonces, el proceso $AR(\Phi, \varepsilon)$ fuertemente causal se puede escribir como:

$$\begin{aligned} \dot{Y}_{i,j} &= \Phi(B_1, B_2)^{-1} \varepsilon_{i,j}, \\ \dot{Y}_{i,j} &= \sum_{(k,l) \in I} a_{k,l} \varepsilon_{i-k,j-l} \end{aligned} \quad (1.5.5)$$

Luego, por la unicidad, se tiene que $a_{k,l} = \lambda_{k,l} \forall (k,l) \in I$.

Observación 1.5.5. *La condición (1.5.3) resulta ser crucial para la identificabilidad de un proceso $AR(\Phi, \varepsilon)$, es decir:*

Sea $\varepsilon = \{\varepsilon_{i,j}\}$ un proceso de innovaciones y $\{\phi_{k,l}\}_{(k,l) \in T}$ los coeficientes del polinomio Φ , entonces si $\sum_{(k,l) \in T} |\phi_{k,l}| < 1$, existe un único proceso fuertemente causal $X = \{X_{i,j}\}$ tal que $\Phi(B_1, B_2)X = \varepsilon$.

1.6. Procesamiento de Imágenes

1.6.1. Interpretación matemática de una imagen

La información visual desempeña un papel muy importante en la comunicación humana y en los últimos años se han desarrollado disciplinas tendientes a tratar este tipo de información. En este sentido, el procesamiento de señales bidimensionales se destaca como una disciplina científica capaz

de interactuar con muchas áreas del conocimiento (Landi et al (2017), Dormann et al (2007), Latha et al (2014), Prastawa et al (2004), Quintana et al (2011)).

Diversas áreas se ven vinculadas con el procesamiento de imágenes tales como la medicina, la agricultura, la epidemiología, la meteorología, la geología y la criminalística entre otras (Zielinski et al (2010), Sahu et al (2015), Smith et al (1986), Quintana et al (2011)). En el contexto de la agricultura, por ejemplo, es posible detectar la existencia de plagas, problemas de riego u otras enfermedades vegetales, a partir de los cambios de intensidad que muestra una imagen aérea provocados por la radiación electromagnética (Latha et al (2014)). En biología y geología, es posible realizar estudios, a partir de imágenes satelitales, de las características del suelo y la vegetación que permiten entender las causas y el comportamiento de incendios o inundaciones, y a partir de ellos idear estrategias para la aplicación de prácticas adecuadas de prevención y control (Landi et al (2017), Dormann et al (2007)).

Mediante un complejo proceso físico, un escenario real tridimensional es transformado en un objeto bidimensional llamado “imagen”. Desde el punto de vista de la matemática una imagen I , es el rango de una función $f_I : \mathbb{Z}^2 \rightarrow \mathbb{R}$, observado en algún subconjunto finito W de \mathbb{Z}^2 , donde $f_I(i, j)$ cuantifica algún atributo de interés de I (también llamado intensidad de la imagen I) en el lugar (i, j) de W , tal como nivel de gris, luminosidad, brillo, etc. Habitualmente W es un grilla rectangular de \mathbb{Z}^2 , llamada ventana de observación de I . Como consecuencia, I puede representarse formalmente mediante una sucesión de números reales $\{f_I(i, j)\}_{(i, j) \in W}$ o mediante una matriz numérica llamada imagen digital.

Los errores que se producen en los procesos de captación y cuantificación hasta la obtención de la imagen digital implican la consideración de elementos aleatorios en la representación formal de una imagen; por ello, el marco más adecuado para la construcción de tal representación es el que brindan los procesos estocásticos a través de la Probabilidad y Estadística. Luego, si Ω es el conjunto de todas las imágenes bidimensionales posibles, \mathcal{A} es una σ -álgebra sobre Ω y P una probabilidad sobre (Ω, \mathcal{A}) , una imagen matemática es la observación en alguna ventana W_M de una realización de un proceso aleatorio aleatorio real $Y = \{Y_{i,j}\}_{(i,j) \in \mathbb{Z}^2}$ definido sobre el espacio probabilístico (Ω, \mathcal{A}, P) , tal que $Y_{i,j}(I)$ representa alguna característica de la imagen I en el lugar (i, j) . En adelante, al hablar de imagen nos estaremos refiriendo a la imagen digital; esto es, a realizaciones de procesos aleatorios 2D reales definidos sobre (Ω, \mathcal{A}, P) y observadas en ventanas finitas rectangulares o cuadradas de la red \mathbb{Z}^2 .

1.6.2. Modelos para imágenes

Los modelos para imágenes intentan representar o explicar la intensidad de la imagen a través de un pequeño número de parámetros. Llamaremos x a la verdadera imagen e y a la imagen observada (x es la realización del verdadero proceso aleatorio X , restringida a una ventana W e y es la realización del proceso aleatorio Y correspondiente a la imagen efectivamente captada, observada también en W). Debido a los errores de captación y cuantificación de la imagen, la verdadera imagen no se observa; sino que se observa una distorsión de la misma. Con el objetivo de reconstruir a partir de y la verdadera imagen x , se postula sobre X , Y y el proceso de error algún modelo matemático. Este modelo envuelve una cierta cantidad de parámetros que se desea estimar y supone el cumplimiento de ciertos requisitos llamados “supuestos del modelo”. Estos supuestos apuntan a formalizar lo que el investigador sabe o conjetura sobre la verdadera imagen y su distorsión, y al mismo tiempo están direccionados a conseguir que el modelo propuesto sea manejable desde el punto de vista teórico y computacional. Sin embargo, se entiende que el modelo y sus supuestos formales son simplificaciones

de la realidad y que su validez es, en el mejor de los casos, sólo aproximada. Muchas veces ocurre que la mayoría de los datos pueden ser explicados por un modelo paramétrico explícito y una pequeña proporción de éstos (que no conocemos), no. En este contexto es que surgen los métodos robustos. El enfoque robusto para el modelado estadístico y el análisis de datos tiene como objetivo derivar procedimientos que produzcan estimaciones fiables de los parámetros, no sólo cuando los datos sigan exactamente una distribución hipotética, sino también cuando sigan “sólo aproximadamente” esta distribución.

Entre los modelos cuyas formulaciones bidimensionales han resultado ser exitosas para representar imágenes se destacan la clase de los modelos SAR introducidos por Whittle (1954) y la clase de los modelos CAR estudiados inicialmente por Besag (1974). Estos modelos han demostrado ser de gran importancia en varias áreas que se benefician del procesamiento de imágenes (Smith et al (1986), Dormann et al (2007), Sain and Cressie (2007)). Un caso particular de los modelos SAR, son los modelos AR-2D los cuales permiten representar la intensidad de una imagen por medio de un pequeño número de parámetros. Estos modelos han sido utilizados ampliamente en el estudio y análisis de distintos tipos de imágenes: Ojeda et al (2010), Bustos et al (2009), Alata and Olivier (2003), Sahu et al (2015), Vallejos and Mardesic (2004).

Existen varios métodos de estimación de los parámetros en modelos AR-2D. Kashyap and Eom (1988) propusieron la clase de los estimadores M 2D que presentan un buen desempeño para la estimación paramétrica en modelos AR-2D contaminados innovativamente, pero resultan inadecuados en el caso de contaminación aditiva. Posteriormente, Allende et al (2001) introdujeron los estimadores GM-2D los cuales muestran buen comportamiento con los dos esquemas de contaminación. Luego, Ojeda et al (2002) propusieron los estimadores RA 2D; esta clase de estimadores supera levemente a la clase de los GM 2D, siendo más compleja su implementación computacional. Su principal ventaja sobre los estimadores GM 2D, radica en que se ha demostrado que son consistentes y normalmente asintóticos. Los estimadores mencionados surgieron como generalizaciones de los estimadores planteados para modelos ARMA de series de tiempo.

En este trabajo se propone un nuevo estimador de los parámetros en modelos AR-2D contaminados, para contaminación aditiva y de reemplazo, introduciendo un modelo auxiliar como estrategia para controlar la contaminación. Esta propuesta surge a partir de la idea de extender a dos dimensiones el estimador planteado para modelos ARMA de series de tiempo desarrollado por Muler et al (2009). Se prueba la consistencia y la normalidad asintótica del nuevo estimador.

1.6.3. Contaminación en procesos AR-2D

Maronna et al (2006) describió algunos modelos para valores atípicos de series temporales, incluidos los de tipo aditivo (AO), de reemplazo (RO) y de innovación o innovativo (IO). Para modelos AR-2D, se llevaron a cabo estudios sobre estos procesos considerando modelos bidimensionales de contaminación para el caso aditivo e innovativo (Kashyap and Eom (1988), Allende et al (2001), Ojeda et al (2002)). En esta sección generalizamos la noción de valores atípicos de reemplazo para procesos espaciales.

Definición 1.6.1. Sea $0 < \alpha < 1$ y sea Y un proceso estacionario AR-2D. Diremos que un proceso Z dado por

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

sigue un modelo bidimensional de outliers de reemplazo (RO), donde ξ^α es un proceso de unos y ceros tal que $P(\xi_{i,j}^\alpha = 1) = \alpha$ y $P(\xi_{i,j}^\alpha = 0) = 1 - \alpha$ y W es un proceso de reemplazo que no es

necesariamente independiente de Y .

Definición 1.6.2. Sea $0 < \alpha < 1$ y sea Y un proceso estacionario AR-2D. Diremos que un proceso Z dado por

$$Z_{i,j} = (1 - \xi_{i,j}^\alpha)Y_{i,j} + \xi_{i,j}^\alpha(Y_{i,j} + \nu_{i,j}) \quad (1.6.1)$$

sigue un modelo bidimensional de outliers aditivo (AO), donde ξ^α es un proceso de unos y ceros tal que $P(\xi_{i,j}^\alpha = 1) = \alpha$ y $P(\xi_{i,j}^\alpha = 0) = 1 - \alpha$ y ν es un proceso estacionario independiente de Y .

Observación 1.6.1. .

i) El modelo AO es un caso particular del modelo RO donde

$$W_{i,j} = Y_{i,j} + \nu_{i,j}.$$

ii) Cuando W no sigue la forma de AO, decimos que el proceso de contaminación sigue un modelo de outlier de reemplazo estricto (Strictly Replacement Outlier model o bien SRO).

iii) El tipo de contaminación AO es útil para analizar y modelar distorsiones en imágenes ópticas tal como se describe en Ponomarenko et al (2015).

1.6.4. Expresividad del modelo AR-2D

Muchos de los escenarios reales que se nos presentan a diario son naturalmente ricos en cuanto a textura, nivel de gris, brillo y color; por ejemplo, las imágenes satelitales de campos cultivados y de concentraciones de población, muestran una gran diversidad de matices en lo que a textura se refiere. Lo mismo sucede con las imágenes de regiones geográficas que permiten la confección de mapas y, en general, con casi todas las fotografías de la tierra. Esta diversidad de texturas que se observa en los escenarios reales conduce a que en el estudio y modelización de imágenes se deba recurrir a modelos expresivos, capaces de representar con pocos parámetros la riqueza y diversidad de los escenarios captados por las imágenes. Entre los modelos para imágenes más utilizados por reunir ambas condiciones, se destacan los modelos AR-2D, por lo cual vienen siendo utilizados exitosamente en diferentes contextos para la representación y procesamiento de imágenes digitales (Baran et al (2004), Vallejos and García-Donato (2006), Bustos et al (2009), Quintana et al (2011), Zielinski et al (2010), Yao and Brockwell (2006), Sadabadi et al (2009)). En la siguiente sección intentamos mostrar la expresividad de estos modelos, a partir de la simulación en R de variantes de un modelo AR-2D con tres parámetros:

$$Y_{i,j} = \phi_{1,0}Y_{i-1,j} + \phi_{1,1}Y_{i-1,j-1} + \phi_{0,1}Y_{i,j-1} + \varepsilon_{i,j}, \quad (1.6.2)$$

donde $\varepsilon \sim N(0, 1)$.

- La Figura 1.3-(a) muestra la imagen generada con el modelo 1.6.2 con parámetros $\phi_{1,0} = 0,5$, $\phi_{1,1} = 0$, $\phi_{0,1} = 0,4999$.
- La Figura 1.3-(b) muestra la imagen generada con el modelo 1.6.2 con parámetros $\phi_{1,0} = 0,5$, $\phi_{1,1} = 0$, $\phi_{0,1} = 0,4$.
- La Figura 1.3-(c) muestra la imagen generada con el modelo 1.6.2 con parámetros $\phi_{1,0} = 0,01$, $\phi_{1,1} = 0,8$, $\phi_{0,1} = 0,01$.
- La Figura 1.3-(d) muestra la imagen generada con el modelo 1.6.2 con parámetros $\phi_{1,0} = 0,15$, $\phi_{1,1} = 0,2$, $\phi_{0,1} = 0,17$.

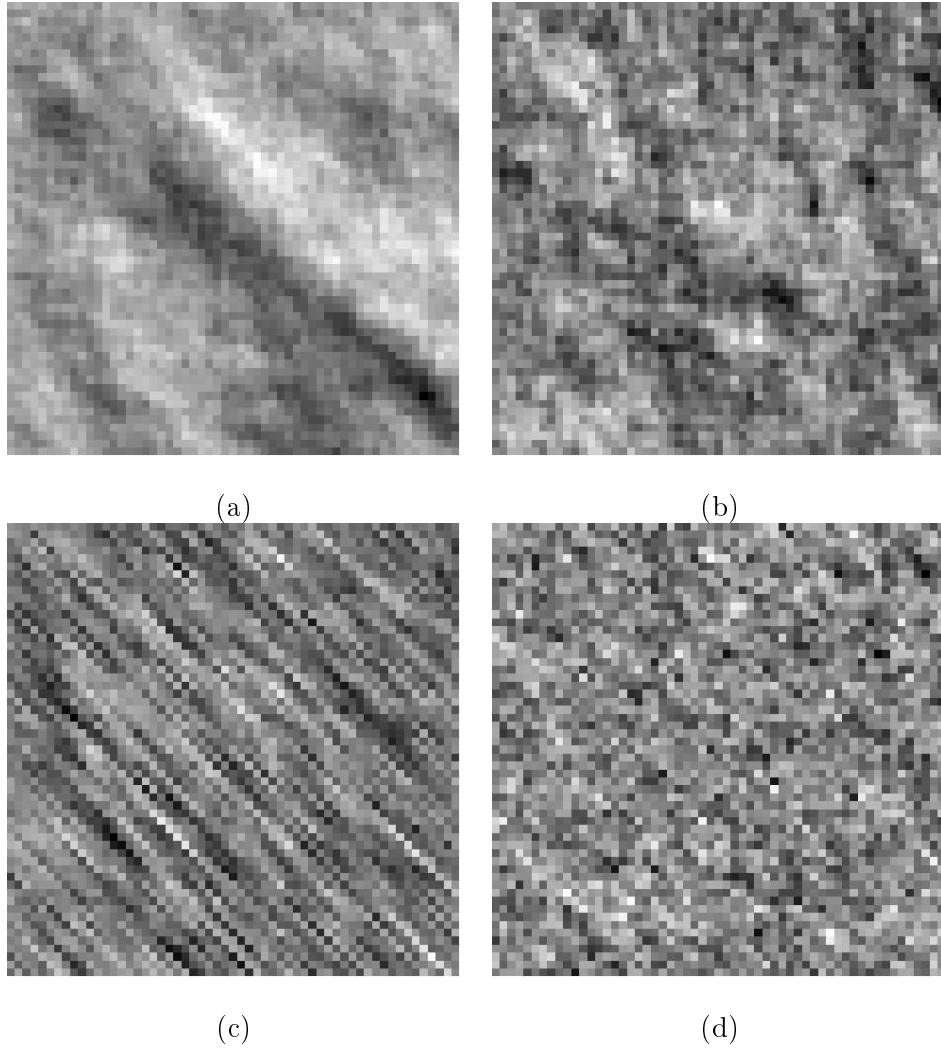


Figura 1.3: Imágenes generadas a partir de procesos AR-2D con tres parámetros. (a) $\phi_{1,0} = 0,5$, $\phi_{1,1} = 0$, $\phi_{0,1} = 0,4999$; (b) $\phi_{1,0} = 0,5$, $\phi_{1,1} = 0$, $\phi_{0,1} = 0,4$; (c) $\phi_{1,0} = 0,01$, $\phi_{1,1} = 0,8$, $\phi_{0,1} = 0,01$; (d) $\phi_{1,0} = 0,15$, $\phi_{1,1} = 0,2$, $\phi_{0,1} = 0,17$.

Capítulo 2

Estimadores BMM para modelos AR Bidimensionales

Debido a que el estimador de mínimos cuadrados de los parámetros de un modelo AR-2D (LS) es muy sensible a la presencia de valores atípicos (Allende et al (1998), Ojeda et al (2002)), se han propuesto varios estimadores alternativos para mitigar el impacto de observaciones contaminadas. Muchas de estas propuestas son extensiones naturales de estimadores robustos estudiados en series de tiempo. Los estimadores robustos conocidos para procesos autorregresivos bidimensionales son los estimadores M, GM y RA, los cuáles han sido estudiados en Kashyap and Eom (1988), Allende et al (1998) y Ojeda et al (2002) respectivamente. La idea central de las alternativas robustas para la estimación paramétrica en el modelo AR 2D es moderar el efecto de las distorsiones que afectan a los datos, controlando los residuos.

Por lo visto en el capítulo 1:

$$I = \{(i, j) \in \mathbb{Z}^2 : i \geq 0, j \geq 0\},$$

$$T = \{(k, l) \in I : k \leq L, l \leq L, (k, l) \neq (0, 0)\},$$

con $L \in \mathbb{N}$, conocido, $L < M$ y

$$W_M = \{(i, j) \in I : i, j \leq M\}.$$

En este trabajo consideramos un proceso Y estacionario AR-2D fuertemente causal definido sobre (Ω, \mathcal{A}, P) y observado en una ventana fuertemente causal W_M . El proceso Y tiene media μ_0 , un proceso de innovación fuertemente causal $\varepsilon = \{\varepsilon_{i,j}\}$ con variables aleatorias $\varepsilon_{i,j}$ i.i.d. con densidad simétrica estrictamente unimodal y polinomio unilateral Φ_0 con soporte en T como en (1.5.2), es decir:

$$Y_{i,j} = \mu_0 + \sum_{(k,l) \in T} \phi_{k,l}^0 (Y_{i-k,j-l} - \mu_0) + \varepsilon_{i,j}, \quad \forall (i, j) \in (W_M \sim T)_L \quad (2.0.1)$$

donde

$$(W_M \sim T)_L = \{(i, j) \in W_M : (i - L, j - L) \in W_M\},$$

o bien de forma compacta:

$$\Phi_0(B_1, B_2)(Y - \mu_0) = \varepsilon.$$

Para no recargar la notación, llamamos $(W_M \sim T)$ a $(W_M \sim T)_L$ asumiendo que el valor de L fue elegido y fijado de antemano.

Como vimos en la sección 1.5.1, Y admite una “Representación de Medias Móviles” de la siguiente manera:

$$\dot{Y}_{i,j} = \sum_{(k,l) \in I} \lambda_{k,l} \varepsilon_{i-k,j-l}, \quad (2.0.2)$$

donde $\lambda_{k,l} \in \mathbb{R} \forall (k,l) \in I$ y $\lambda_{0,0} = 1$.

Los parámetros del modelo (2.0.1) se asumen desconocidos con espacio paramétrico

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

donde

$$\phi = (\phi_{1,0}, \phi_{1,1}, \phi_{0,1}, \phi_{2,0}, \phi_{2,1}, \phi_{2,2}, \phi_{1,2}, \phi_{0,2}, \dots, \phi_{L,0}, \dots, \phi_{L,L}, \phi_{L-1,L}, \dots, \phi_{0,L})$$

de acuerdo al orden espiral \preceq definido sobre I en el capítulo 1 y

$$\mathcal{B}_0 = \{\phi \in \mathbb{R}^{(L+1)^2-1} : \sum_{(k,l) \in T} |\phi_{k,l}| \leq 1 - \epsilon\}$$

para algún $\epsilon > 0$ fijo. Con esta notación $\beta_0 = (\phi_0, \mu_0)$.

En este modelo AR-2D, $\forall \beta \in \mathcal{B}$ se definieron las funciones residuales $\varepsilon_{i,j}(\beta)$ en $(i,j) \in (W_M \sim T)$ como

$$\varepsilon_{i,j}(\beta) = \dot{Y}_{i,j} - \phi' \tilde{Y}_{i,j} \quad (2.0.3)$$

y $\varepsilon_{i,j}(\beta) = 0$ en caso contrario, donde

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

siendo $\dot{Y}_{s,t} = Y_{s,t} - \mu$. La expresión (2.0.3) es equivalente a

$$\varepsilon_{i,j}(\beta) = (Y_{i,j} - \mu) - \sum_{(k,l) \in T} \phi_{k,l} (Y_{i-k,j-l} - \mu) \quad (2.0.4)$$

cuando $(i,j) \in (W_M \sim T)$. Es importante observar que $\varepsilon_{i,j} = \varepsilon_{i,j}(\beta_0) \forall (i,j) \in (W_M \sim T)$.

A partir de aquí nos referimos a este modelo AR-2D (2.0.1) como modelo puro o central.

Como vimos en la sección (1.6.3), podemos modelar un proceso AR-2D contaminado con una fracción α de datos atípicos por:

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

donde Y es un proceso AR-2D puro, W un proceso de reemplazo y ξ^α es un proceso de unos y ceros tal que $P(\xi_{i,j}^\alpha = 1) = \alpha$ y $P(\xi_{i,j}^\alpha = 0) = 1 - \alpha$. La robustez está relacionada con la posibilidad de estimar de manera precisa los parámetros del modelo central Y cuando en realidad observamos un proceso contaminado Z ; para ello, definimos una nueva familia de modelos que ayudan a controlar el efecto de los datos atípicos sobre la estimación de los parámetros.

2.1. Modelo BIP-AR 2D

En esta sección se presenta una nueva clase de modelos acotados no lineales derivados de los modelos AR-2D: los modelos AR-2D de propagación innovativa acotada o por su nombre en inglés “bounded innovation propagation AR 2D model” (BIP-AR 2D). Esta clase de modelos es una generalización bidimensional del modelo presentado para series de tiempo por Muler et al (2009). El modelo BIP-AR 2D surge de la necesidad de estimar los parámetros del modelo AR-2D puro de la mejor manera posible cuando se observa un proceso contaminado, controlando mediante una función acotada los valores atípicos que pueden propagarse en las innovaciones.

Sea Y como en (2.0.1) con proceso de innovaciones ε , $\rho : \mathbb{R} \rightarrow \mathbb{R}$ una función tal que $\rho(0) = 0$, $\rho(x) = \rho(-x)$, $\rho(x)$ es continua, no constante y no decreciente en $|x|$ y b una constante en \mathbb{R} tal que $E(\rho(Z)) = b$ cuando Z tiene densidad simétrica estrictamente unimodal. Se define la M-escala σ de ε como la solución de la ecuación

$$E\left(\rho\left(\frac{\varepsilon_{i,j}}{\sigma}\right)\right) = b. \quad (2.1.1)$$

Luego, para definir nuestro estimador, se utilizó la siguiente familia de modelos auxiliares, llamada familia BIP-AR 2D:

$$\dot{X}_{i,j} = \sum_{(k,l) \in I \setminus \{(0,0)\}} \lambda_{k,l} \sigma \eta\left(\frac{\varepsilon_{i-k,j-l}}{\sigma}\right) + \varepsilon_{i,j}, \quad (2.1.2)$$

donde $\lambda_{k,l}$ son como en (2.0.2) y dependen de β_0 , $\varepsilon_{i,j}$'s son las variables aleatorias del proceso de innovaciones ε , η es una función impar y acotada y σ es la M-escala robusta de ε . Se puede destacar que en el caso en que $\varepsilon \sim N(0, \sigma_\varepsilon^2)$, la M-escala σ coincide con el desvío estándar del proceso ε ; es decir, $\sigma = \sigma_\varepsilon$. Esto se debe a la elección de b de manera tal que $b = E(\rho(Z))$ cuando Z tiene densidad simétrica estrictamente unimodal. Como $\varepsilon_{i,j}/\sigma_\varepsilon \sim N(0, 1)$, $E(\rho(\varepsilon_{i,j}/\sigma_\varepsilon)) = b$; luego, σ_ε es la M-escala de ε correspondiente a b y ρ .

Elegimos η acotada y tal que $\eta(x) = x$ para $|x| \leq k$ para algún $k > 0$. En el modelo (2.1.2), η no modifica los valores de $\varepsilon_{i,j}/\sigma$ cuando los mismos están en el intervalo $[-k, k]$; y los controla o acota si estos superan en valor absoluto a k . En cualquier caso, como $\lambda_{i,j} \rightarrow 0$ cuando $(i, j) \rightarrow \infty$ (según el orden espiral planteado en la sección 1.3), se espera que el efecto de observaciones atípicas en $\varepsilon_{i,j}$ desaparezca en futuras observaciones de ε .

Note que (2.1.2) puede ser escrita como

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

y multiplicando ambos miembros por $\Phi_0(B_1, B_2)$, obtenemos

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

lo cuál es equivalente a

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

En la ventana fuertemente causal W_M , de la expresión anterior, las funciones residuales $\varepsilon_{i,j}^b(\beta, \sigma)$ del modelo BIP-AR 2D pueden ser definidas recursivamente para $(i, j) \in (W_M \sim T)$ por

$$\begin{aligned} \varepsilon_{i,j}^b(\beta, \sigma) &= X_{i,j} - \mu - \sum_{(k,l) \in T} \phi_{(k,l)} (X_{i-k,j-l} - \mu) - \sum_{(k,l) \in T} \phi_{(k,l)} \sigma \eta \left(\frac{\varepsilon_{i-k,j-l}^b(\beta, \sigma)}{\sigma} \right) \\ &+ \sum_{(k,l) \in T} \phi_{(k,l)} \varepsilon_{i-k,j-l}^b(\beta, \sigma), \end{aligned} \quad (2.1.3)$$

y $\varepsilon_{i,j}^b(\beta, \sigma) = 0$ en cualquier otro caso.

2.2. Estimador de Mínimos Cuadrados en el modelo AR 2D

El estimador de mínimos cuadrados $\hat{\beta}_{LS}$ de β_0 en el modelo (2.0.1), se define como:

$$\hat{\beta}_{LS} = \arg \min_{\beta \in \mathcal{B}} \sum_{(i,j) \in (W_M \sim T)} \varepsilon_{i,j}^2(\beta). \quad (2.2.1)$$

Una manera de obtener $\hat{\beta}_{LS}$, es derivar (2.2.1) con respecto a $\phi_{k,l}$, $\forall (k, l) \in T$ y con respecto a μ obteniendo las siguientes ecuaciones:

$$\sum_{(i,j) \in (W_M \sim T)} \left[\varepsilon_{i,j}(\beta) \left(\frac{\partial \varepsilon_{i,j}(\beta)}{\partial \phi_{k,l}} \right) \right] \Big|_{\beta = \hat{\beta}_{LS}} = 0, \quad \forall (k, l) \in T \quad y$$

$$\sum_{(i,j) \in (W_M \sim T)} \left[\varepsilon_{i,j}(\beta) \left(\frac{\partial \varepsilon_{i,j}(\beta)}{\partial \mu} \right) \right] \Big|_{\beta = \hat{\beta}_{LS}} = 0.$$

Luego, como

$$\frac{\partial \varepsilon_{i,j}(\beta)}{\partial \phi_{k,l}} = -\dot{Y}_{i-k,j-l}, \quad \forall (k, l) \in T \quad y \quad (2.2.2)$$

$$\frac{\partial \varepsilon_{i,j}(\beta)}{\partial \mu} = -1 + \sum_{(k,l) \in T} \phi_{k,l} =: \xi, \quad (2.2.3)$$

se tiene que el estimador $\hat{\beta}_{LS}$ se obtiene a partir de las ecuaciones:

$$\sum_{(i,j) \in (W_M \sim T)} \left[\varepsilon_{i,j}(\beta) \dot{Y}_{i-k,j-l} \right] \Big|_{\beta = \hat{\beta}_{LS}} = 0, \quad \forall (k, l) \in T \quad y \quad (2.2.4)$$

$$\sum_{(i,j) \in (W_M \sim T)} [\varepsilon_{i,j}(\beta)\xi]_{\beta=\hat{\beta}_{LS}} = 0. \quad (2.2.5)$$

Como se verá en el capítulo 3 estos estimadores son muy eficientes para estimar los parámetros de un modelo AR-2D puro, no así en el caso de un modelo AR-2D con contaminación. Este hecho ha sido verificado por diversos autores (Kashyap and Eom (1988), Allende et al (1998), Ojeda (1999)).

2.3. Estimadores M, GM y RA en el modelo AR 2D

La alternativa a la estimación LS en modelos AR-2D contaminados la ofrecen los estimadores robustos. En esta sección, se presentan las propuestas conocidas y relacionadas al procesamiento de imágenes.

- Sea (β, σ) el vector de parámetros donde $\beta = (\phi, \mu)$ es el vector de parámetros del modelo AR-2D y σ es la M-escala de $\varepsilon_{i,j}$ como fue definida en (2.1.1). Se define el M-estimador del vector de parámetros (β, σ) por la minimización en $\mathcal{B} \times \mathbb{R}^+$ de la función objetivo:

$$M_N(\beta, \sigma) = \frac{1}{N} \sum_{(i,j) \in W_M \sim T} \rho\left(\frac{\varepsilon_{i,j}(\beta)}{\sigma}\right) \sigma,$$

donde $N = \#(W_M \sim T) = (M - L + 1)^2$, ρ es una función continua, diferenciable, convexa y simétrica con respecto al origen con $\rho(0) = 0$ y además ρ tiene derivada continua y acotada. En este estimador es posible estimar simultáneamente los parámetros β y σ o bien estimar preliminarmente la M-escala σ por algún estimador $\hat{\sigma}$ y luego estimar los parámetros β minimizando la función $M_N(\beta, \hat{\sigma})$.

- Una clase más general de estimadores robustos para el modelo AR 2D definido por (1.5.2), corresponde a los estimadores GM. Estos estimadores asignan convenientemente pesos a las distintas observaciones del proceso. Sea (β, σ) el vector de parámetros donde $\beta = (\phi, \mu)$ es el vector de parámetros del modelo AR-2D y σ es la M-escala de $\varepsilon_{i,j}$ como fue definida en (2.1.1). Similar a la definición del M estimador, se define el estimador GM del vector de parámetros (β, σ) por la minimización en $\mathcal{B} \times \mathbb{R}^+$ de la función objetivo:

$$M_N^G(\beta, \sigma) = \frac{1}{N} \sum_{(i,j) \in W_M \sim T} l_{i,j} t_{i,j} \rho\left(\frac{\varepsilon_{i,j}(\beta)}{l_{i,j} \sigma}\right) \sigma,$$

donde $N = \#(W_M \sim T) = (M - L + 1)^2$, ρ es una función continua, diferenciable, convexa y simétrica con respecto al origen con $\rho(0) = 0$ y además ρ tiene derivada continua, acotada ψ ; $l_{i,j}$ y $t_{i,j}$ son los pesos asignados a la variable $Y_{i,j}$ del proceso Y . Los pesos $l_{i,j}$ y $t_{i,j}$ pueden elegirse de diferentes maneras, determinando en cada caso un estimador GM particular. Según la elección de los pesos, los principales tipos de estimadores GM son los GM tipo Mallows y los GM tipo Schweppe (Vallejos (1998)). Para este estimador es posible estimar simultáneamente los parámetros β y σ o bien estimar preliminarmente la M-escala σ por algún estimador $\hat{\sigma}$ y luego estimar los parámetros β minimizando la función $M_N^G(\beta, \hat{\sigma})$.

Las propiedades asintóticas del estimador GM no han sido aún estudiadas.

- Por último, el estimador robusto RA ha demostrado tener un desempeño aceptable a la hora de estimar los parámetros de un modelo como (1.5.2) y además posee propiedades estadísticas de interés como la consistencia y la normalidad asintótica (Ojeda et al (2002)). Sin embargo la complejidad de la definición de dicho estimador y el alto costo computacional lo vuelven poco atractivo. La idea de este método es robustecer las ecuaciones (2.2.4) y (2.2.5) para obtener un estimador más resistente a perturbaciones en las observaciones.

El estimador RA del vector de parámetros (β, σ) donde $\beta = (\phi, \mu)$ es el vector de parámetros del modelo AR-2D y σ es la M-escala de $\varepsilon_{i,j}$ definida en (2.1.1), se obtiene a partir de las siguientes ecuaciones:

$$\sum_{(s,t) \in D_M(i,j)} \lambda_{s,t} \varphi_{s+i,t+j}(\beta, \sigma) = 0, \quad \forall (i,j) \in T,$$

$$\sum_{(k,l) \in (W_M \sim T)} \psi \left(\frac{\varepsilon_{k,l}(\beta)}{\hat{\sigma}} \right) = 0 \quad y$$

$$\hat{\sigma} = \frac{\text{Med}(|\varepsilon_{k,l}(\hat{\beta}_{LS})| : (k,l) \in (W_M \sim T))}{0,6745}.$$

donde $0,6745 = \text{Med}(|Z|)$ con Z variable aleatoria con distribución $N(0, 1)$, $\hat{\beta}_{LS}$ es el estimador de mínimos cuadrados (usado como estimador inicial), $D_M(i, j) = \cup_{(k,l) \in (W_M \sim T)} \{(s, t) \in I : (k - i - s, l - j - t) \in (W_M \sim T)\}$,

$$\varphi_{u,v}(\beta, \sigma) = \sum_{(k,l) \in (W_M \sim T)} \tau \left(\frac{\varepsilon_{k,l}(\beta)}{\hat{\sigma}}, \frac{\varepsilon_{k-u,l-v}(\beta)}{\hat{\sigma}} \right) \quad \forall (u, v) \in \mathbb{Z}^2$$

con τ función continua, acotada e impar en cada una de sus variables y ψ función continua, acotada e impar. Entre las posibles funciones ψ , se encuentran: ψ de Huber, ψ de Tuckey, ψ de Hampel y ψ de Andrews. En la implementación computacional usamos la función ψ de Huber. Como se puede ver, σ es estimado de manera independiente; otra posibilidad es la estimación simultánea de todos los parámetros. Ahora bien, a los efectos de la implementación computacional del estimador RA, la estimación independiente de σ , simplifica la rutina computacional.

2.4. Estimador BMM 2D para procesos AR-2D

En esta subsección se presenta un nuevo estimador de los parámetros en modelos AR-2D contaminados como otra alternativa a los métodos robustos existentes que sea competitiva con respecto a éstos y muestre propiedades asintóticas deseables. Esta propuesta se basa en realizar una doble M-estimación de los parámetros en el modelo AR-2D (1.5.2) pero con la diferencia de que utilizamos por un lado las funciones residuales del modelo AR-2D (2.0.4) y, por otro lado, las funciones residuales del modelo BIP-AR 2D (2.1.3). La idea principal yace en computar en una primera etapa una estimación robusta de la escala de las innovaciones del modelo central, y en una segunda etapa, usar esta escala estimada para calcular una M estimación de los parámetros en el modelo autorregresivo bidimensional puro. Concretamente, nuestra propuesta calcula, en una primera etapa, M estimaciones de la escala obtenidas considerando las funciones residuales de los modelos AR-2D y BIP-AR 2D, quedándonos con la menor escala estimada. Luego, en una segunda etapa, se obtienen

M estimaciones de los parámetros de autorregresión del modelo AR-2D puro a partir de la escala estimada en la etapa previa y considerando las funciones residuales de los dos modelos. Finalmente, elegimos la mejor estimación. Este estimador es una generalización bidimensional del estimador propuesto para series de tiempo por Muler et al (2009) en el caso de un proceso AR-2D fuertemente causal estacionario.

Una manera de estimar de forma robusta la escala en modelos de regresión fue introducida por Huber (1964). Según esta propuesta, dada una muestra $\mathbf{u} = (u_1, \dots, u_N)$ con $u_i \in \mathbb{R}$, una M-estimación de la escala, $S_N(\mathbf{u})$, es definida por el valor $s \in (0, \infty)$ que satisface

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{u_i}{s}\right) = b, \quad (2.4.1)$$

donde ρ es una función tal que $\rho(0) = 0$, $\rho(x) = \rho(-x)$, $\rho(x)$ es continua, no constante y no decreciente en $|x|$, y $b = E(\rho(Z))$ cuando Z tiene densidad simétrica estrictamente unimodal. Notemos que la M-estimación de la escala s depende de la muestra \mathbf{u} .

Sea Y un proceso AR-2D puro con proceso de innovaciones ε ; y sea $\{y_{i,j}\}$ una sucesión de datos generada por este proceso observada en la ventana fuertemente causal W_M .

Si tenemos la certeza de que los datos no están contaminados y las variables aleatorias del proceso de innovaciones distribuyen como una normal, la mejor manera de estimar los parámetros del modelo AR-2D subyacente es a través del estimador LS. Sin embargo, si sospechamos que los datos $\{y_{i,j}\}_{(i,j) \in W_M}$ pueden estar contaminados, lo más conveniente es recurrir a un método robusto de estimación de los parámetros. Como se vió en la sección 2.3, estos métodos robustos propuestos se basan en minimizar una cierta función conveniente de los residuos y no directamente la suma de los cuadrados de los residuos. Esta función intenta controlar los residuos que pudieran estar contaminados, afectando así la estimación de los parámetros. En esta tesis trabajamos con dos alternativas para expresar los residuos del modelo, (2.0.4) y (2.1.3) y, consecuentemente, dispondremos de dos estrategias de minimización a partir de funciones convenientes de estas dos formas de expresión de los residuos.

Partimos de (2.1.1) intentando estimar σ a través de la estrategia presentada por Huber (2.4.1) utilizando los datos y las dos expresiones del residuo en el modelo AR-2D.

Consideramos la siguiente propiedad sobre funciones $\rho : \mathbb{R} \rightarrow \mathbb{R}$:

P1: $\rho(0) = 0$, $\rho(x) = \rho(-x)$, $\rho(x)$ es continua, acotada, no constante y no decreciente en $|x|$.

A continuación definimos la estimación BMM 2D en el modelo AR-2D siguiendo las etapas que se detallan seguidamente:

Primera Etapa: Se consideraron dos estimaciones de σ : la primera usando los residuos del modelo AR-2D y la otra usando los residuos del modelo BIP-AR 2D. Elegimos finalmente la más pequeña de ellas.

Sea ρ_1 una función que satisface **P1** y tal que $b = E(\rho_1(Z))$ cuando Z tiene densidad simétrica

estrictamente unimodal. Entonces definimos una estimación $\hat{\beta}_S$ de $\beta_0 \in \mathcal{B}$ como

$$\hat{\beta}_S = \arg \min_{\beta \in \mathcal{B}} S_N(\boldsymbol{\varepsilon}_N(\beta)) \quad (2.4.2)$$

y la correspondiente estimación de σ está dada por

$$s_N = S_N(\boldsymbol{\varepsilon}_N(\hat{\beta}_S)), \quad (2.4.3)$$

donde

$$\boldsymbol{\varepsilon}_N(\beta) = (\varepsilon_{M-1,M}(\beta), \varepsilon_{M-1,M-1}(\beta), \varepsilon_{M,M-1}(\beta), \varepsilon_{M-2,M}(\beta), \dots, \varepsilon_{L,L}(\beta), \varepsilon_{L+1,L}(\beta), \dots, \varepsilon_{M,L}(\beta)),$$

con $\varepsilon_{i,j}(\beta) = (y_{i,j} - \mu) - \sum_{(k,l) \in T} \phi_{k,l}(y_{i-k,j-l} - \mu)$ como en (2.0.4) y $S_N(\boldsymbol{\varepsilon}_N(\beta))$ es la M-estimación de la escala basada en ρ_1 , b y la muestra $\boldsymbol{\varepsilon}_N(\beta)$, es decir:

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

como en (2.4.1).

Para obtener la estimación de σ para el modelo BIP-AR, se definió $\hat{\beta}_S^b$ por la minimización de $S_N(\boldsymbol{\varepsilon}_N^b(\beta, \hat{\sigma}(\phi)))$ sobre \mathcal{B} , donde $\hat{\sigma}(\phi)$ es una función con dominio en el espacio paramétrico \mathcal{B}_0 que estima la escala σ para cada ϕ como si este fuera el verdadero valor del vector del parámetro del modelo (2.1.2) y los $\varepsilon_{i,j}$ fueran normales. Para este caso, como la M-escala σ coincide con el desvío estándar de $\varepsilon_{i,j}$, de 2.1.2 se tiene que

$$\sigma^2 = \frac{\sigma_y^2}{1 + \kappa^2 \sum_{(k,l) \in I \setminus \{(0,0)\}} \lambda_{k,l}^2(\phi)}, \quad (2.4.4)$$

donde $\kappa^2 = \text{var}(\eta(\varepsilon_{i,j}/\sigma))$ y $\sigma_y^2 = \text{var}(Y_{i,j})$. Sea $\hat{\sigma}_y^2$ una estimación robusta de σ_y^2 tal que $\hat{\sigma}_y \rightarrow \sigma_y$ a.e. y $\kappa^2 = \text{Var}(\eta(Z))$ donde $Z \sim N(0, 1)$. Entonces definimos la función dependiente de $\phi \in \mathcal{B}$ como

$$\hat{\sigma}^2(\phi) = \frac{\hat{\sigma}_y^2}{1 + \kappa^2 \sum_{(k,l) \in I \setminus \{(0,0)\}} \lambda_{k,l}^2(\phi)}. \quad (2.4.5)$$

Es decir,

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

Luego, la escala estimada s_N^b correspondiente al modelo BIP-AR 2D fue definida por

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

donde, por simplicidad, denotamos $\tilde{\sigma} = \hat{\sigma}(\phi)$ y

$$\boldsymbol{\varepsilon}_N^b(\beta, \tilde{\sigma}) = (\varepsilon_{M-1,L}^b(\beta, \tilde{\sigma}), \varepsilon_{M-1,M-1}^b(\beta, \tilde{\sigma}), \varepsilon_{M,M-1}^b(\beta, \tilde{\sigma}), \dots, \varepsilon_{L,L}^b(\beta, \tilde{\sigma}), \varepsilon_{L+1,L}^b(\beta, \tilde{\sigma}), \dots, \varepsilon_{M,L}^b(\beta, \tilde{\sigma})),$$

$$\text{con } \varepsilon_{i,j}^b(\beta, \tilde{\sigma}) = \dot{y}_{i,j} - \sum_{(k,l) \in T} \phi_{(k,l)} \dot{y}_{i-k,j-l} - \sum_{(k,l) \in T} \phi_{(k,l)} \tilde{\sigma} \eta \left(\frac{\varepsilon_{i-k,j-l}^b(\beta, \tilde{\sigma})}{\tilde{\sigma}} \right) + \sum_{(k,l) \in T} \phi_{(k,l)} \varepsilon_{i-k,j-l}^b(\beta, \tilde{\sigma}),$$

definida como en (2.1.3) y $S_N(\boldsymbol{\varepsilon}_N^b(\beta, \hat{\sigma}(\phi)))$ es la M-estimación de la escala basada en ρ_1 , b y la muestra $\boldsymbol{\varepsilon}_N^b(\beta, \hat{\sigma}(\phi))$, es decir:

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

como en (2.4.1).

Finalmente, nuestra estimación de σ fue

$$s_N^* = \min(s_N, s_N^b).$$

Segunda Etapa: Si bien hemos conseguido estimaciones para β en la primera etapa, nuestro objetivo es mejorar estas estimaciones. Consideremos una función ρ_2 que satisface **P1** y tal que $\rho_2 \leq \rho_1$. Sean M_N y M_N^b funciones definidas sobre \mathcal{B} dadas por:

$$M_N(\beta) = \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*} \right), \quad (2.4.6)$$

y

$$M_N^b(\beta) = \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_2 \left(\frac{\varepsilon_{i,j}^b(\beta, s_N^*)}{s_N^*} \right). \quad (2.4.7)$$

Las correspondientes M estimaciones de los parámetros para cada función fueron:

$$\hat{\beta}_M = \arg \min_{\beta \in \mathcal{B}} M_N(\beta) \quad \text{y} \quad \hat{\beta}_M^b = \arg \min_{\beta \in \mathcal{B}} M_N^b(\beta).$$

Entonces, definimos la estimación BMM 2D como:

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

Observación 2.4.1. .

- 1) Notemos que la definición de la estimación BMM 2D depende fuertemente de la muestra. Si partimos de una muestra aleatoria, en lugar de hablar de estimaciones BMM 2D nos referiremos a estimadores BMM 2D.
- 2) La expresión (2.4.4) se obtuvo de tomar varianza en la ecuación (2.1.2). Como la función η es acotada, los términos $\lambda_{k,l}(\phi)\sigma\eta\left(\frac{\varepsilon_{i-k,j-l}}{\sigma}\right)$ de la suma infinita están acotados por una función integrable. Luego, como los términos $\lambda_{k,l}(\phi)\sigma\eta\left(\frac{\varepsilon_{i-k,j-l}}{\sigma}\right)$ son v.a.i.i.d., por el Teorema de Convergencia Dominada de Lebesgue, la varianza de la suma infinita de esas variables es igual a la suma infinita de las varianzas de cada una de ellas. Entonces, se obtiene que:

$$\begin{aligned} \text{var}(\dot{Y}_{i,j}) &= \text{var}(Y_{i,j}) = \text{var} \left(\sum_{(k,l) \in I \setminus \{(0,0)\}} \lambda_{k,l}(\phi)\sigma\eta\left(\frac{\varepsilon_{i-k,j-l}}{\sigma}\right) + \varepsilon_{i,j} \right) \\ &= \sum_{(k,l) \in I \setminus \{(0,0)\}} \lambda_{k,l}^2(\phi)\sigma^2 \text{var} \left(\eta\left(\frac{\varepsilon_{i-k,j-l}}{\sigma}\right) \right) + \sigma^2 \\ \sigma_y^2 &= \sigma^2 \left(\kappa^2 \sum_{(k,l) \in I \setminus \{(0,0)\}} \lambda_{k,l}^2(\phi) + 1 \right). \end{aligned}$$

- 3) *En las simulaciones computacionales, el valor de κ en la expresión (2.4.5), fue estimado por simulaciones de Monte Carlo de $\text{var}(\eta(Z))$ cuando $Z \sim N(0,1)$.*
- 4) *En el modelo AR-2D fuertemente causal con tres parámetros, los $\lambda_{k,l}$'s tienen una expresión conocida y que puede ser calculada según el trabajo de Basu and Reinsel (1993). Una expresión general de los $\lambda_{k,l}$'s para un modelo con más parámetros se puede obtener calculando los coeficientes de la expansión multinomial de Taylor de $\Phi(B_1, B_2)^{-1}$.*

Capítulo 3

Comportamiento del Estimador BMM

El principal objetivo de este capítulo es analizar el rendimiento del nuevo estimador BMM 2D para estimar los parámetros en un modelo AR-2D comparado con los estimadores LS, M, GM y RA bidimensionales. Para ello se llevaron a cabo varios experimentos Monte Carlo donde se estimaron los parámetros bajo el modelo puro y bajo distintos tipos de distorsiones del mismo usando el estimador BMM 2D y sus competidores. Al final del capítulo se muestran algunas aplicaciones al procesamiento de imágenes.

3.1. El modelo

Se consideró un modelo AR-2D de tres parámetros y media 0, dado por:

$$Y_{i,j} = \phi_{1,0}^0 Y_{i-1,j} + \phi_{1,1}^0 Y_{i-1,j-1} + \phi_{0,1}^0 Y_{i,j-1} + \varepsilon_{i,j}, \quad (3.1.1)$$

con $\phi_{1,0}^0 = 0,15$, $\phi_{1,1}^0 = 0,2$, $\phi_{0,1}^0 = 0,17$ y $\varepsilon = \{\varepsilon_{i,j}\}$ un proceso de ruido blanco donde las variables aleatorias independientes idénticamente distribuidas $\varepsilon_{i,j} \sim N(0, 1)$. El polinomio unilaterial asociado al modelo es:

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

Es importante mencionar que el conjunto de parámetros fue elegido aleatoriamente satisfaciendo la condición $|\phi_{1,0}^0| + |\phi_{1,1}^0| + |\phi_{0,1}^0| < 1$.

El estudio se realizó bajo cinco diferentes condiciones del modelo (Casos); en el Caso I, el modelo no fue contaminado, mientras que en los Casos II, III, IV y V, el modelo fue contaminado como en (1.6.1):

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

- Caso I) Modelo no contaminado como en (3.1.1), donde ε es un proceso tal que $\varepsilon_{i,j} \sim N(0, 1)$ para todo i, j .
- Caso II) Contaminación con outlier de tipo aditivo (AO), donde el proceso ν es independiente del proceso Y y $\nu_{i,j} \sim N(0, 50)$ para todo i, j .

- Caso III) Contaminación de reemplazo estricto (SRO), donde el proceso de reemplazo W sigue una distribución t-Student con 2.3 grados de libertad.
- Caso IV) Contaminación de reemplazo estricto (SRO), donde el proceso de reemplazo W es otro proceso autorregresivo, independiente del proceso Y , con parámetros $\tilde{\phi}_{1,0} = 0,1$, $\tilde{\phi}_{1,1} = 0,3$ y $\tilde{\phi}_{0,1} = 0,2$.
- Caso V) Contaminación de reemplazo estricto (SRO), donde el proceso de reemplazo W es un proceso de ruido blanco con $W_{i,j} \sim N(0, 50)$ para todo i, j .

3.2. Estimación de los parámetros

En cada una de las cinco variantes planteadas, los parámetros $\phi_{1,0}^0$, $\phi_{1,1}^0$ y $\phi_{0,1}^0$ fueron estimados por los cinco procedimientos presentados en el capítulo anterior (LS, M, GM, RA y BMM). En cada experimento, se generaron 500 simulaciones del modelo, y se calcularon el valor medio muestral, el error cuadrático medio (ECM) y la varianza muestral. Para los modelos contaminados se consideraron cuatro niveles de contaminación: 5 %, 10 %, 15 % y 20 %. Además, el estudio se realizó considerando diferentes tamaños de ventana W_M : 8×8 , 16×16 , 32×32 y 57×57 . Se dan a continuación algunos detalles técnicos para el cómputo de la estimación BMM 2D.

Empecemos con la estimación de $\hat{\beta}_S$ y $\hat{\beta}_S^b$. De acuerdo a la definición 2.4.1, podemos escribir $S_N^2(\boldsymbol{\varepsilon}_N(\beta)) = \sum_{(i,j) \in (W_M \sim T)} r_{i,j}^2(\beta)$ donde

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

De esta manera,

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

Luego, para calcular $\hat{\beta}_S$ podemos usar cualquier algoritmo de mínimos cuadrados no lineal; en nuestro caso utilizamos el algoritmo de Levenberg-Marquardt implementado en la función *nls.lm* del paquete *minpack.lm* de R. Este algoritmo interpola entre el algoritmo Gauss-Newton y el método de descenso (Marquardt (1963)).

Similarmente transformamos el problema de minimizar $S_N(\boldsymbol{\varepsilon}_N^b(\beta, \hat{\sigma}(\boldsymbol{\phi})))$ en un problema de mínimos cuadrados no lineales.

Para el cómputo de $\hat{\beta}_M$ y $\hat{\beta}_M^b$ en la segunda etapa, utilizamos nuevamente el algoritmo de Levenberg-Marquardt usando una idea similar a la anterior y tomando como estimación inicial la

mejor estimación calculada en la primera etapa.

En las simulaciones consideramos las siguientes funciones:

$$\rho_2(x) = \begin{cases} 0,5x^2, & \text{si } |x| \leq 2; \\ 0,002x^8 - 0,052x^6 + 0,432x^4 - 0,972x^2 + 1,792, & \text{si } 2 < |x| \leq 3; \\ 3,25, & \text{si } 3 < |x| \end{cases}$$

$\rho_1(x) = \rho_2(\frac{x}{0,405})$ y $\eta = \rho_2'$. La función ρ_1 fue elegida tal que si $b = \text{máx}(\rho_1)/2$, entonces $b = E(\rho_1(Z))$ cuando $Z \sim \mathcal{N}(0, 1)$ y así la escala coincide con el desvío estándar para muestras normales.

Para calcular las M-estimaciones se utilizó la misma función ρ_2 . Además, para la implementación del estimador GM se fijaron los pesos de acuerdo a Allende et al (2001) como:

$$l_{i,j} = 1, \quad \forall i, j \quad \text{y}$$

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

donde ψ_H es la siguiente función de Huber (Kashyap and Eom (1988)):

$$\psi_H(x) = \begin{cases} x, & \text{si } |x| \leq 1,5; \\ 1,5, & \text{si } 1,5 < x; \\ -1,5, & \text{si } x < -1,5. \end{cases}$$

3.3. Experimentos

En el primer experimento, estudiamos la performance del estimador BMM 2D para el modelo no contaminado (Caso I). Todos los métodos estimaron los parámetros bastante bien. La tabla 3.4 muestra los resultados obtenidos para los cuatro diferentes tamaños de ventana considerados. En la Figura 3.1, se muestran los correspondientes boxplots de los residuos para ventanas de tamaño 16×16 , 32×32 y 57×57 . Los gráficos de boxplots para el tamaño de ventana 8×8 fueron omitidos de la figura por no arrojar resultados relevantes como se observa en la tabla 3.4. En este caso, es conveniente usar el método LS debido a su simplicidad y velocidad de cálculo.

El segundo experimento fue desarrollado en el contexto del Caso II. Se analizó la capacidad del método BMM 2D para estimar los parámetros del modelo, considerando un 10% de contaminación aditiva, y tamaños de ventana 8×8 , 16×16 , 32×32 y 57×57 , en comparación con los métodos LS, M, GM y RA. La tabla 3.5 muestra los valores estimados para $\phi_{1,0}^0$, $\phi_{1,1}^0$ y $\phi_{0,1}^0$, usando los cinco diferentes procedimientos analizados. La Figura 3.2 exhibe los correspondientes boxplots de los residuos, omitiendo los boxplots para el caso de ventanas de tamaño 8×8 como en el Caso I. Para tamaños de ventana 32×32 y 57×57 , se puede ver que el estimador BMM es mejor ya que sus valores son más cercanos a los verdaderos valores de los parámetros que las estimaciones producidas por los otros métodos mencionados. Además, el estimador BMM tuvo la menor varianza y el menor ECM. Cuando el tamaño de ventana fue 8×8 o 16×16 , el mejor rendimiento correspondió a los estimadores GM y RA; sin embargo, los valores BMM estimados fueron más parecidos en general a las

estimaciones GM. Una afirmación análoga es válida para la varianza muestral y el ECM del BMM 2D. También se observa que para cualquier tamaño de ventana, el estimador M tuvo la menor varianza muestral pero sus estimaciones fueron erróneas cuando las comparamos con los otros métodos.

El tercer experimento también se refiere al Caso II. Se fijó el tamaño de ventana en 32×32 y se varió el nivel de contaminación aditiva, considerando cuatro niveles: 5%, 10%, 15% y 20%. El método BMM fue el mejor en la mayoría de los casos estudiados, seguido por el estimador RA aunque este último presentó mayor dispersión. Este comportamiento fue deducido de la comparación de los valores estimados por el método BMM con las estimaciones respectivas obtenidas por los otros procedimientos. Los valores de las medidas de dispersión también señalaron que el estimador BMM fue la metodología más precisa. Los resultados pueden verse en la tabla 3.6. Además, de la Figura 3.3 se puede notar que para una ventana de tamaño 32×32 usando cualquiera de los cinco estimadores, el parámetro $\phi_{1,1}^0$ fue estimado con menos precisión que $\phi_{1,0}^0$ y $\phi_{0,1}^0$; mientras que $\phi_{1,1}^0$ fue sobreestimado por todos los métodos, para todos los niveles de contaminación, el estimador RA fue el único método que subestimó $\phi_{1,0}^0$ y $\phi_{0,1}^0$, independientemente del nivel de contaminación.

El cuarto experimento está relacionado al Caso III. El proceso de contaminación fue de reemplazo, donde el proceso W sigue una distribución t-Student con 2.3 g.l. Las simulaciones fueron realizadas para una ventana de tamaño 57×57 . La tabla 3.7 y la figura 3.4 muestran los resultados. Los boxplots exhiben al método BMM como el estimador de mejor desempeño, seguido por los métodos GM, RA, M y LS, en este orden. También se notó que en todos los métodos las estimaciones se deterioran a medida que el nivel de contaminación aumenta, siendo más estable la estimación BMM 2D.

El quinto experimento fue ejecutado en el contexto del Caso IV, donde el proceso de reemplazo W fue un proceso autorregresivo. Se fijó el tamaño de ventana en 32×32 , variando el nivel de contaminación (5%, 10%, 15% y 20%). La tabla 3.8 muestra estos resultados. Además, la Figura 3.5 exhibe los boxplots de los correspondientes residuos. Se puede ver un patrón similar al del cuarto experimento; excepto que en este caso el aumento en el nivel de contaminación deterioró de manera similar a todas las estimaciones.

Finalmente, el sexto experimento fue llevado a cabo de acuerdo al Caso V. El proceso de reemplazo fue un ruido blanco con varianza 50. Como en el quinto experimento, fijamos el tamaño de ventana en 32×32 , variando el nivel de contaminación (5%, 10%, 15% y 20%). La tabla 3.9 muestra los valores estimados obtenidos. Los correspondientes boxplots de los residuos se exhiben en la Figura 3.6. Los valores de los parámetros $\phi_{1,0}^0$ y $\phi_{0,1}^0$ fueron sobreestimados por todos los métodos, excluido el estimador RA que los subestimó. El estimador BMM fue menos afectado por el proceso de contaminación. Los estimadores LS y M fueron menos exactos que los estimadores GM, RA y BMM. Comparativamente, el estimador RA presentó la varianza más alta, mientras que el estimador GM, aunque bastante preciso, se deterioró más que el estimador BMM a medida que aumentó el nivel de contaminación.

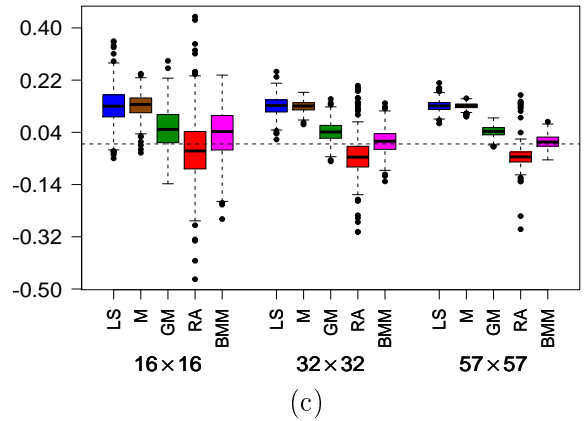
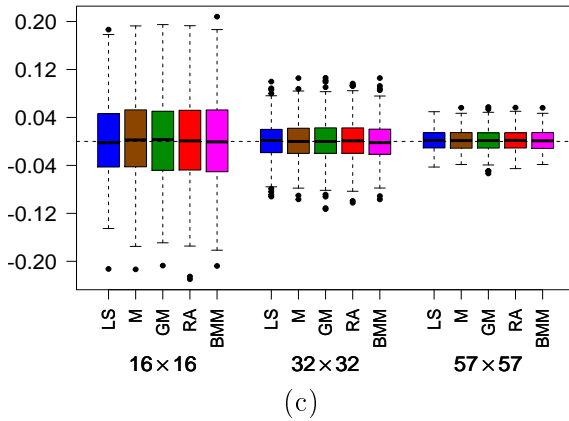
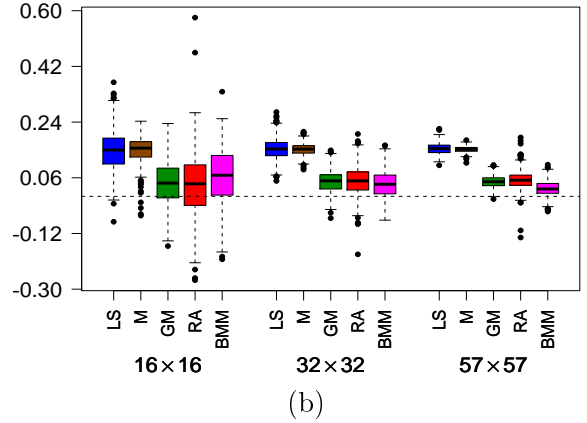
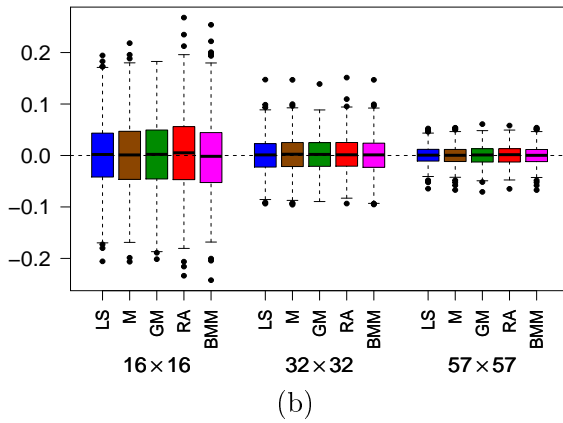
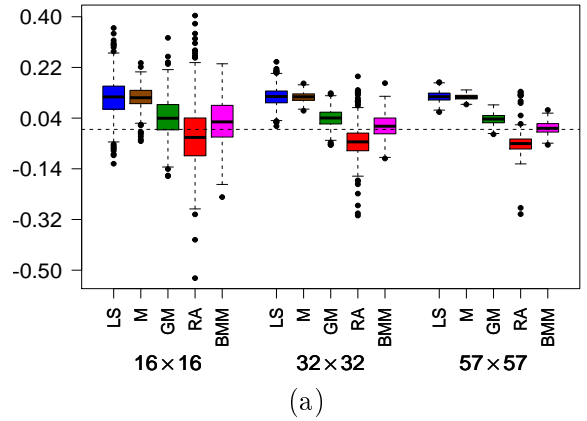
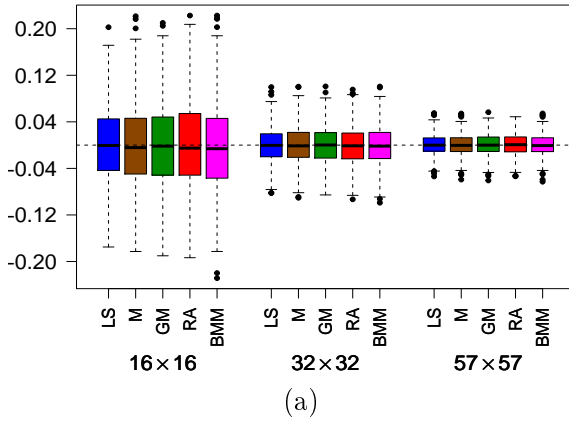


Figura 3.1: Boxplots de las estimaciones de los residuos LS, M, GM, RA y BMM para (a) $\phi_{1,0}^0 = 0,15$, (b) $\phi_{1,1}^0 = 0,2$ y (c) $\phi_{0,1}^0 = 0,17$; modelo (3.1.1) sin contaminación, variando el tamaño de ventana.

Figura 3.2: Boxplots de las estimaciones de los residuos LS, M, GM, RA y BMM para (a) $\phi_{1,0}^0 = 0,15$, (b) $\phi_{1,1}^0 = 0,2$ y (c) $\phi_{0,1}^0 = 0,17$, variando el tamaño de ventana. Modelo (3.1.1) con contaminación aditiva a un nivel del 10% con ruido normal.

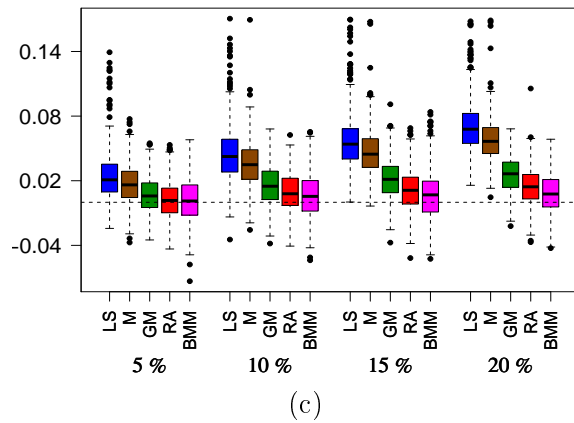
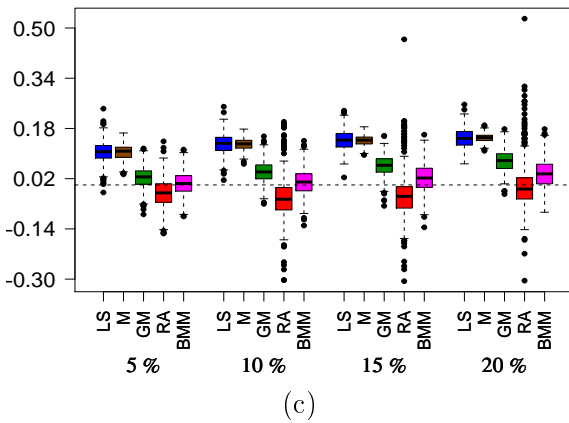
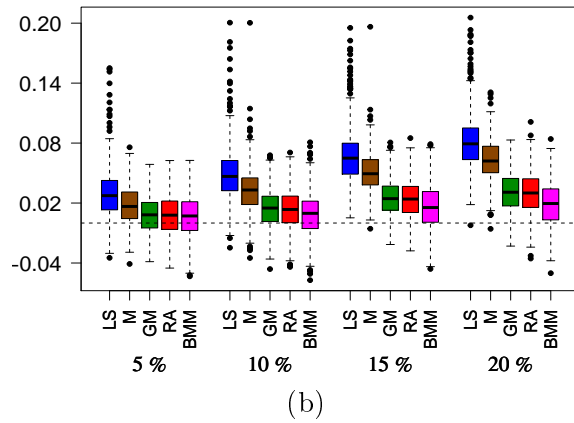
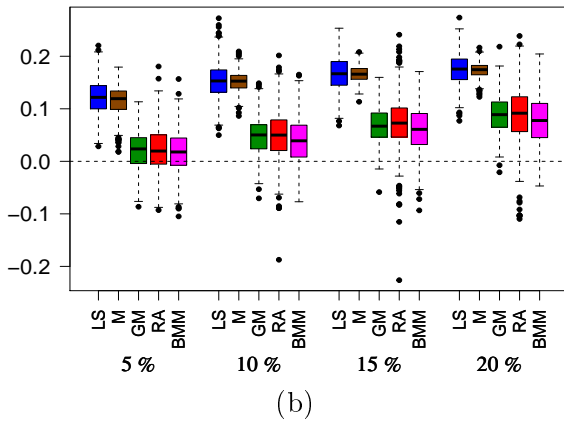
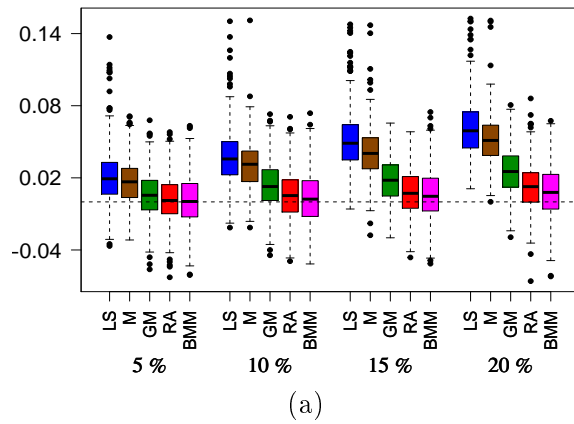
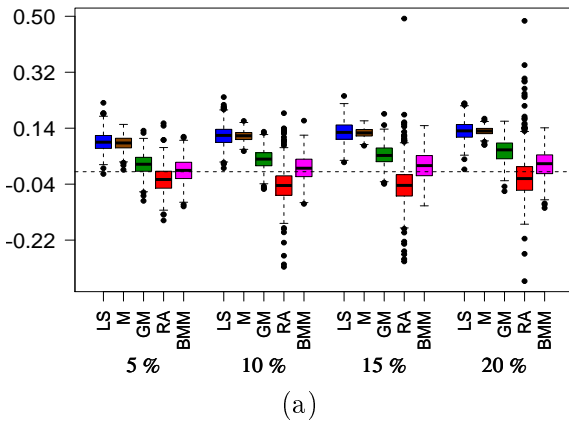


Figura 3.3: Boxplots de las estimaciones de los residuos LS, M, GM, RA y BMM para (a) $\phi_{1,0}^0 = 0,15$, (b) $\phi_{1,1}^0 = 0,2$ y (c) $\phi_{0,1}^0 = 0,17$ en el modelo 3.1.1, con contaminación aditiva, variando el nivel de contaminación con tamaño de ventana 32×32 . El proceso de contaminación es un ruido normal con $\sigma^2 = 50$.

Figura 3.4: Boxplots de las estimaciones de los residuos LS, M, GM, RA y BMM para (a) $\phi_{1,0}^0 = 0,15$, (b) $\phi_{1,1}^0 = 0,2$ y (c) $\phi_{0,1}^0 = 0,17$ en el modelo 3.1.1, con contaminación de reemplazo, variando el nivel de contaminación con tamaño de ventana 57×57 . El proceso de contaminación sigue una distribución t-Student con 2.3 grados de libertad.

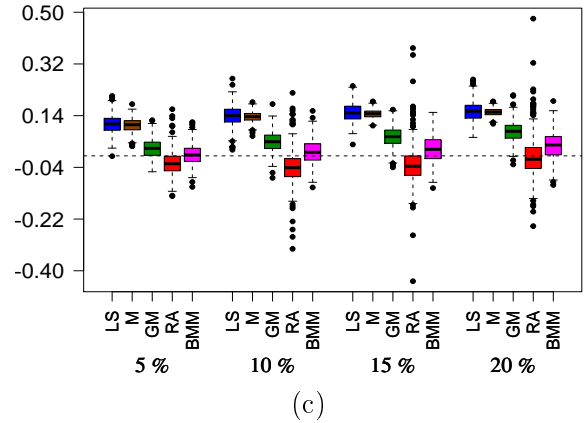
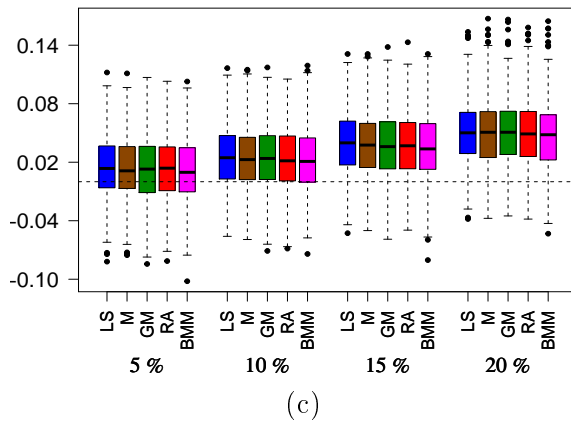
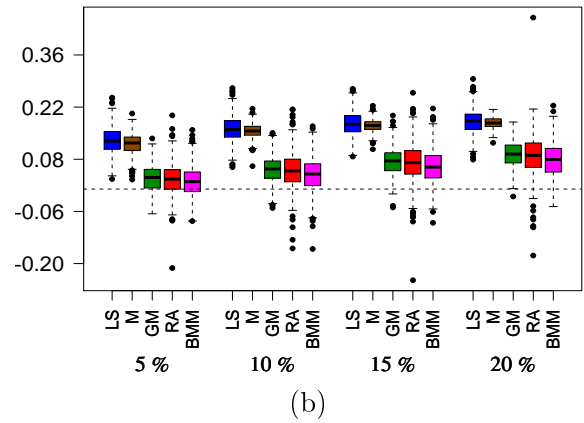
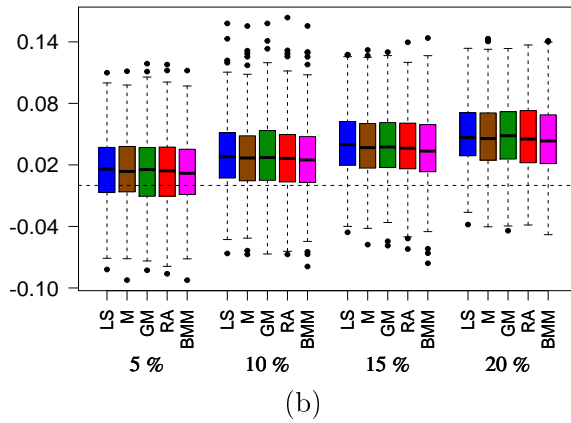
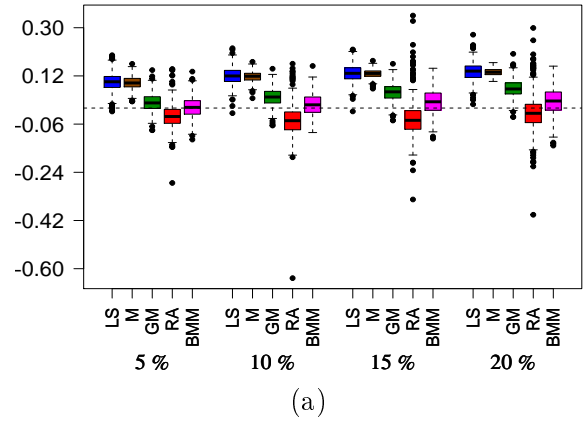
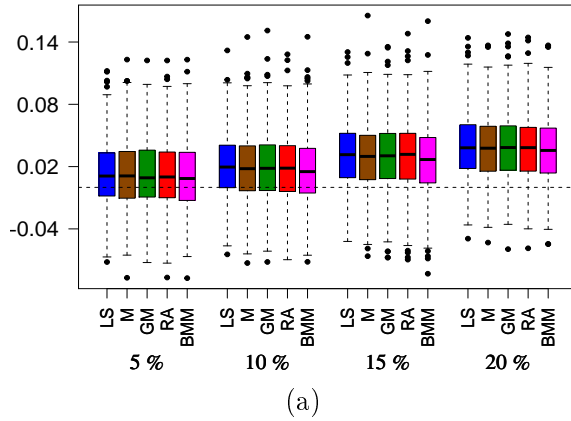


Figura 3.5: Boxplots de las estimaciones de los residuos LS, M, GM, RA y BMM para (a) $\phi_{1,0}^0 = 0,15$, (b) $\phi_{1,1}^0 = 0,2$ y (c) $\phi_{0,1}^0 = 0,17$ en el modelo 3.1.1, variando el nivel de contaminación con ventana de tamaño 32×32 . El proceso de contaminación es de tipo de reemplazo por un proceso AR con parámetros $\tilde{\phi}_{1,0} = 0,1$, $\tilde{\phi}_{1,1} = 0,3$ y $\tilde{\phi}_{0,1} = 0,2$.

Figura 3.6: Boxplots de las estimaciones de los residuos LS, M, GM, RA y BMM para (a) $\phi_{1,0}^0 = 0,15$, (b) $\phi_{1,1}^0 = 0,2$ y (c) $\phi_{0,1}^0 = 0,17$ en el modelo 3.1.1, variando el nivel de contaminación con ventana de tamaño 32×32 . El proceso de contaminación es de tipo de reemplazo con un ruido blanco de varianza 50.

3.3.1. Evaluación del Tiempo Computacional

Todas las rutinas computacionales fueron desarrolladas con el software estadístico R y fueron llevadas a cabo en el servidor JupiterAce de FaMAF-UNC. Este tiene un procesador de 12-cores 2.40GHz Intel Xeon E5-2620v3, con 128 GiB 2133MHz de memoria RAM DDR4. En la figura 3.7 mostramos el tiempo de ejecución dado como el logaritmo del tiempo de una sola simulación para cada estimador y para cada tamaño de ventana en el Caso II. El tiempo fue expresado en segundos. El gráfico muestra que el costo computacional del estimador RA es el más alto; por ejemplo, en una ventana de tamaño 32×32 , el tiempo de ejecución de la estimación RA fue de 43.812 segundos, mientras que el costo computacional para BMM, GM, M y LS fue 2.936, 0.552, 0.516 y 0.436 segundos, respectivamente. Este resultado muestra que, aunque el estimador RA es uno de los mejores competidores del estimador BMM debido a su exactitud y buenas propiedades asintóticas, éste exhibe su costo computacional como una desventaja. Este hecho convierte al estimador RA en un estimador poco atractivo para el procesamiento de imágenes de gran tamaño.

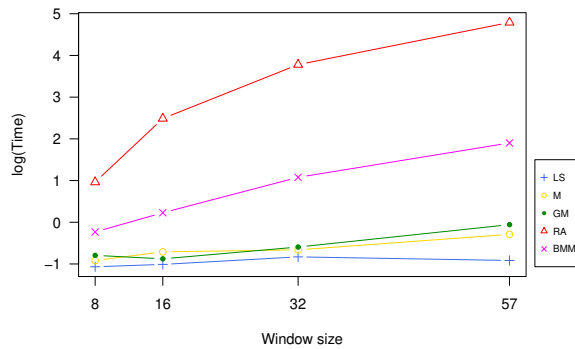


Figura 3.7: Logaritmo del tiempo de estimación (en segundos) cuando el proceso tiene contaminación aditiva de $\sigma^2 = 50$ de acuerdo al tamaño de ventana.

3.4. Aplicación a imágenes reales

El análisis de imágenes contaminadas es de gran interés en varias áreas de investigación (Allende and Galbiati (2004), Bhandari et al (2015), Bustos et al (2009), Gottardo et al (2006), Guyon (1995), Huang and Lee (2006), Kashyap and Eom (1988), Ponomarenko et al (2009), Ponomarenko et al (2015)), siendo reelevante la reducción del ruido que se produce en los procesos de captación física de la imagen y su transmisión electrónica.

En Ojeda et al (2010), se presentaron dos algoritmos para procesamiento de imágenes basados en el modelo unilateral AR-2D con dos parámetros. El primero de los algoritmos produce una aproximación local de las imágenes, y el segundo, es un algoritmo de segmentación. En este trabajo, se propuso utilizar una variante de estos algoritmos usando un proceso AR-2D unilateral con tres parámetros (3.1.1), en lugar de dos. Se llamó a los algoritmos modificados Algoritmo 1 y Algoritmo 2. Estos se aplicaron a la reconstrucción y segmentación de imágenes usando los estimadores de los parámetros LS, GM y BMM 2D en el modelo (3.1.1). Luego, se inspeccionó y comparó el rendimiento de estos estimadores en los Algoritmos 1 y 2 en imágenes contaminadas. Para comparar las

imágenes generadas por los algoritmos y, por lo tanto, el rendimiento de los diferentes estimadores, se calcularon tres índices usados en la literatura; el índice SSIM (Wang and Bovik (2002)), el índice CQ (Ojeda et al (2012)), y el índice CQmax (Ojeda et al (2018)). Estos índices toman valores entre -1 y 1. Cuando el valor del módulo del índice sea cercano a 1 diremos que hay mayor similitud entre las imágenes analizadas. Mientras que si el valor del módulo del índice es cercano a cero diremos que las imágenes son disímiles.

Luego, se presentaron dos experimentos numéricos usando la imagen “Lenna”, la cual fue tomada de la base de datos de imágenes USC-SIPI image database <http://sipi.usc.edu/database/>. En la Figura 3.8-(I), se muestra la imagen original.

A continuación se presentan los Algoritmos 1 y 2. Para más detalles acerca de la notación se puede ver el trabajo Ojeda et al (2010).

Algoritmo 1 Aproximación local de imágenes usando procesos AR-2D.

Require: Imagen original Z .

Ensure: Imagen aproximada \hat{Z} de la imagen original Z

- 1: Defina X como $X = Z - \bar{Z}$
- 2: Genere el bloque $B_X(i_b, j_b)$
- 3: Compute las estimaciones $\hat{\phi}_{1,0}^{(i_b, j_b)}$, $\hat{\phi}_{1,1}^{(i_b, j_b)}$ y $\hat{\phi}_{0,1}^{(i_b, j_b)}$ de $\phi_{1,0}$, $\phi_{1,1}$ y $\phi_{0,1}$ correspondientes al bloque $B_X(i_b, j_b)$ extendido a $B'_X(i_b, j_b) = [X_{r,s}]_{(k-1)(i_b-1) \leq r \leq (k-1)i_b, (k-1)(j_b-1) \leq s \leq (k-1)j_b}$
- 4: Defina \hat{X} en el bloque $B_X(i_b, j_b)$ por

$$\hat{X}_{r,s} = \hat{\phi}_{1,0}^{(i_b, j_b)} X_{r-1,s} + \hat{\phi}_{1,1}^{(i_b, j_b)} X_{r-1,s-1} + \hat{\phi}_{0,1}^{(i_b, j_b)} X_{r,s-1}$$

donde $(k-1)(i_b-1) + 1 \leq r \leq (k-1)i_b$ y $(k-1)(j_b-1) + 1 \leq s \leq (k-1)j_b$

- 5: Defina \hat{Z} como $\hat{Z} = \hat{X} + \bar{Z}$
-

Algoritmo 2 Segmentación

Require: Imagen original Z

Ensure: Imagen segmentada W

- 1: Genere una imagen aproximada \hat{Z} de Z con el Algoritmo 1.
 - 2: Compute la imagen residual W definida como $W = Z - \hat{Z}$.
-

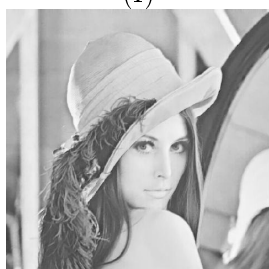
En el primer experimento, se aplicó el Algoritmo 1 para evaluar la capacidad del estimador BMM 2D en la representación de imágenes. Se ajustó localmente un proceso AR-2D a la imagen original para diferentes tamaños de ventana, y se estimaron los parámetros del modelo con el estimador BMM 2D. La Fig. 3.8, (a), (b), (c) y (d) exhibe las imágenes reconstruidas por el estimador BMM 2D usando los tamaños de ventana 8×8 , 16×16 , 32×32 y 57×57 respectivamente. Para todos los tamaños de ventana, las imágenes reconstruidas BMM fueron visualmente buenas; aunque un análisis cuantitativo de la similitud entre cada imagen BMM reconstruida y la imagen original muestra diferencias. Se calcularon los índices SSIM, CQ(1,1) y CQ_{máx}, entre cada imagen reconstruida y la imagen original. Los tres índices revelaron que la similitud decrece cuando el tamaño de ventana aumenta (Tabla 3.1); así, el mejor ajuste fue obtenido con tamaños de ventana pequeños. Este resultado refleja que la estimación BMM 2D en el ajuste local de modelos AR-2D a imágenes es útil para reproducir imágenes digitales. Luego, se aplicó el Algoritmo 2 y se generaron cuatro imágenes diferencia (e), (f), (g) y (h), mostradas en la Fig. 3.8. Se observó que la imagen diferencia (h) resalta más los bordes que las otras. Esto nos dice que cuando se realiza la

reconstrucción con una ventana de tamaño 57×57 , (Fig. 3.8 (d)), una gran cantidad de información se pierde, lo cuál se ve reflejado en la imagen diferencia (Fig. 3.8 (h)).

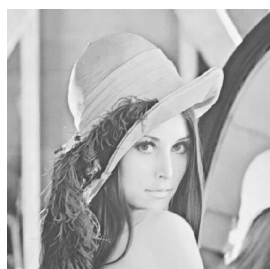


(I)

Figura 3.8: Imagen (I): imagen original Lena. Abajo: La primera fila tiene las reconstrucciones hechas con el estimador BMM ajustadas por tamaños de ventana 8×8 , 16×16 , 32×32 y 57×57 respectivamente ((a) - (d)). La segunda fila tiene las respectivas imágenes diferencia ((e) - (h)) con respecto a la imagen original (I).



(a)



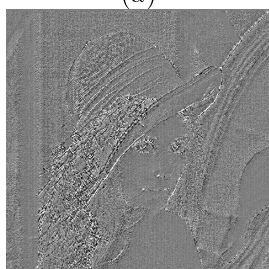
(b)



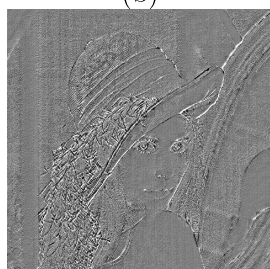
(c)



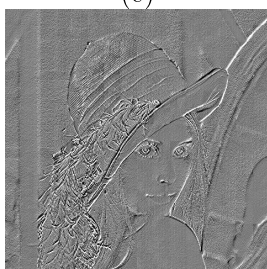
(d)



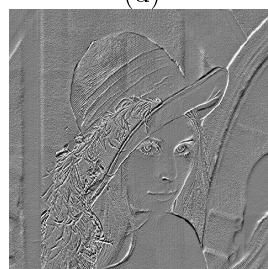
(e)



(f)



(g)



(h)

Cuadro 3.1: Índices SSIM, CQ y CQ_{max} entre la imagen original y cada una de las imágenes reconstruidas BMM (a), (b), (c) y (d) de la Figura 3.8.

Tamaño de ventana	SSIM	CQ(1,1)	$CQ_{m\acute{a}x}$
8×8	0.9948914	0.8582201	0.9706984
16×16	0.9827996	0.8309626	0.9544317
32×32	0.9779204	0.8151581	0.9462133
57×57	0.9762065	0.8073910	0.9423786

En el segundo experimento, la imagen original fue contaminada aditivamente al 10 % (Fig. 3.9 (II)), y se utilizó como entrada en el Algoritmo 1. Se obtuvieron cuatro imágenes reconstruidas usando los estimadores LS, GM y BMM y tamaños de ventana 8×8 y 57×57 . Luego, se aplicó el Algoritmo 2. En las primeras dos columnas de la Figura 3.9, se pueden apreciar los resultados obtenidos considerando ventanas de tamaño 8×8 . Visualmente, no hay grandes diferencias entre las imágenes reconstruidas. Cuando se analiza la tabla 3.2, se verifica este hecho pues los índices calculados son comparables unos con otros siendo levemente más alta la similitud $CQ(1,1)$ y $CQ_{máx}$ para la reconstrucción LS. Por otro lado, la tercera y cuarta columna de la Figura 3.9 muestra los resultados obtenidos por el ajuste de ventanas de tamaño 57×57 . Se observa que la imagen (I), correspondiente a la diferencia entre la imagen restaurada con el método BMM Fig. 3.9-(i) y la imagen contaminada con ruido aditivo (Fig. 3.9-(II)), resalta ligeramente más los bordes.

Cuadro 3.2: Similaridad entre la imagen original y las reconstrucciones de Lena contaminadas aditivamente usando ventanas de tamaño 8×8 (Figura 3.9-(I) vs. figuras 3.9-(a),(b),(c)).

Estimación	SSIM	$CQ(1,1)$	$CQ_{máx}$
LS	0.9836079	0.8416351	0.9588826
GM	0.9390820	0.7821954	0.9103257
BMM	0.9846007	0.8328356	0.9577393

En el tercer experimento, la imagen original fue contaminada aditivamente con $\alpha = 10, 20, 30, 40$ y 50 por ciento. La primera columna de la Fig. 3.10 ((a1) a (a5)) muestra las correspondientes imágenes contaminadas. La segunda y tercera columnas muestran los resultados de la aplicación de los Algoritmos 1 y 2, respectivamente. Los estudios se llevaron a cabo considerando un tamaño de ventana 8×8 con ruido aditivo $\nu \sim N(0, 50)$. La tabla 3.3 muestra la similitud medida entre la imagen original (I) y las reconstrucciones que aparecen en la segunda columna ((b1) a (b5)). Al igual que en el experimento anterior, visualmente, no se observan diferencias importantes con respecto a la imagen original en las imágenes reconstruidas. Sin embargo, se observó que, a medida que aumenta el nivel de contaminación, los bordes de las diferentes imágenes se vuelven más prominentes ((c1) a (c5)). Esto puede ser corroborado por los valores en la Tabla 3.3, donde para todos los índices estudiados, a medida que la contaminación aumenta, la similitud disminuye.

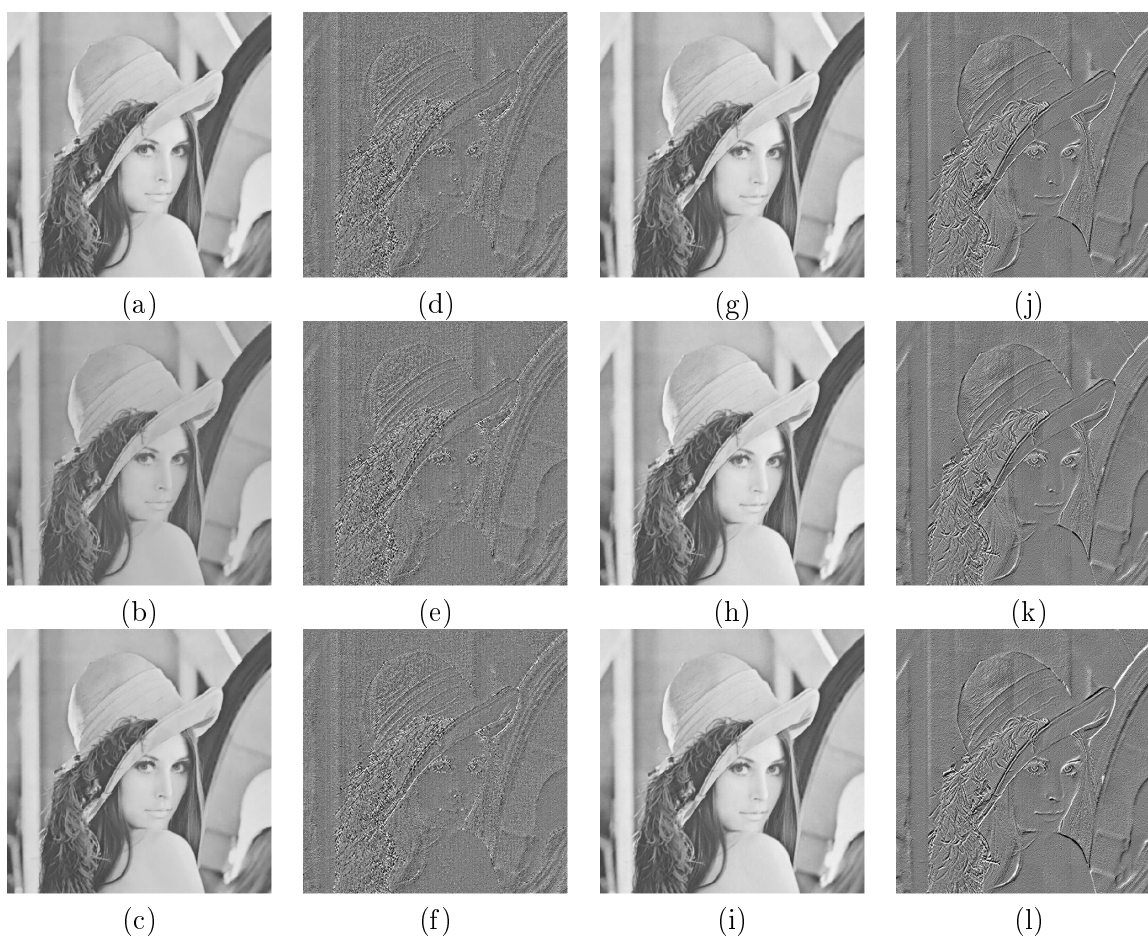
Cuadro 3.3: Similitud entre la imagen original (Fig. 3.10) y las reconstrucciones de Lena (Fig. 3.10, (b1) a (b5)) con contaminación aditiva usando un tamaño de ventana 8×8 para diferentes porcentajes de contaminación ($\nu \sim N(0, 50)$).

	Índice	SSIM	$CQ(1,1)$	$CQ_{máx}$
	10 %	0.9846007	0.8328356	0.9577393
Nivel	20 %	0.9870200	0.8228507	0.9576280
	30 %	0.9793508	0.8045561	0.9480293
Contaminación	40 %	0.9858553	0.8017031	0.9526419
	50 %	0.9745754	0.7828139	0.9398709

Todos los resultados de este capítulo han sido publicados en el trabajo Britos and Ojeda (2018).



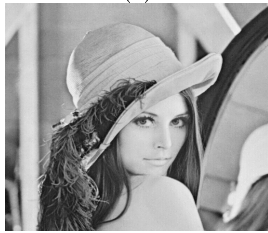
Figura 3.9: Imagen (I): imagen original Lena; Imagen (II): imagen con 10 % de contaminación aditiva y $\tau^2 = 50$. En la primera fila, ajustes hechos LS; en la segunda fila, con GM; y en la tercera fila, con BMM. Columnas 1 y 2 corresponden a ajustes con tamaño de ventana 8×8 , y columnas 3 y 4 con ventanas de tamaño 57×57 . Las columnas 1 y 3 son las reconstrucciones hechas con el Algoritmo 1 y las columnas 2 y 4 son las imágenes segmentadas por el Algoritmo 2.





(I)

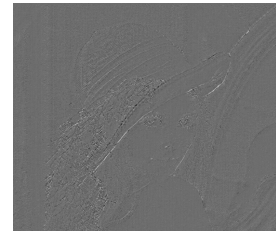
Figura 3.10: Imagen (I): imagen original de Lena. La primera columna muestra las imágenes contaminadas aditivamente al 10, 20, 30, 40 y 50 % respectivamente ($\nu \sim N(0, 50)$). En la segunda columna, se observan sus correspondientes reconstrucciones con el estimador BMM. En la tercera columna se muestran las imágenes diferencia entre (I) y las reconstrucciones.



(a1)



(b1)



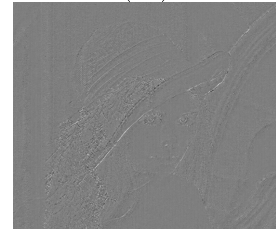
(c1)



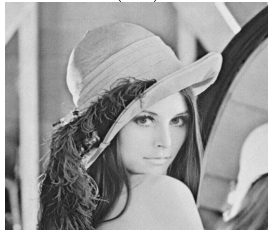
(a2)



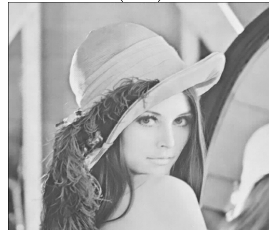
(b2)



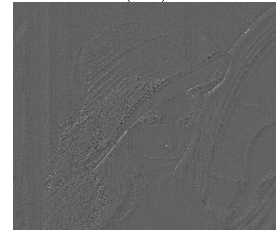
(c2)



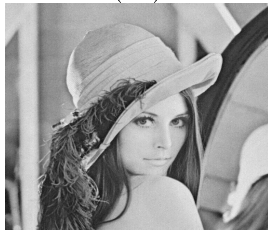
(a3)



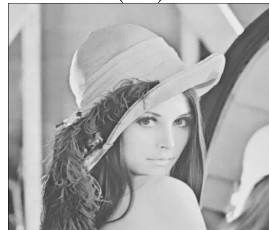
(b3)



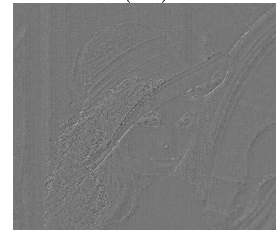
(c3)



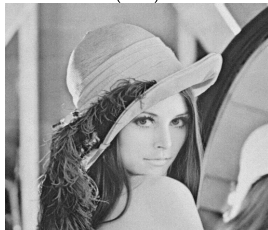
(a4)



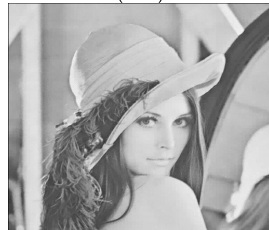
(b4)



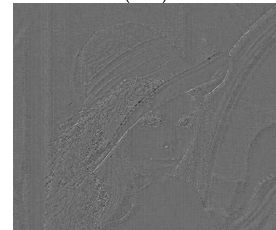
(c4)



(a5)



(b5)



(c5)

Cuadro 3.4: Estimación de $\phi_{1,0} = 0,15$, $\phi_{1,1} = 0,2$ y $\phi_{0,1} = 0,17$ en un modelo AR-2D sin contaminación.

N		$\phi_{1,0}$						$\phi_{1,1}$								
		LS	M	GM	RA	BMM-2D	LS	M	GM	RA	BMM-2D					
8	$\hat{\phi}$	0.14642	0.14638	0.14758	0.14797	0.15260	0.16328	0.16474	0.16422	0.16959	0.16968	0.18854	0.18636	0.18907	0.19012	0.19083
	ecm_{ϕ}	0.01832	0.01955	0.02082	0.02509	0.02789	0.01658	0.01871	0.01914	0.02366	0.02766	0.02046	0.02252	0.02255	0.03273	0.03233
	$\hat{\sigma}_{\phi}^2$	0.01835	0.01957	0.02085	0.02513	0.02794	0.01657	0.01872	0.01915	0.02370	0.02771	0.02037	0.02238	0.02247	0.03270	0.03231
16	$\hat{\phi}$	0.14998	0.15056	0.15081	0.15153	0.15390	0.16794	0.16598	0.16706	0.16771	0.16953	0.19751	0.19826	0.19811	0.19560	0.20218
	ecm_{ϕ}	0.00430	0.00502	0.00498	0.00525	0.00586	0.00408	0.00458	0.00489	0.00497	0.00512	0.00453	0.00485	0.00520	0.00589	0.00537
	$\hat{\sigma}_{\phi}^2$	0.00430	0.00503	0.00499	0.00526	0.00586	0.00408	0.00457	0.00490	0.00497	0.00513	0.00454	0.00485	0.00521	0.00589	0.00537
32	$\hat{\phi}$	0.15014	0.15025	0.14993	0.15053	0.15127	0.16881	0.16884	0.16906	0.16930	0.17019	0.19897	0.19842	0.19851	0.19828	0.19945
	ecm_{ϕ}	0.00099	0.00108	0.00110	0.00113	0.00115	0.00099	0.00106	0.00114	0.00111	0.00108	0.00110	0.00120	0.00123	0.00128	0.00124
	$\hat{\sigma}_{\phi}^2$	0.00099	0.00108	0.00110	0.00113	0.00115	0.00099	0.00106	0.00114	0.00111	0.00109	0.00110	0.00120	0.00123	0.00128	0.00124
57	$\hat{\phi}$	0.14937	0.14937	0.14951	0.14942	0.14956	0.16864	0.16840	0.16845	0.16851	0.16858	0.19934	0.19971	0.19962	0.19944	0.19944
	ecm_{ϕ}	0.00031	0.00032	0.00033	0.00033	0.00033	0.00030	0.00031	0.00032	0.00033	0.00031	0.00031	0.00033	0.00034	0.00035	0.00033
	$\hat{\sigma}_{\phi}^2$	0.00031	0.00032	0.00033	0.00033	0.00033	0.00030	0.00031	0.00032	0.00033	0.00031	0.00031	0.00033	0.00034	0.00035	0.00033

Cuadro 3.5: Estimación de $\phi_{1,0} = 0,15$, $\phi_{1,1} = 0,2$ y $\phi_{0,1} = 0,17$ en el modelo AR-2D con contaminación aditiva de media $\mu = 0$ y $\tau^2 = 50$.

N		$\phi_{1,0}$						$\phi_{1,1}$								
		LS	M	GM	RA	BMM-2D	LS	M	GM	RA	BMM-2D					
8	$\hat{\phi}$	0.04048	0.05018	0.11614	0.12502	0.10983	0.04915	0.04851	0.12014	0.12216	0.11082	0.06266	0.06973	0.15827	0.12456	0.11873
	ecm_{ϕ}	0.03170	0.02101	0.02208	0.06115	0.02689	0.03560	0.02513	0.02582	0.06229	0.02848	0.04177	0.02690	0.02690	0.05767	0.02989
	$\hat{\sigma}_{\phi}^2$	0.01975	0.01107	0.02098	0.06065	0.02533	0.02104	0.01039	0.02338	0.06012	0.02503	0.02296	0.00995	0.02521	0.05208	0.02333
16	$\hat{\phi}$	0.03640	0.03670	0.10622	0.17498	0.12390	0.03880	0.03902	0.11819	0.19017	0.13301	0.05228	0.05043	0.15777	0.16202	0.13817
	ecm_{ϕ}	0.01750	0.01434	0.00669	0.01389	0.00729	0.02094	0.01872	0.00731	0.01249	0.00884	0.02581	0.02413	0.00677	0.01153	0.01208
	$\hat{\sigma}_{\phi}^2$	0.00460	0.00151	0.00478	0.01329	0.00662	0.00374	0.00157	0.00463	0.01211	0.00749	0.00399	0.00176	0.00499	0.01010	0.00827
32	$\hat{\phi}$	0.03426	0.03537	0.11038	0.19341	0.13797	0.03831	0.04006	0.12820	0.21266	0.16113	0.04690	0.04852	0.15217	0.15033	0.16094
	ecm_{ϕ}	0.01446	0.01344	0.00259	0.00522	0.00180	0.01843	0.01721	0.00291	0.00588	0.00180	0.02460	0.02329	0.00352	0.00475	0.00354
	$\hat{\sigma}_{\phi}^2$	0.00107	0.00030	0.00103	0.00335	0.00166	0.00110	0.00033	0.00116	0.00407	0.00173	0.00117	0.00034	0.00124	0.00229	0.00202
57	$\hat{\phi}$	0.03397	0.03567	0.11288	0.20015	0.14520	0.03860	0.03888	0.12653	0.21397	0.16261	0.04609	0.04812	0.15209	0.14793	0.17444
	ecm_{ϕ}	0.01378	0.01316	0.00172	0.00386	0.00049	0.01759	0.01729	0.00222	0.00332	0.00057	0.02401	0.02317	0.00266	0.00357	0.00125
	$\hat{\sigma}_{\phi}^2$	0.00032	0.00009	0.00034	0.00135	0.00047	0.00032	0.00010	0.00033	0.00139	0.00052	0.00032	0.00010	0.00037	0.00086	0.00060

Cuadro 3.6: Estimación de $\phi_{1,0} = 0,15$, $\phi_{1,1} = 0,2$ y $\phi_{0,1} = 0,17$ en el modelo AR-2D con contaminación aditiva y ventana de tamaño 32×32 .

%		$\phi_{1,0}$				$\phi_{0,1}$				$\phi_{1,1}$						
		LS	M	GM	RA	BMM-2D	LS	M	GM	RA	BMM-2D	LS	M	GM	RA	BMM-2D
5%	$\hat{\phi}$	0.05354	0.05835	0.12624	0.17486	0.14533	0.06474	0.06533	0.14652	0.19550	0.16565	0.07732	0.08454	0.17818	0.17824	0.18322
	ecm_{ϕ}	0.01035	0.00893	0.00164	0.00235	0.00139	0.012162	0.01153	0.00173	0.00245	0.00155	0.01613	0.01408	0.00164	0.00209	0.00188
	$\hat{\sigma}_{\phi}^2$	0.00104	0.00053	0.00108	0.00174	0.00137	0.00108	0.00058	0.00119	0.00181	0.00153	0.00108	0.00075	0.00116	0.00162	0.00160
10%	$\hat{\phi}$	0.03426	0.03537	0.11038	0.19341	0.13797	0.03831	0.04006	0.12820	0.21266	0.16113	0.04690	0.04852	0.15217	0.15033	0.16094
	ecm_{ϕ}	0.01446	0.01344	0.00259	0.00522	0.00180	0.01843	0.01721	0.00291	0.00588	0.00180	0.02460	0.02329	0.00352	0.00475	0.00354
	$\hat{\sigma}_{\phi}^2$	0.00107	0.00030	0.00103	0.00335	0.00166	0.00110	0.00033	0.00116	0.00407	0.00173	0.00117	0.00034	0.00124	0.00229	0.00202
15%	$\hat{\phi}$	0.02257	0.02493	0.09578	0.19227	0.13160	0.02793	0.02815	0.10832	0.20706	0.14783	0.03318	0.03889	0.13131	0.12771	0.14057
	ecm_{ϕ}	0.01729	0.01585	0.00397	0.00606	0.00232	0.02115	0.02035	0.00484	0.00628	0.00249	0.02881	0.02780	0.00582	0.00782	0.00553
	$\hat{\sigma}_{\phi}^2$	0.00106	0.00021	0.00103	0.00428	0.00198	0.00096	0.00023	0.00103	0.00492	0.00200	0.00099	0.00021	0.00110	0.00259	0.00200
20%	$\hat{\phi}$	0.01766	0.01886	0.08255	0.16534	0.12508	0.02144	0.02021	0.09349	0.17371	0.13534	0.02511	0.02647	0.11064	0.11122	0.12291
	ecm_{ϕ}	0.01852	0.01737	0.00580	0.00612	0.00273	0.02306	0.02259	0.00704	0.00566	0.00336	0.03155	0.03031	0.00919	0.01082	0.00820
	$\hat{\sigma}_{\phi}^2$	0.00101	0.00017	0.00125	0.00589	0.00211	0.00100	0.00016	0.00118	0.00566	0.00216	0.00096	0.00020	0.00121	0.00295	0.00226

Cuadro 3.7: Estimación de $\phi_{1,0} = 0,15$, $\phi_{1,1} = 0,2$ y $\phi_{0,1} = 0,17$ en un modelo AR-2D con contaminación t-Student con 2.3 g.l.; ventana de tamaño 57.

%		$\phi_{1,0}$				$\phi_{0,1}$				$\phi_{1,1}$						
		LS	M	GM	RA	BMM-2D	LS	M	GM	RA	BMM-2D	LS	M	GM	RA	BMM-2D
5%	$\hat{\phi}$	0.12843	0.13359	0.14418	0.14803	0.14869	0.14606	0.15290	0.16332	0.16762	0.16841	0.17005	0.18190	0.19197	0.19234	0.19414
	ecm_{ϕ}	0.00099	0.00063	0.00038	0.00035	0.00044	0.00113	0.00064	0.00036	0.00033	0.00044	0.00157	0.00070	0.00042	0.00044	0.00047
	$\hat{\sigma}_{\phi}^2$	0.00053	0.00036	0.00035	0.00035	0.00043	0.00055	0.00035	0.00032	0.00033	0.00043	0.00067	0.00037	0.00036	0.00038	0.00044
10%	$\hat{\phi}$	0.11243	0.11936	0.13663	0.14450	0.14653	0.12495	0.13453	0.15399	0.16093	0.16367	0.15040	0.16673	0.18506	0.18625	0.19089
	ecm_{ϕ}	0.00193	0.00133	0.00056	0.00042	0.00048	0.00271	0.00169	0.00060	0.00042	0.00049	0.00322	0.00159	0.00059	0.00057	0.00056
	$\hat{\sigma}_{\phi}^2$	0.00052	0.00039	0.00038	0.00039	0.00047	0.00068	0.00043	0.00035	0.00034	0.00045	0.00076	0.00048	0.00037	0.00038	0.00047
15%	$\hat{\phi}$	0.09909	0.10950	0.13231	0.14231	0.14422	0.11271	0.12404	0.14860	0.15876	0.16274	0.13272	0.14937	0.17524	0.17647	0.18357
	ecm_{ϕ}	0.00318	0.00207	0.00064	0.00041	0.00050	0.00396	0.00257	0.00081	0.00049	0.00058	0.00530	0.00300	0.00096	0.00092	0.00075
	$\hat{\sigma}_{\phi}^2$	0.00059	0.00043	0.00033	0.00035	0.00047	0.00068	0.00046	0.00035	0.00037	0.00053	0.00078	0.00044	0.00035	0.00037	0.00049
20%	$\hat{\phi}$	0.08895	0.09830	0.12530	0.13771	0.14167	0.09862	0.11246	0.14363	0.15533	0.16191	0.11802	0.13740	0.16826	0.16956	0.18116
	ecm_{ϕ}	0.00426	0.00308	0.00095	0.00052	0.00052	0.00556	0.00370	0.00099	0.00053	0.00042	0.00753	0.00434	0.00137	0.00132	0.00087
	$\hat{\sigma}_{\phi}^2$	0.00053	0.00040	0.00034	0.00037	0.00045	0.00061	0.00039	0.00029	0.00031	0.00036	0.00081	0.00042	0.00037	0.00040	0.00052

Cuadro 3.8: Estimación de $\phi_{1,0} = 0,15$, $\phi_{1,1} = 0,2$ y $\phi_{0,1} = 0,17$ en un modelo AR-2D con contaminación de remplazo con proceso de remplazo AR-2D de parámetros $\tilde{\phi}_{1,0} = 0,1$, $\tilde{\phi}_{1,1} = 0,3$ y $\tilde{\phi}_{0,1} = 0,2$; tamaño de ventana 32×32 .

%		$\phi_{1,0}$					$\phi_{1,1}$									
		LS	M	GM	RA	BMM-2D	LS	M	GM	RA	BMM-2D					
5%	$\hat{\phi}$	0.13738	0.13788	0.13763	0.13795	0.13951	0.15559	0.15667	0.15660	0.15707	0.15873	0.18439	0.18481	0.18530	0.18601	0.18667
	ecm_{ϕ}	0.00115	0.00119	0.00120	0.00125	0.00116	0.00123	0.00125	0.00133	0.00128	0.00127	0.00126	0.00131	0.00140	0.00138	0.00128
	$\hat{\sigma}_{\phi}^2$	0.00099	0.00104	0.00105	0.00111	0.00105	0.00102	0.00108	0.00115	0.00112	0.00114	0.00102	0.00108	0.00119	0.00118	0.00111
10%	$\hat{\phi}$	0.12977	0.13134	0.13044	0.13186	0.13395	0.14436	0.14571	0.14477	0.14603	0.14756	0.17055	0.17244	0.17132	0.17310	0.17487
	ecm_{ϕ}	0.00136	0.00140	0.00145	0.00139	0.00136	0.00166	0.00165	0.00173	0.00170	0.00160	0.00192	0.00191	0.00199	0.00192	0.00187
	$\hat{\sigma}_{\phi}^2$	0.00096	0.00105	0.00107	0.00106	0.00111	0.00100	0.00107	0.00109	0.00113	0.00110	0.00106	0.00116	0.00117	0.00120	0.00124
15%	$\hat{\phi}$	0.11879	0.12068	0.11975	0.12008	0.12333	0.13053	0.13242	0.13282	0.13331	0.13515	0.15953	0.16200	0.16142	0.16245	0.16436
	ecm_{ϕ}	0.00197	0.00194	0.00203	0.00200	0.00188	0.00256	0.00249	0.00251	0.00247	0.00234	0.00255	0.00245	0.00249	0.00250	0.00236
	$\hat{\sigma}_{\phi}^2$	0.00100	0.00108	0.00112	0.00111	0.00117	0.00100	0.00108	0.00113	0.00113	0.00113	0.00091	0.00101	0.00100	0.00109	0.00110
20%	$\hat{\phi}$	0.11002	0.11141	0.11104	0.11190	0.11384	0.11989	0.12046	0.11980	0.12060	0.12338	0.15032	0.15283	0.15146	0.15289	0.15556
	ecm_{ϕ}	0.00257	0.00251	0.00253	0.00250	0.00238	0.00356	0.00362	0.00369	0.00364	0.00344	0.00344	0.00337	0.00348	0.00344	0.00317
	$\hat{\sigma}_{\phi}^2$	0.00097	0.00102	0.00101	0.00105	0.00107	0.00105	0.00117	0.00117	0.00120	0.00127	0.00097	0.00115	0.00113	0.00123	0.00120

Cuadro 3.9: Estimación de $\phi_{1,0} = 0,15$, $\phi_{1,1} = 0,2$ y $\phi_{0,1} = 0,17$ en un modelo AR-2D con contaminación de remplazo con proceso de remplazo ruido blanco de varianza 50; ventana de tamaño 32.

%		$\phi_{1,0}$					$\phi_{1,1}$									
		LS	M	GM	RA	BMM-2D	LS	M	GM	RA	BMM-2D					
5%	$\hat{\phi}$	0.05141	0.05608	0.12961	0.18022	0.14770	0.05960	0.06437	0.14453	0.19548	0.16614	0.06973	0.07972	0.17067	0.17307	0.17888
	ecm_{ϕ}	0.01085	0.00936	0.00152	0.00282	0.00141	0.01315	0.01174	0.00167	0.00230	0.00131	0.01828	0.01528	0.00203	0.00249	0.00205
	$\hat{\sigma}_{\phi}^2$	0.00113	0.00054	0.00110	0.00191	0.00141	0.00097	0.00058	0.00103	0.00166	0.00130	0.00132	0.00082	0.00117	0.00177	0.00161
10%	$\hat{\phi}$	0.02990	0.03274	0.10976	0.19717	0.13843	0.03144	0.03431	0.12180	0.20953	0.15686	0.03895	0.04424	0.14775	0.14994	0.16107
	ecm_{ϕ}	0.01550	0.01409	0.00274	0.00560	0.00178	0.02037	0.01871	0.00349	0.00467	0.00182	0.02701	0.02462	0.00392	0.00485	0.00347
	$\hat{\sigma}_{\phi}^2$	0.00108	0.00034	0.00112	0.00338	0.00165	0.00117	0.00030	0.00116	0.00311	0.00165	0.00107	0.00036	0.00119	0.00235	0.00196
15%	$\hat{\phi}$	0.01994	0.02190	0.09124	0.19055	0.12807	0.01905	0.02397	0.10452	0.20163	0.14848	0.02439	0.02998	0.12630	0.12875	0.14057
	ecm_{ϕ}	0.01796	0.01666	0.00453	0.00617	0.00237	0.02382	0.02154	0.00540	0.00536	0.00248	0.03185	0.02916	0.00673	0.00769	0.00576
	$\hat{\sigma}_{\phi}^2$	0.00104	0.00025	0.00108	0.00454	0.00190	0.00103	0.00022	0.00112	0.00436	0.00202	0.00101	0.00025	0.00130	0.00262	0.00224
20%	$\hat{\phi}$	0.01369	0.01590	0.07760	0.16684	0.12467	0.01536	0.01802	0.08600	0.17541	0.13500	0.01779	0.02225	0.10617	0.10972	0.12265
	ecm_{ϕ}	0.01970	0.01816	0.00627	0.00464	0.00308	0.02498	0.02326	0.00824	0.00463	0.00345	0.03430	0.03180	0.00991	0.01091	0.00807
	$\hat{\sigma}_{\phi}^2$	0.00112	0.00018	0.00103	0.00436	0.00245	0.00106	0.00016	0.00119	0.00461	0.00223	0.00110	0.00021	0.00111	0.00277	0.00209

Capítulo 4

Propiedades Teóricas del Estimador BMM Bidimensional

4.1. Preliminares

Los principales resultados de este capítulo se encuentran en el Teorema 4.3.4 y el Teorema 4.4.2 los cuales demuestran la consistencia y la normalidad asintótica (respectivamente) del estimador BMM 2D para procesos AR-2D puro. Para demostrar el Teorema 4.3.4 probamos primero la consistencia del estimador $\hat{\beta}_S$ y luego la consistencia del estimador $\hat{\beta}_M$ (Teoremas 4.3.1 y 4.3.3, respectivamente). El Teorema 4.3.2 relaciona las propiedades del estimador $\hat{\beta}_S$ con el estimador BMM 2D, $\hat{\beta}_M^*$. Por otra parte, para demostrar el Teorema 4.4.2 necesitaremos probar antes la normalidad asintótica del estimador $\hat{\beta}_M$ (Teorema 4.4.1).

Estos resultados dependen de varios lemas que serán enunciados en este capítulo y cuyas demostraciones se encuentran en el Apéndice. Las estrategias de prueba están inspiradas en algunas ideas presentadas en Muler et al (2009) aportando en este trabajo demostraciones minuciosas de las afirmaciones para la versión bidimensional.

En los siguientes cuadros se presentan sendos esquemas que muestran el orden en que se realizan las demostraciones de la consistencia y normalidad asintótica del estimador BMM para estimar los parámetros en el modelo AR-2D.

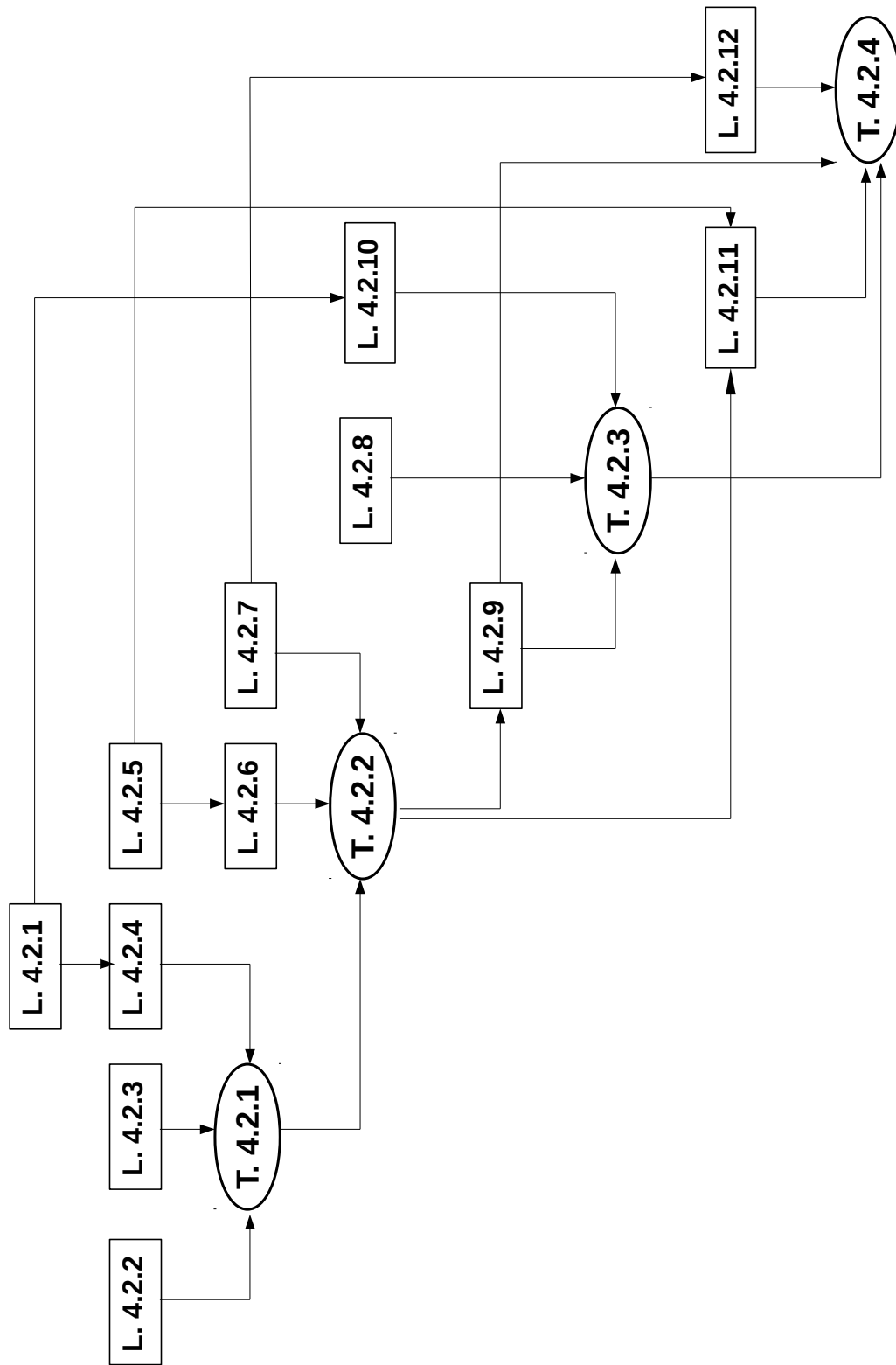


Figura 4.1: Esquema de demostración de la consistencia del estimador BMM 2D

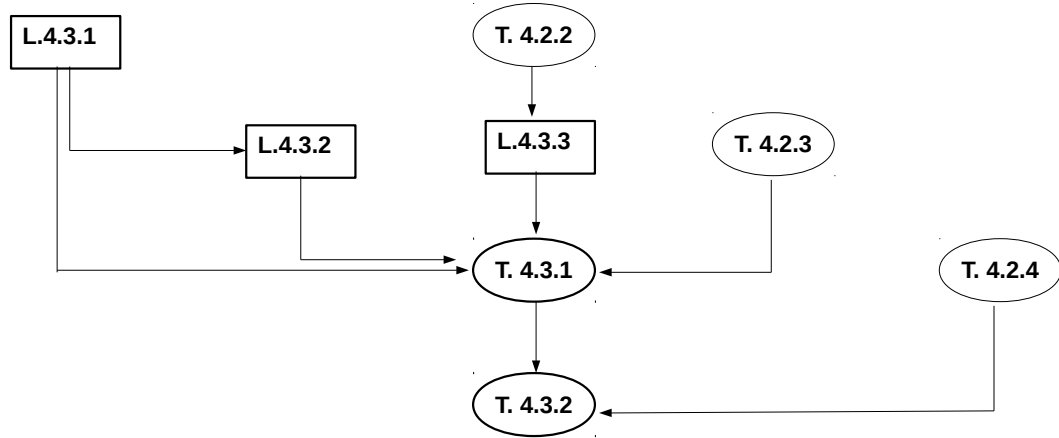


Figura 4.2: Esquema de demostración de la normalidad asintótica del estimador BMM 2D

Asumiremos que se cumplen las siguientes condiciones:

- P1** $\rho(0) = 0$, $\rho(x) = \rho(-x)$, $\rho(x)$ es continua, acotada, no constante y no decreciente en $|x|$.
- P2** El proceso $Y = \{Y_{i,j}\}_{(i,j) \in \mathbb{Z}^2}$ es un proceso AR-2D estacionario (en el sentido estricto) y ergódico definido sobre (Ω, \mathcal{A}, P) con parámetros $\beta_0 = (\phi_0, \mu_0) \in \mathcal{B}$ y proceso de innovaciones $\varepsilon = \{\varepsilon_{i,j}\}_{(i,j) \in \mathbb{Z}^2}$.
- P3** Las variables aleatorias $\varepsilon_{i,j}$, del proceso de innovación ε , tienen una distribución absolutamente continua con una densidad simétrica y estrictamente unimodal.
- P4** $P(\varepsilon_{i,j} \in C) < 1$ para cualquier compacto C .
- P5** La función η es continua, impar y acotada.

4.2. Resultados que se desprenden del estimador BMM 2D

Debido a que las derivadas de primer orden de $\varepsilon_{i,j}(\beta)$ son como en (2.2.2) y (2.2.3), se obtuvo que sus derivadas de segundo orden son:

$$\frac{\partial^2 \varepsilon_{i,j}(\beta)}{\partial \phi_{m,n} \partial \phi_{k,l}} = 0, \quad \forall (k,l), (m,n) \in T, \quad (4.2.1)$$

$$\frac{\partial^2 \varepsilon_{i,j}(\beta)}{\partial \phi_{k,l} \partial \mu} = 1, \quad \forall (k,l) \in T \quad y \quad (4.2.2)$$

$$\frac{\partial^2 \varepsilon_{i,j}(\beta)}{\partial^2 \mu} = 0. \quad (4.2.3)$$

Definimos el vector $\nabla(\varepsilon_{i,j}(\beta))$ que contiene a las derivadas de primer orden de $\varepsilon_{i,j}(\beta)$ como:

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

La siguiente definición nos permitire obtener algunos resultados que necesitaremos para probar los teoremas de este capítulo.

Definición 4.2.1. *Dada la función ρ_1 que satisface **P1**, se define $s(\beta)$ como la función $s : \mathcal{B} \rightarrow \mathbb{R}$ dada por*

$$E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)} \right) \right) = b, \quad (4.2.4)$$

donde b es tal que $b = E(\rho_1(Z))$ cuando Z tiene densidad simétrica y estrictamente unimodal.

Denotaremos s_0 como el valor real positivo tal que $s_0 = s(\beta_0)$.

Dada la función ρ_2 que surge en la segunda etapa de la definición de la estimación BMM 2D, se define $\psi_2 := \rho_2'$ y se obtuvieron las siguientes propiedades:

Propiedad 1.

$$\nabla \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \right) = \frac{1}{s_0} \psi_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \cdot \nabla(\varepsilon_{i,j}(\beta)). \quad (4.2.5)$$

Esta propiedad se debe a que:

$$\frac{\partial}{\partial \phi_{k,l}} \rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) = -\frac{1}{s_0} \psi_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \dot{Y}_{i-k,j-l}, \quad \forall (k,l) \in T \quad (4.2.6)$$

por (2.2.2) y debido a que

$$\frac{\partial}{\partial \mu} \rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) = \frac{1}{s_0} \psi_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \xi \quad (4.2.7)$$

por (2.2.3) donde $\xi = -1 + \sum_{(k,l) \in T} \phi_{k,l}$.

Propiedad 2.

$$\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \right) = \frac{1}{s_0^2} \psi_2' \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \cdot \nabla (\varepsilon_{i,j}(\beta)) \nabla (\varepsilon_{i,j}(\beta))^t + \frac{1}{s_0} \psi_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \cdot \nabla^2 (\varepsilon_{i,j}(\beta)). \quad (4.2.8)$$

Esta propiedad resulta de los hechos

$$\frac{\partial^2}{\partial \phi_{m,n} \partial \phi_{k,l}} \rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) = \frac{1}{s_0^2} \psi_2' \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \dot{Y}_{i-k,j-l} \dot{Y}_{i-m,j-n}, \quad \forall (k,l), (m,n) \in T, \quad (4.2.9)$$

$$\frac{\partial^2}{\partial \mu \partial \phi_{k,l}} \rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) = -\frac{\xi}{s_0^2} \psi_2' \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \dot{Y}_{i-k,j-l} + \frac{1}{s_0} \psi_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right), \quad \forall (k,l) \in T \quad (4.2.10)$$

y

$$\frac{\partial^2}{\partial \mu^2} \rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) = \frac{\xi^2}{s_0^2} \psi_2' \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \quad (4.2.11)$$

de acuerdo a lo visto en la propiedad (1).

Se tienen, además, los siguientes resultados:

Propiedad 3. .

a)

$$E \left(\psi_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) = 0. \quad (4.2.12)$$

b)

$$E \left[\nabla \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \right) \right] = E \left(\frac{1}{s_0} \psi_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \right) \cdot E [\nabla (\varepsilon_{i,j}(\beta))] = 0. \quad (4.2.13)$$

La propiedad 3-(a) se debe a que ψ_2 es impar y la distribución de $\varepsilon_{i,j}$ es simétrica. La propiedad 3-(b) resulta de (4.2.5) y (4.2.12) y del hecho de que $\nabla (\varepsilon_{i,j}(\beta))$ es independiente de $\psi_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right)$.

Sea

$$V_0 = E \left[\nabla \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \cdot \nabla \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right)^t \right]. \quad (4.2.14)$$

Propiedad 4.

$$V_0 = E \left[\frac{1}{s_0^2} \psi_2^2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \cdot \nabla (\varepsilon_{i,j}(\beta)) \cdot \nabla (\varepsilon_{i,j}(\beta))^t \right] = E \left(\frac{1}{s_0^2} \psi_2^2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \right) \cdot E [\nabla (\varepsilon_{i,j}(\beta)) \nabla (\varepsilon_{i,j}(\beta))^t] \quad (4.2.15)$$

donde

$$\begin{aligned}
 E [\nabla (\varepsilon_{i,j}(\beta)) \nabla (\varepsilon_{i,j}(\beta))^t] &= E \begin{pmatrix} (\dot{Y}_{i-s,j-t} \dot{Y}_{i-m,j-n})_{(s,t),(m,n) \in T} & (-\xi \dot{Y}_{i-s,j-t})_{(s,t) \in T} \\ (-\xi \dot{Y}_{i-s,j-t})_{(s,t) \in T} & \xi^2 \end{pmatrix} \\
 &= \begin{pmatrix} (E(\dot{Y}_{i-s,j-t} \dot{Y}_{i-m,j-n}))_{(s,t),(m,n) \in T} & 0 \\ 0 & \xi^2 \end{pmatrix} \\
 &= \begin{pmatrix} \tilde{C} & 0_{L \times 1} \\ 0_{1 \times L} & \xi^2 \end{pmatrix} \tag{4.2.16}
 \end{aligned}$$

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

Esta última propiedad se obtuvo por (4.2.5) y la independencia de $\nabla (\varepsilon_{i,j}(\beta))$ de $\psi_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right)$.

Cabe observar que

$$\begin{aligned}
 E(\dot{Y}_{i-s,j-t} \dot{Y}_{i-m,j-n}) &= E((\Phi_0(B_1, B_2)^{-1} \varepsilon_{i-s,j-t})(\Phi_0(B_1, B_2)^{-1} \varepsilon_{i-m,j-n})) \\
 &= E\left(\sum_{(k,l) \in I} \lambda_{k,l} \varepsilon_{i-s-k,j-t-l} \sum_{(q,r) \in I} \lambda_{q,r} \varepsilon_{i-m-q,j-n-r} \right) \\
 &= E\left(\sum_{(k,l) \in I} \sum_{(q,r) \in I} \lambda_{k,l} \lambda_{q,r} \varepsilon_{i-s-k,j-t-l} \varepsilon_{i-m-q,j-n-r} \right) \\
 &= \sum_{(k,l) \in I} \sum_{(q,r) \in I} \lambda_{k,l} \lambda_{q,r} E(\varepsilon_{i-s-k,j-t-l} \varepsilon_{i-m-q,j-n-r}). \tag{4.2.17}
 \end{aligned}$$

Como las variables $\varepsilon_{i,j}$ son i.i.d. con media 0, en la suma (4.2.17) sólo sobrevivirán los términos con igual subíndice, es decir, $q = k + s - m$ y $r = l + t - n$. Luego, (4.2.17) nos queda:

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

Por lo tanto, \tilde{C} nos queda

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

donde

$$C = \begin{pmatrix} \sum_{(k,l) \in I} \lambda_{k,l}^2 & \sum_{(k,l) \in I} \lambda_{k,l} \lambda_{k,l-1} & \sum_{(k,l) \in I} \lambda_{k,l} \lambda_{k+1,l-1} & \cdots & \sum_{(k,l) \in I} \lambda_{k,l} \lambda_{k+1,l-L} \\ \sum_{(k,l) \in I} \lambda_{k,l} \lambda_{k,l-1} & \sum_{(k,l) \in I} \lambda_{k,l}^2 & \sum_{(k,l) \in I} \lambda_{k,l} \lambda_{k-1,l} & \cdots & \sum_{(k,l) \in I} \lambda_{k,l} \lambda_{k+1,l+1-L} \\ \sum_{(k,l) \in I} \lambda_{k,l} \lambda_{k+1,l-1} & \sum_{(k,l) \in I} \lambda_{k,l} \lambda_{k-1,l} & \sum_{(k,l) \in I} \lambda_{k,l}^2 & \cdots & \sum_{(k,l) \in I} \lambda_{k,l} \lambda_{k,l+1-L} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{(k,l) \in I} \lambda_{k,l} \lambda_{k+1,l-L} & \cdots & \cdots & \cdots & \sum_{(k,l) \in I} \lambda_{k,l}^2 \end{pmatrix}. \quad (4.2.18)$$

4.3. Consistencia

El siguiente teorema establece la consistencia del estimador $\hat{\beta}_S$ obtenido en la primera etapa de la definición del estimador BMM 2D.

Teorema 4.3.1. *Sea Y proceso que satisface **P2** con proceso de innovaciones satisfaciendo **P3**. Sea ρ_1 tal que satisface **P1** con $\sup \rho_1 > b$ y $\psi_1 = \rho'_1$ acotada y continua. Entonces:*

- i) $\hat{\beta}_S$ es fuertemente consistente para β_0 , i.e., $\hat{\beta}_S \xrightarrow[N \rightarrow \infty]{} \beta_0$ a.e..
- ii) $s_N \xrightarrow[N \rightarrow \infty]{} s_0$ a.e..

El próximo teorema establece que el estimador de la escala s_N , obtenido en la primera etapa de la definición del estimador BMM 2D, converge casi seguramente a la M-escala en el verdadero parámetro ($s_0 = s(\beta_0)$).

Teorema 4.3.2. *Sea Y proceso que satisface la condición **P2**, con proceso de innovaciones ε que satisface **P3** y **P4**. Sea ρ_1 tal que satisface **P1** con $\sup \rho_1 > b$. Supongamos que $\psi_1 = \rho'_1$ es acotada, continua y sea η tal que se satisface **P5**. Entonces, si Y no es un ruido blanco,*

$$s_N^* = \min(s_N, s_N^b) \rightarrow s_0 \quad a.e..$$

En el siguiente teorema se demuestra que el estimador $\hat{\beta}_M$, obtenido en la segunda etapa de la definición del estimador BMM 2D, es consistente para estimar β_0 .

Teorema 4.3.3. *Sea Y proceso que satisface **P2** con proceso de innovaciones ε que satisface **P3**. Sean ρ_1 y ρ_2 satisfaciendo **P1**. Sean $\psi_i = \rho'_i$ acotadas y continuas con $i = 1, 2$ y tal que $\sup \rho_1 > b$. Entonces,*

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

Finalmente, en el teorema que sigue se demuestra la consistencia del estimador BMM 2D, $\hat{\beta}_M^*$.

Teorema 4.3.4. *Suponga que se satisfacen las suposiciones del Teorema 4.3.3, **P4** y **P5**. Entonces si el proceso Y no es un ruido blanco, con probabilidad 1 existe un N_0 tal que $\hat{\beta}_M^* = \hat{\beta}_M \forall N \geq N_0$*

y entonces

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

A continuación enunciamos los lemas necesarios y demostramos los teoremas 4.3.1, 4.3.2, 4.3.3 y 4.3.4. Las demostraciones de los lemas pueden verse en el Apéndice.

El siguiente Lema nos permitió demostrar los Lemas 4.3.4 y 4.3.10.

Lema 4.3.1. *Sea Y un proceso que satisface **P2** con proceso de innovaciones ε satisfaciendo **P3**. Entonces, para cualquier $d > 0$ tenemos que existe $W^0 = \{W_{i,j}^0\}_{(i,j) \in \mathbb{Z}^2}$ proceso estacionario y ergódico definido sobre (Ω, \mathcal{A}, P) tal que*

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

Además, $E(|W_{i,j}^0|^2) < \infty$.

Los siguientes tres lemas, fueron necesarios para probar el Teorema 4.3.1. El primero establece propiedades sobre la función que da la M-escala $s(\beta)$ para diferentes parámetros $\beta \in \mathcal{B}$. Los dos lemas que le siguen establecen relaciones entre la función $s(\beta)$ y el estimador de la escala con los residuos del modelo AR-2D ($S_N(\varepsilon_N(\beta))$).

Lema 4.3.2. *Sea Y un proceso que satisface **P2** con proceso de innovaciones ε satisfaciendo **P3**. Asuma que ρ_1 es una función satisfaciendo **P1** y que la función s es como en (4.2.4). Entonces,*

i) si $\beta \neq \beta_0$ tenemos que $s_0 = s(\beta_0) < s(\beta)$.

ii) s es continua.

Lema 4.3.3. *Bajo las suposiciones del Teorema 4.3.1, para cualquier $d > 0$ se satisface:*

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

Lema 4.3.4. *Bajo las suposiciones del Teorema 4.3.1, existe $d > 0$ tal que se satisface:*

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

A continuación demostramos el Teorema 4.3.1.

Demostración Teorema 4.3.1:

Probemos primero algunos resultados preliminares.

Sea $\epsilon > 0$ tan chico como se quiera y d como en el Lema 4.3.4. Como se demostró en el Lema 4.3.2, $s(\beta)$ es continua y alcanza un mínimo absoluto en $\beta_0 \in \mathcal{B}$. Veamos primero que existe un $\gamma > 0$ tal que

$$\min_{\beta \in \mathcal{B}_0 \times [-d, d], \|\beta - \beta_0\| \geq \epsilon} s(\beta) \geq s_0 + \gamma. \quad (4.3.1)$$

Como \mathcal{B}_0 es compacto y $s(\beta)$ es continua, $\forall \beta \in \mathcal{B}_0 \times [-d, d]$ y $\|\beta - \beta_0\| \geq \epsilon$ se tiene que existe un $0 < \gamma < 1$ tal que $s(\beta) - s(\beta_0) > \gamma$, i.e., $s(\beta) > \gamma + s(\beta_0) = \gamma + s_0$ entonces

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

Por el Lema 4.3.3, $\exists N_1$ tal que $\forall N > N_1$, $\sup_{\beta \in \mathcal{B}_0 \times [-d, d]} |S_N(\epsilon_N(\beta)) - s(\beta)| < \gamma/4$ a.e..

$$\begin{aligned} \Rightarrow -S_N(\epsilon_N(\beta)) + s(\beta) &\leq | -S_N(\epsilon_N(\beta)) + s(\beta) | < \frac{\gamma}{4}, \quad \forall \beta \in \mathcal{B}_0 \times [-d, d]; \\ \Rightarrow s(\beta) - \frac{\gamma}{4} &< S_N(\epsilon_N(\beta)), \quad \forall \beta \in \mathcal{B}_0 \times [-d, d]; \\ \Rightarrow s(\beta) - \frac{\gamma}{4} &< S_N(\epsilon_N(\beta)), \quad \forall \beta \in \mathcal{B}_0 \times [-d, d] \quad \wedge \quad \|\beta - \beta_0\| \geq \epsilon. \end{aligned}$$

Tomando mínimo en la última expresión y por la ecuación (4.3.1) tenemos que

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

Por lo tanto,

$$\min_{\beta \in \mathcal{B}_0 \times [-d, d], \|\beta - \beta_0\| \geq \epsilon} S_N(\epsilon_N(\beta)) > s_0 + \frac{\gamma}{2} \quad a.e.. \quad (4.3.2)$$

Como $\beta_0 \in \mathcal{B}_0 \times [-d, d]$ ($\phi_0 \in \mathcal{B}_0$ y $|\mu_0| \leq d$) y como por el Lema 4.3.3

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

entonces

$$\begin{aligned} & |S_N(\epsilon_N(\beta_0)) - s(\beta_0)| < \frac{\gamma}{4} \quad a.e.. \\ \Rightarrow & S_N(\epsilon_N(\beta_0)) - s(\beta_0) < \frac{\gamma}{4} \quad a.e., \\ \Rightarrow & S_N(\epsilon_N(\beta_0)) < \frac{\gamma}{4} + s(\beta_0) = \frac{\gamma}{4} + s_0 \quad a.e.. \end{aligned}$$

Por lo que

$$S_N(\epsilon_N(\beta_0)) < \frac{\gamma}{4} + s_0 \quad a.e.. \quad (4.3.3)$$

Por el Lema 4.3.4, $\liminf_{N \rightarrow \infty} \inf_{|\mu| > d, \phi \in \mathcal{B}_0} S_N(\epsilon_N(\beta)) > s_0 + 1$ a.e., es decir,

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

Como $B_N := \inf_{k \geq N} \left(\inf_{|\mu| > d, \phi \in \mathcal{B}_0} (S_k(\epsilon_k(\beta))) \right)$ es una sucesión creciente, $\exists N_2$ tal que $\forall N \geq N_2$, $B_N \geq s_0 + \gamma$ a.e. ($0 < \gamma < 1$).

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

En particular,

$$\inf_{|\mu| > d, \phi \in \mathcal{B}_0} (S_N(\epsilon_N(\beta))) \geq s_0 + \gamma \quad a.e., \quad \forall N \geq N_2. \quad (4.3.4)$$

Veamos ahora (i), es decir, $\hat{\beta}_S \rightarrow \beta_0$ a.e..

Dado $\epsilon > 0$, sea $N_0 = \max(N_1, N_2)$. Entonces, si $N \geq N_0$, se satisface (4.3.1), (4.3.2), (4.3.3) y (4.3.4).

De (4.3.2) y (4.3.4) se tiene que

$$\min_{\beta \in \mathcal{B}, \|\beta - \beta_0\| \geq \epsilon} S_N(\epsilon_N(\beta)) \geq s_0 + \frac{\gamma}{2} \quad a.e.. \quad (4.3.5)$$

Además, por la definición de $\hat{\beta}_S$, se cumple que

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

En particular, cuando $\beta = \beta_0$,

$$S_N(\epsilon_N(\beta_0)) \geq S_N(\epsilon_N(\hat{\beta}_S)). \quad (4.3.6)$$

Es decir, por (4.3.3) y (4.3.6),

$$s_0 + \frac{\gamma}{4} > S_N(\epsilon_N(\beta_0)) \geq S_N(\epsilon_N(\hat{\beta}_S)) \quad a.e.. \quad (4.3.7)$$

Si $\|\hat{\beta}_S - \beta_0\| \geq \epsilon$, entonces por (4.3.5)

$$S_N(\epsilon_N(\hat{\beta}_S)) \geq s_0 + \frac{\gamma}{2} \quad a.e., \quad (4.3.8)$$

y luego por (4.3.5) y (4.3.7) se tendría que

$$s_0 + \frac{\gamma}{4} > S_N(\epsilon_N(\beta_0)) \geq S_N(\epsilon_N(\hat{\beta}_S)) \geq s_0 + \frac{\gamma}{2} \quad a.e., \quad \text{lo cual es absurdo.}$$

Por lo tanto, debe ser que $\forall N > N_0 = \max(N_1, N_2)$, $\|\hat{\beta}_S - \beta_0\| \leq \epsilon \quad a.e.$, es decir,

$$\hat{\beta}_S \xrightarrow[N \rightarrow \infty]{} \beta_0 \quad a.e..$$

Veamos ahora (ii), es decir, $s_N \xrightarrow[N \rightarrow \infty]{} s_0 = s(\beta_0) \quad a.e..$

Sumando y restando $s(\hat{\beta}_S)$ se tiene que

$$|s_N - s_0| \leq |S_N(\epsilon_N(\hat{\beta}_S)) - s(\hat{\beta}_S)| + |s(\hat{\beta}_S) - s_0|.$$

Por la continuidad de $s(\beta)$ y como $\hat{\beta}_S \xrightarrow[N \rightarrow \infty]{} \beta_0 \quad a.e.$, entonces $s(\hat{\beta}_S) \xrightarrow[N \rightarrow \infty]{} s(\beta_0) = s_0 \quad a.e..$

Además, como $\hat{\beta}_S \xrightarrow[N \rightarrow \infty]{} \beta_0 \quad a.e.$, para $N \gg 0$, $\hat{\beta}_S \in \mathcal{B}_0 \times [-d, d]$. Luego, por el Lema 4.3.3, $|S_N(\epsilon_N(\hat{\beta}_S)) - s(\hat{\beta}_S)| \xrightarrow[N \rightarrow \infty]{} 0 \quad a.e..$

Finalmente se tiene que

$$s_N \xrightarrow[N \rightarrow \infty]{} s_0 \quad a.e.$$

y, por lo tanto, el teorema queda demostrado. \square

Los siguientes tres lemas permitieron demostrar el Teorema 4.3.2. El Lema 4.3.5 obtuvo una cota para los residuos en el modelo BIP-AR 2D la cuál permitió probar los lemas 4.3.6 y 4.3.11. Los resultados de los lemas 4.3.6 y 4.3.7 se utilizaron de forma directa en la demostración del Teorema 4.3.2. Además, el lema 4.3.7 contribuyó a demostrar el Lema 4.3.12. Estos lemas (4.3.6 y 4.3.7) es-

tablecieron relaciones entre la M-escala en los verdaderos parámetros y los M-estimadores de escala bajo residuos de un modelo BIP-AR 2D.

Lema 4.3.5. *Sea Y un proceso que satisface la condición **P2**. Dado $d > 0$ y $\tilde{\sigma} > 0$, existen constantes $C > 0$ y $D > 0$ tal que*

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

Lema 4.3.6. *Bajo las suposiciones del Teorema 4.3.2, dado $d > 0$, existe $\delta > 0$ tal que*

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

Lema 4.3.7. *Bajo las suposiciones del Teorema 4.3.2, existe $d > 0$ tal que*

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

A continuación demostramos el Teorema 4.3.2.

Demostración Teorema 4.3.2:

Sea $\tilde{\delta} > 0$ obtenido por el Lema 4.3.6.

De los Lemas 4.3.6 y 4.3.7 tenemos que existe $\delta = \min(\tilde{\delta}, 1)$ tal que

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

Luego, $\exists N_1$ tal que $\forall N > N_1$ se tiene que

$$\inf_{\beta \in \mathcal{B}} S_N(\varepsilon_N^b(\beta, \hat{\sigma}(\phi))) \geq s_0 + \delta \quad a.e.,$$

entonces,

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

En particular,

$$S_N(\varepsilon_N^b(\hat{\beta}_S^b, \hat{\sigma}(\hat{\phi}_S^b))) \geq s_0 + \delta \quad a.e., \quad \forall N > N_1,$$

es decir,

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

Por Teorema 4.3.1 (ii), $s_N \rightarrow s_0 \quad a.e.$ Entonces, existe N_2 tal que $\forall N > N_2$, $|s_N - s_0| < \delta \quad a.e.$, es decir,

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

Luego, si $N > \max(N_1, N_2)$, se satisface que $s_N < \delta + s_0 \leq s_N^b \quad a.e.$

Por lo tanto,

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

y como $s_N \rightarrow s_0 \quad a.e.$, entonces $s_N^* \rightarrow s_0 \quad a.e.$ y el teorema queda demostrado. □

Los próximos tres lemas intervinieron de manera directa en la demostración del Teorema 4.3.3. El Lema 4.3.8 probó propiedades sobre la función esperanza de los residuos del modelo AR-2D con la función $\rho_2(m(\beta))$. Los lemas 4.3.9 y 4.3.10 establecieron relaciones entre la función $m(\beta)$ y la función objetivo que determina la M-estimación de los parámetros del modelo AR-2D utilizando las funciones residuales de dicho modelo.

Lema 4.3.8. *Sea Y un proceso que satisface la condición **P2**, con proceso de innovaciones ε que satisface **P3**. Asumamos que ρ_2 , dada en la segunda etapa de la definición del estimador BMM satisface la condición **P1**. Sea $m : \mathcal{B} \rightarrow \mathbb{R}$ función definida por:*

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

Entonces,

i)

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

ii) m es una función continua.

Lema 4.3.9. *Sea Y proceso que satisface la condición **P2**. Supongamos que ρ_2 la condición **P1**. Definamos*

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

tal como en (2.4.6). Entonces,

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

Lema 4.3.10. *Bajo las suposiciones del Teorema 4.3.3, existen $d > 0$ y $\delta > 0$ tal que*

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

donde $m(\beta_0)$ es definida como en el Lema 4.3.8 y $M_N(\beta)$ como en el Lema 4.3.9.

A continuación demostramos el Teorema 4.3.3.

Demostración Teorema 4.3.3:

Sea $\epsilon > 0$ tan chico como se quiera y d y δ como en el Lema 4.3.10.

Por el Teorema de Convergencia Dominada de Lebesgue, la función $m(\beta)$ definida en 4.3.8 es continua (ρ_2 y $\varepsilon_{i,j}(\beta)$ son continuas y además ρ_2 es acotada).

Por el Lema 4.3.8, $m(\beta)$ alcanza un mínimo absoluto en β_0 .

Como $m(\beta)$ es continua $\forall \beta \in \mathcal{B}$, en particular es continua en β_0 . Es decir, $\exists 0 < \gamma < \delta$ tal que si $\|\beta - \beta_0\| \geq \epsilon$ entonces $m(\beta) - m(\beta_0) = |m(\beta) - m(\beta_0)| > \gamma$. Luego,

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

es decir,

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

Consecuentemente,

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

con lo cual:

$$\min_{\beta \in \mathcal{B}_0 \times [-d, d], \|\beta - \beta_0\| \geq \epsilon} m(\beta) > \gamma + m(\beta_0). \quad (4.3.9)$$

Por Lema 4.3.9, $\exists N_1$ tal que $\forall N > N_1$, $\sup_{\beta \in \mathcal{B}_0 \times [-d, d]} |M_N(\beta) - m(\beta)| < \frac{\gamma}{4}$ a.e.. Luego,

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

y por lo tanto,

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

Entonces,

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

Por lo tanto, de la ecuación anterior y usando 4.3.9 se tiene:

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

Por lo tanto,

$$\min_{\beta \in \mathcal{B}_0 \times [-d, d], \|\beta - \beta_0\| \geq \epsilon} M_N(\beta) > m(\beta_0) + \frac{\gamma}{2} \quad a.e., \quad \forall N > N_1. \quad (4.3.10)$$

Como $\beta_0 \in \mathcal{B}_0 \times [-d, d]$ y por el Lema 4.3.9,

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

es decir,

$$M_N(\beta_0) < \frac{\gamma}{4} + m(\beta_0) \quad a.e., \quad \forall N > N_1. \quad (4.3.11)$$

Luego, por el Lema 4.3.10,

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

es decir,

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

Debido a que $\inf_{k \geq N} (\inf_{|\mu| > d, \phi \in \mathcal{B}_0} M_k(\beta))$ es una sucesión creciente, $\exists N_2$ tal que $\forall N \geq N_2$:

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

entonces,

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

En particular:

$$\inf_{|\mu|>d, \phi \in \mathcal{B}_0} M_N(\beta) \geq m(\beta_0) + \delta > m(\beta_0) + \frac{\gamma}{2} \quad a.e., \quad \forall N \geq N_2. \quad (4.3.12)$$

Veamos que $\hat{\beta}_M \rightarrow \beta_0$ *a.e.* Dado $\epsilon > 0$, sea $N_0 = \max(N_1, N_2)$. Si $N > N_0$, se satisfacen (4.3.9), (4.3.10), (4.3.11) y (4.3.12).

De (4.3.10) y (4.3.12) se tiene que

$$\inf_{\beta \in \mathcal{B}, \|\beta - \beta_0\| \geq \epsilon} M_N(\beta) \geq m(\beta_0) + \frac{\gamma}{2} \quad a.e.. \quad (4.3.13)$$

Además, por la definición de $\hat{\beta}_M$, se cumple que

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

En particular,

$$M_N(\beta_0) \geq M_N(\hat{\beta}_M). \quad (4.3.14)$$

Si fuera que $\|\hat{\beta}_M - \beta_0\| \geq \epsilon$, entonces de (4.3.13) se tendría

$$M_N(\hat{\beta}_M) \geq m(\beta_0) + \frac{\gamma}{2}. \quad (4.3.15)$$

Luego, de (4.3.11), (4.3.14) y (4.3.15) se tendría que

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

Por lo tanto, debe ser que $\forall N > N_0 = \max(N_1, N_2)$, $\|\hat{\beta}_M - \beta_0\| < \epsilon$ *a.e.*, es decir, $\hat{\beta}_M \rightarrow \beta_0$ *a.e.* y el teorema queda demostrado. □

Los próximos dos lemas establecieron relaciones entre la función objetivo que determina la M-estimación de los parámetros del modelo AR-2D utilizando las funciones residuales del modelo BIP-AR 2D y la función $m(\beta)$. Estos lemas permitieron demostrar el teorema final: Teorema 4.3.4.

Lema 4.3.11. *Bajo las suposiciones del Teorema 4.3.3, para todo $d > 0$, existe $\delta > 0$ tal que*

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

Lema 4.3.12. *Bajo las suposiciones del Teorema 4.3.3, existen $d > 0$ y $\delta > 0$ tal que*

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

Por último, demostraremos el Teorema 4.3.4.

Demostración Teorema 4.3.4:

Dado $d > 0$ y $\delta_1 > 0$ como en el Lema 4.3.12, sea $\delta_2 > 0$ como en el Lema 4.3.11. Entonces, existe $\delta = \min(\delta_1, \delta_2) > 0$ tal que se cumple que

$$\liminf_{N \rightarrow \infty} \inf_{\beta \in \mathcal{B}} M_N^b(\beta) \geq m(\beta_0) + \delta. \quad (4.3.16)$$

Por Teorema 4.3.3, $\hat{\beta}_M \rightarrow \beta_0$ *a.e.*.

Veamos primero que

$$M_N(\hat{\beta}_M) \rightarrow m(\beta_0) \quad a.e.. \quad (4.3.17)$$

Sumando y restando $m(\hat{\beta}_M)$, tenemos que

$$|M_N(\hat{\beta}_M) - m(\beta_0)| \leq |M_N(\hat{\beta}_M) - m(\hat{\beta}_M)| + |m(\hat{\beta}_M) - m(\beta_0)|.$$

Por la continuidad de $m(\beta)$ y como $\hat{\beta}_M \rightarrow \beta_0$ *a.e.*, entonces $m(\hat{\beta}_M) \rightarrow m(\beta_0)$ *a.e.*, por lo que $|m(\hat{\beta}_M) - m(\beta_0)| \rightarrow 0$ *a.e.*.

Además, por el Lema 4.3.9, $\lim_{N \rightarrow \infty} \sup_{\beta \in \mathcal{B}_0 \times [-d, d]} |M_N(\beta) - m(\beta)| = 0$ *a.e.*. Como $\hat{\beta}_M \rightarrow \beta_0$ *a.e.* y $\beta_0 \in \mathcal{B} \times [-d, d]$, a partir de un $N \gg 0$, $\hat{\beta}_M \in \mathcal{B} \times [-d, d]$. Entonces, $|M_N(\hat{\beta}_M) - m(\hat{\beta}_M)| \rightarrow 0$. Por lo tanto, $|M_N(\hat{\beta}_M) - m(\beta_0)| \rightarrow 0$ *a.e.*.

Por (4.3.16), $\exists N_1$ tal que $\forall N > N_1$, $M_N^b(\beta) \geq \inf_{\beta \in \mathcal{B}} M_N^b(\beta) \geq m(\beta_0) + \delta$. En particular, $m(\beta_0) + \delta \leq M_N^b(\hat{\beta}_M)$.

Por (4.3.17), $\exists N_2$ tal que $\forall N > N_2$, $|M_N(\hat{\beta}_M) - m(\beta_0)| < \delta$, entonces $M_N(\hat{\beta}_M) < m(\beta_0) + \delta$.

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

Luego, por la definición de $\hat{\beta}_M^*$, $\hat{\beta}_M^* = \hat{\beta}_M$, $\forall N > \max(N_1, N_2)$ y como $\hat{\beta}_M \rightarrow \beta_0$ a.e., entonces

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

y el teorema queda demostrado. □

4.4. Normalidad Asintótica

El siguiente teorema establece la consistencia del estimador $\hat{\beta}_M^*$ obtenido en la segunda etapa de la definición del estimador BMM 2D.

Teorema 4.4.1. *Suponga que valen las suposiciones del Teorema 4.3.3. Más aún, suponga que ψ_2' es una función continua y acotada, $\sigma_\varepsilon^2 = E(\varepsilon_{i,j}) < \infty$ y la matrix $C = (C_{i,j})$ de dimensión $[(L+1)^2 - 1] \times [(L+1)^2 - 1]$ simétrica definida por*

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

es no singular. Entonces,

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

donde

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

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

Los siguientes lemas permitieron probar la normalidad asintótica del estimador $\hat{\beta}_M$.

Lema 4.4.1. *Bajo las suposiciones del Teorema 4.4.1, se tiene*

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

donde

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

Lema 4.4.2. *Bajo las suposiciones del Teorema 4.4.1, tenemos*

$$\frac{1}{\sqrt{N}} \left\| \sum_{(i,j) \in (W_M \sim T)} \left[\nabla \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_N^*} \right) \right) - \nabla \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \right] \right\| \rightarrow 0 \quad \text{en probabilidad.}$$

Lema 4.4.3. *Bajo las suposiciones del Teorema 4.4.1, tenemos que $\forall d > 0$:*

i)

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

donde $\|A\|$ denota la norma l_2 de la matriz A .

ii)

$$E \left[\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \right] = \frac{1}{s_0^2} E \left(\psi_2' \left(\frac{\varepsilon_{i,j}}{s_0} \right) \right) \cdot E \left(\nabla(\varepsilon_{i,j}(\beta_0)) \cdot \nabla(\varepsilon_{i,j}(\beta_0))^t \right).$$

Demostración Teorema 4.4.1:

Por definición de $\hat{\beta}_M$ se cumple que

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

Por lo que $\hat{\beta}_M$ satisfice:

$$\sum_{(i,j) \in (W_M \sim T)} \nabla \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\hat{\beta}_M)}{s_N^*} \right) \right) = 0.$$

Dados β_0 y $\hat{\beta}_M$ en \mathcal{B} , entonces por el Teorema del valor medio se obtiene:

$$\begin{aligned} 0 &= \sum_{(i,j) \in (W_M \sim T)} \nabla \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\hat{\beta}_M)}{s_N^*} \right) \right) \\ &= \sum_{(i,j) \in (W_M \sim T)} \nabla \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_N^*} \right) \right) + \sum_{(i,j) \in (W_M \sim T)} \nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_N^*} \right) \right) (\hat{\beta}_M - \beta_0) \end{aligned} \quad (4.4.1)$$

donde $\tilde{\beta}$ es un punto intermedio entre β_0 y $\hat{\beta}_M$, es decir, $\tilde{\beta} = \beta_0 + \theta(\hat{\beta}_M - \beta_0)$ con $0 < \theta < 1$.

Por el Teorema 4.3.3, tenemos que $\hat{\beta}_M \rightarrow \beta_0$ *a.e.* y, por lo tanto, $\tilde{\beta} \rightarrow \beta_0$ *a.e.*.

Tomemos $d > 0$ tal que $d > |\mu_0|$, entonces, con probabilidad 1, existe N_0 tal que $\hat{\beta}_M \in \mathcal{B}_0 \times [-d, d]$, $\forall N \geq N_0$.

Sea

$$A_N = \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_N^*} \right) \right).$$

Veamos que

$$\lim_{N \rightarrow \infty} A_N = E \left[\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \right] \quad a.e.. \quad (4.4.2)$$

Sea $\epsilon > 0$,

$$\begin{aligned} & \left\| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_N^*} \right) \right) - E \left[\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \right] \right\| \leq \\ & \left\| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_N^*} \right) \right) - E \left[\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_0} \right) \right) \right] \right\| \\ & + \left\| E \left[\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_0} \right) \right) \right] - E \left[\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \right] \right\|. \end{aligned} \quad (4.4.3)$$

Como $\tilde{\beta} \rightarrow \beta_0$ *a.e.*, para un $N \gg 0$, $\tilde{\beta} \in \mathcal{B} \times [-d, d]$. Luego, por el Lema 4.4.3-(i), $\exists N_0$ tal que $\forall N > N_0$, el primer término de la desigualdad (4.4.3) nos queda:

$$\left\| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_N^*} \right) \right) - E \left[\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_0} \right) \right) \right] \right\| < \epsilon/2 \quad a.e.. \quad (4.4.4)$$

Como la función $\rho_2'' = \psi_2'$ es continua y acotada y las funciones residuales $\varepsilon_{i,j}(\beta)$ son continuas, por el Teorema de Convergencia Dominada de Lebesgue, resulta que $E \left[\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \right) \right]$ es una función continua como función de β . Luego, como $\tilde{\beta} \rightarrow \beta_0$ *a.e.*, se tiene que $\exists N_1$ tal que $\forall N > N_1$, el segundo término de la desigualdad (4.4.3) nos queda:

$$\left\| E \left[\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \right] - E \left[\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\tilde{\beta})}{s_0} \right) \right) \right] \right\| < \epsilon/2 \quad a.e.. \quad (4.4.5)$$

Por lo tanto, de (4.4.4) y (4.4.5), para todo $N > \max(N_0, N_1)$ la desigualdad (4.4.3) queda menor que ϵ . Luego, (4.4.2) se satisface.

Por Lema 4.4.3-(ii), $A := E \left[\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \right]$ es no singular y como $A_N \rightarrow A$ *a.e.* entonces para $N \gg 0$ se tiene que A_N es no singular.

Por otro lado, dividiendo (4.4.1) por N y llamando

$$C_N := \frac{1}{\sqrt{N}} \sum_{(i,j) \in (W_M \sim T)} \nabla \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_N^*} \right) \right),$$

se obtiene que la ecuación (4.4.1) es equivalente a:

$$\begin{aligned} \frac{\sqrt{N}}{N} C_N + A_N (\hat{\beta}_M - \beta_0) &= 0 \\ \frac{1}{\sqrt{N}} C_N + A_N (\hat{\beta}_M - \beta_0) &= 0 \\ C_N + \sqrt{N} A_N (\hat{\beta}_M - \beta_0) &= 0 \\ A_N \sqrt{N} (\hat{\beta}_M - \beta_0) &= -C_N. \end{aligned}$$

Entonces para $N \gg 0$ tal que A_N sea no singular, se tiene que

$$\sqrt{N} (\hat{\beta}_M - \beta_0) = -A_N^{-1} C_N.$$

Luego, para probar el teorema basta ver que $A_N^{-1} C_N \xrightarrow{D} \mathcal{N}(0, D)$.

Veamos primero que $C_N \xrightarrow{D} \mathcal{N}(0, V_0)$ con V_0 como en (4.2.14).

Sea $Z_N := \frac{1}{\sqrt{N}} \sum_{(i,j) \in (W_M \sim T)} \nabla \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right)$. Entonces, por el Lema 4.4.2 se tiene que

$$\|C_N - Z_N\| \rightarrow 0 \quad \text{en probabilidad,}$$

es decir,

$$C_N - Z_N \rightarrow 0 \quad \text{en probabilidad.}$$

Además, el Lema 4.4.1 dice que $Z_N \xrightarrow{D} Z$ con $Z \sim \mathcal{N}(0, V_0)$. Luego, por Slutsky,

$$Z_N + (C_N - Z_N) \xrightarrow{D} Z + 0 = Z,$$

es decir,

$$C_N \xrightarrow{D} Z \text{ con } Z \sim \mathcal{N}(0, V_0). \quad (4.4.6)$$

Para ir finalizando la demostración veamos que $A_N^{-1}C_N \xrightarrow{D} \mathcal{N}(0, A^{-1}V_0(A^{-1})^t)$.

Como $h : \mathbb{R}^{(L+1)^2} \rightarrow \mathbb{R}^{(L+1)^2}$ definida por $h(X) = A^{-1}X$ es una función medible y continua en $\mathbb{R}^{(L+1)^2}$ y debido a (4.4.6) se tiene por el Teorema 29.2 de Billingsley (2013) que

$$h(C_N) = A^{-1}C_N \xrightarrow{D} h(Z) = A^{-1}Z \text{ con } A^{-1}Z \sim \mathcal{N}(0, A^{-1}V_0(A^{-1})^t).$$

Por otra parte, como C_N es acotada (ρ'_2 es acotada) y $A_N^{-1} \rightarrow A^{-1}$ *a.e.* (invertir una matriz es una función continua) se tiene que

$$\|A^{-1}C_N - A_N^{-1}C_N\| \rightarrow 0 \text{ en probabilidad.}$$

Entonces,

$$A^{-1}C_N - A_N^{-1}C_N \rightarrow 0 \text{ en probabilidad.} \quad (4.4.7)$$

Luego, como $A_N^{-1}C_N = (A_N^{-1}C_N - A^{-1}C_N) + A^{-1}C_N$, por Teorema 5.1.5 de Lehmann (2004) y usando (4.4.7) y (4.4.6) se tiene que

$$A_N^{-1}C_N \xrightarrow{D} \mathcal{N}(0, A^{-1}V_0(A^{-1})^t)$$

Por lo tanto,

$$\sqrt{N}(\hat{\beta}_M - \beta_0) \xrightarrow{D} \mathcal{N}(0, A^{-1}V_0(A^{-1})^t).$$

Queda ver que $D = A^{-1}V_0(A^{-1})^t$.

Por (4.2.15) se tiene que

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

Y por (4.2.8) se tiene

$$\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) = \frac{1}{s_0^2} \psi_2' \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \cdot \nabla (\varepsilon_{i,j}(\beta_0)) + \frac{1}{s_0} \psi_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \cdot \nabla^2 (\varepsilon_{i,j}(\beta_0)).$$

Entonces,

$$\begin{aligned} A &= E \left[\frac{1}{s_0^2} \psi_2' \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \cdot \nabla (\varepsilon_{i,j}(\beta_0)) \cdot \nabla (\varepsilon_{i,j}(\beta_0))^t + \frac{1}{s_0} \psi_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \cdot \nabla^2 (\varepsilon_{i,j}(\beta_0)) \right] \\ &= E \left(\frac{1}{s_0^2} \psi_2' \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \cdot E [\nabla (\varepsilon_{i,j}(\beta_0)) \nabla (\varepsilon_{i,j}(\beta_0))^t] + E \left(\frac{1}{s_0} \psi_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \cdot E [\nabla^2 (\varepsilon_{i,j}(\beta_0))]. \end{aligned}$$

Luego, por (4.2.12),

$$\begin{aligned} A &= E \left(\frac{1}{s_0^2} \psi_2' \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \cdot E [\nabla (\varepsilon_{i,j}(\beta_0)) \nabla (\varepsilon_{i,j}(\beta_0))^t], \\ \Rightarrow A^{-1} &= \frac{1}{E \left(\frac{1}{s_0^2} \psi_2' \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right)} \cdot E [\nabla (\varepsilon_{i,j}(\beta_0)) \nabla (\varepsilon_{i,j}(\beta_0))^t]^{-1}. \end{aligned}$$

Como además $E [\nabla (\varepsilon_{i,j}(\beta_0)) \nabla (\varepsilon_{i,j}(\beta_0))^t]$ es simétrica, entonces $E [\nabla (\varepsilon_{i,j}(\beta_0)) \nabla (\varepsilon_{i,j}(\beta_0))^t]^{-1}$ es simétrica. Entonces,

$$\begin{aligned} A^{-1} V_0 (A^{-1})' &= A^{-1} V_0 A^{-1} = \frac{E \left(\frac{1}{s_0^2} \psi_2'^2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right)}{E \left(\frac{1}{s_0^2} \psi_2' \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right)^2} \cdot E [\nabla (\varepsilon_{i,j}(\beta_0)) \nabla (\varepsilon_{i,j}(\beta_0))^t]^{-1} \\ &= \frac{s_0^2 E \left(\psi_2'^2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right)}{E \left(\psi_2' \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right)^2} \cdot E [\nabla (\varepsilon_{i,j}(\beta_0)) \nabla (\varepsilon_{i,j}(\beta_0))^t]^{-1}. \end{aligned}$$

Por último, como se vió en (4.2.16),

$$E [\nabla (\varepsilon_{i,j}(\beta)) \nabla (\varepsilon_{i,j}(\beta))^t] = \begin{pmatrix} \sigma_\varepsilon^2 C & 0_{L \times 1} \\ 0_{1 \times L} & \xi^2 \end{pmatrix}.$$

Luego,

$$E [\nabla (\varepsilon_{i,j}(\beta_0)) \nabla (\varepsilon_{i,j}(\beta_0))^t]^{-1} = \begin{pmatrix} \sigma_\varepsilon^{-2} C^{-1} & 0 \\ 0 & \xi_0^{-2} \end{pmatrix}$$

y por lo tanto el teorema queda demostrado. □

Finalmente, el siguiente teorema prueba la normalidad asintótica del estimador BMM 2D $\hat{\beta}_M^*$:

Teorema 4.4.2. *Suponga que valen las suposiciones del Teorema 4.4.1, P4 y P5. Entonces,*

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

donde D es definida como en el Teorema 4.4.1.

Demostración Teorema 4.4.2:

Por la demostración del Teorema 4.3.4, $\exists N_0$ tal que si $N > N_0$ entonces $\hat{\beta}_M^* = \hat{\beta}_M$.

Sean

$$Z_N = \sqrt{N}(\hat{\beta}_M^* - \beta_0) \text{ con } F_{Z_N} \text{ su función de distribución}$$

y

$$Y_N = \sqrt{N}(\hat{\beta}_M - \beta_0) \text{ con } F_{Y_N} \text{ su función de distribución.}$$

Entonces, $Z_N = Y_N$ para $N > N_0$ y entonces $F_{Z_N} = F_{Y_N}$ para $N > N_0$.

Por el Teorema 4.4.1, $\exists N_1$ tal que si $N > N_1$, $|F_{Y_N}(x) - F_Z(x)| < \varepsilon \forall x$ donde F_Z es la distribución de una $\mathcal{N}(0, D)$.

Luego, para $N > \max(N_0, N_1)$, se cumple que $|F_{Z_N}(x) - F(x)| < \varepsilon \forall x$.

Por lo tanto $Z_N \xrightarrow{D} Z$ y el teorema queda demostrado.

□

Conclusiones y Trabajos Futuros

En este trabajo se ha propuesto un nuevo estimador para los parámetros del modelo autorregresivo bidimensional (AR-2D) con contaminación al que se ha llamado BMM 2D. La iniciativa permitió generalizar para procesos AR-2D la versión unidimensional del estimador BMM desarrollado para series de tiempo contaminadas presentado por Muler et al (2009).

El nuevo estimador permite la estimación de los parámetros del modelo bajo contaminación de reemplazo, que hasta ahora solo había sido definida para series de tiempo. Este tipo de contaminación incluye a la contaminación aditiva que es frecuente en el tratamiento y análisis de imágenes digitales.

En el Capítulo 3 se analizó vía simulación y estudios de Monte Carlo el comportamiento del estimador BMM 2D para el caso de un modelo AR-2D con tres parámetros, puro y bajo diferentes esquemas y niveles de contaminación. Se verificó que el estimador propuesto resulta comparable con el estimador LS en el caso en que el modelo no está contaminado (modelo puro) y las variables aleatorias del proceso de innovaciones distribuyen como una normal; mientras que bajo contaminación, el estimador BMM compitió con éxito con respecto a otras propuestas robustas (estimadores M, GM y RA bidimensionales). Para el caso analizado el estimador BMM 2D resultó superior en exactitud y precisión con respecto al estimador M, siendo mejor en precisión que el estimador RA y tanto o más exacto que el estimador GM. Si bien el estimador GM resultó en precisión y exactitud el mejor competidor de la propuesta BMM 2D, no se tiene conocimiento del comportamiento asintótico del estimador GM; por el contrario en el Capítulo 4 de este trabajo se probó que el estimador BMM es consistente y normalmente asintótico. Estos resultados lo posicionan en un escalón superior con respecto a la propuesta GM 2D. Por otro lado, si bien el estimador RA es consistente y normalmente asintótico, presenta con respecto al estimador BMM 2D mayores dificultades a la hora de su implementación y un mayor costo computacional como se vió en el Capítulo 3. Además, del estudio de simulación resultó que las estimaciones del RA son menos precisas que las obtenidas a partir del estimador BMM 2D.

Para indagar sobre las propiedades prácticas del estimador en problemas que utilizan el modelo AR-2D vinculado a análisis y procesamiento de imágenes ópticas, se realizaron tres experimentos, a partir de una adaptación de los algoritmos presentados en Ojeda et al (2010) para representar y segmentar imágenes. Con este propósito se ajustaron localmente a las imágenes modelos AR-2D con tres parámetros y se estimaron los mismos usando el estimador BMM 2D. A la luz de los resultados obtenidos en la sección (3.4), concluimos que el estimador propuesto en esta tesis, vía los algoritmos adaptados de Ojeda et al (2010) es útil para estimar los parámetros de los modelos locales permitiendo resaltar bordes y contornos de las imágenes.

Dentro de las tareas pendientes queda por un lado un estudio comparativo más profundo del estimador BMM 2D con sus competidores robustos GM y RA en ajustes locales de modelos AR-2D a imágenes en problemas de análisis y procesamiento de imágenes en general. Por otro, se plantea como un trabajo a futuro el estudio de las propiedades teóricas de robustez del estimador BMM 2D: punto de quiebre, máximo sesgo asintótico y curva de influencia. Estos conceptos no han sido abordados para el estimador BMM 2D en este trabajo ni tampoco se conocen estudios de estas propiedades para las propuestas M, GM y RA.

Apéndice

Este apéndice presenta las demostraciones de los lemas presentados en el Capítulo 4. Además, se muestran los códigos programados para los diferentes experimentos presentados en este trabajo.

A.1. Demostraciones de lemas

A continuación demostraremos el Lema 4.3.1.

Demostración Lema 4.3.1:

Veamos (i). Una expresión para las funciones residuales $\varepsilon_{i,j}(\beta)$ es la que se observa en (2.0.4), la cuál es equivalente a

$$\varepsilon_{i,j}(\beta) = Y_{i,j} - \sum_{(k,l) \in T} \phi_{k,l} Y_{i-k,j-l} + \mu \xi, \quad (\text{A.1.1})$$

donde $\xi = -1 + \sum_{(k,l) \in T} \phi_{k,l}$ (ver (2.2.3)). Entonces

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

Como $\beta \in \mathcal{B}_0 \times [-d, d]$, entonces $\sum_{(k,l) \in T} |\phi_{k,l}| < 1$ y $|\mu| \leq d$, entonces $|\xi| < 1$. Luego, como $\#(T) < \infty$, se tiene que

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

Definimos $W^0 = \{W_{i,j}^0\}_{(i,j) \in \mathbb{Z}^2}$ tal que $W_{i,j}^0 := |Y_{i,j}| + \sum_{(k,l) \in T} |Y_{i-k,j-l}| + 2d$. Además, como Y es estacionario con momento de segundo orden finito, W^0 es estacionario y $E((W_{i,j}^0)^2) < \infty$.

Veamos que W^0 es ergódico. Sea $g : (\mathbb{R}^{\mathbb{Z}^2}, \mathcal{B}^{\mathbb{Z}^2}) \rightarrow (\mathbb{R}^{\mathbb{Z}^2}, \mathcal{B}^{\mathbb{Z}^2})$ función medible definida por $g(X) = |B^{(0,0)}(X)| + \sum_{(k,l) \in T} |B^{(k,l)}(X)| + 2d$, donde $|B^{(k,l)}(X)|(i, j) = |B^{(k,l)}(X)(i, j)|$. Entonces $W^0 = g(Y)$. Dado $A \in \mathcal{I} = \{A' \in \mathcal{B}^{\mathbb{Z}^2} : A' \text{ es } B^{(k,l)}\text{-invariante } \forall (k,l) \in \mathbb{Z}^2\}$, queremos ver que $P_{W^0}(A) = 0$ ó 1 .

Como $P_{W^0}(A) = P(W^0 \in A) = P(g(Y) \in A) = P(Y^{-1}(g^{-1}(A))) = P(Y \in g^{-1}(A))$, para ver lo que queremos basta ver que $g^{-1}(A) \in \mathcal{I}$.

Veamos que primero que $g^{-1}(A)$ está contenido en $B^{(s,t)}(g^{-1}(A)) = \{B^{(s,t)}(X) : g(X) \in A\}$, $\forall (s,t) \in \mathbb{Z}^2$. Sea $X \in g^{-1}(A)$, luego $g(X) \in A$, es decir, $|B^{(0,0)}(X)| + \sum_{(k,l) \in T} |B^{(k,l)}(X)| + 2d \in A$. Aplicando $B^{(-s,-t)}$ se tiene $B^{(-s,-t)}|B^{(0,0)}(X)| + \sum_{(k,l) \in T} B^{(-s,-t)}|B^{(k,l)}(X)| + 2d \in B^{(-s,-t)}(A) = A$ (pues $A \in \mathcal{I}$). Luego, $|B^{(0,0)}(B^{(-s,-t)}(X))| + \sum_{(k,l) \in T} |B^{(k,l)}(B^{(-s,-t)}(X))| + 2d \in A$, es decir, $g(B^{(-s,-t)}(X)) \in A$. Por lo tanto, $X = B^{(s,t)}(B^{(-s,-t)}(X)) \in B^{(s,t)}(g^{-1}(A))$. Luego, $g^{-1}(A) \subseteq B^{(s,t)}(g^{-1}(A))$ como queríamos. De manera análoga se prueba que $B^{(s,t)}(g^{-1}(A))$ está contenido en $g^{-1}(A)$. Por lo que $g^{-1}(A) \in \mathcal{I}$ y $P(Y \in g^{-1}(A)) = 0$ ó 1 .

□

Demostración Lema 4.3.2:

Probemos (i). Note que por cumplirse **P1** siempre podemos elegir a $s(\beta)$ como una solución positiva de la ecuación (4.2.4) ya que si s es solución, $|s|$ también es solución.

Sea $\beta = (\phi, \mu) \neq \beta_0 = (\phi_0, \mu_0)$. Tenemos

$$\begin{aligned} \varepsilon_{i,j}(\beta) &= \Phi(B_1, B_2)(Y_{i,j} - \mu) \\ &= \Phi(B_1, B_2)(Y_{i,j} - \mu_0) + \Phi(B_1, B_2)(\mu_0 - \mu) \\ &= \Phi(B_1, B_2)\Phi_0(B_1, B_2)^{-1}\varepsilon_{i,j} + \left(1 - \sum_{(k,l) \in T} \phi_{k,l}\right) (\mu_0 - \mu) \\ &= \omega(B_1, B_2)\varepsilon_{i,j} + c.(\mu_0 - \mu), \end{aligned} \tag{A.1.2}$$

donde $\omega(B_1, B_2) := \Phi(B_1, B_2)\Phi_0(B_1, B_2)^{-1}$ y $c = -\xi$. Como $\beta, \beta_0 \in \mathcal{B}$, entonces $\Phi(z_1, z_2)\Phi_0(z_1, z_2)^{-1}$ puede escribirse sobre D^* como una suma de serie de potencias: $1 + \sum_{(k,l) \in I \setminus \{(0,0)\}} w_{k,l} z_1^k z_2^l$ (ver Guyon (1995)).

Luego,

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

Sea

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

Entonces, por la ecuación (A.1.2),

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

Entonces

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

y luego,

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

Sea $S(p, q)$ definida como

$$S(p, q) = E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j} + p}{q} \right) \right), \quad (\text{A.1.3})$$

para todo $p, q \in \mathbb{R}$ con $q \neq 0$. Note que $S(p, q)$ es decreciente en $|q|$. El Lema 3.1 de Yohai (1985), muestra que si se satisfacen **P1** y **P3**, entonces para todo $p, q \neq 0$ se tiene

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

y la igualdad se cumple si y sólo si $q \neq 0$. Luego,

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

y la igualdad vale si y sólo si $\Delta_{i,j}(\beta) = 0$ a.e.. Debido a la identificabilidad del modelo AR-2D, esto ocurre si y sólo si $\beta = \beta_0$. Entonces $\beta \neq \beta_0$ implica

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

y por lo tanto, como $S(p, q)$ es decreciente en $|q|$ y $s(\beta)$ es una función positiva, tenemos $s_0 = s(\beta_0) < s(\beta)$.

Demostremos ahora (ii).

Sea $\epsilon > 0$ tan chico como se quiera. Como la función $S(p, q)$ definida en la ecuación (A.1.3) es decreciente en $|q|$, sea $\beta_1 \in \mathcal{B}$ y $s_1 = s(\beta_1) > 0$, ocurrirá que:

$$E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1 + \epsilon} \right) \right) = S(\Delta_{i,j}(\beta_1), s_1 + \epsilon) < S(\Delta_{i,j}(\beta_1), s_1) = E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1} \right) \right) = b \quad (\text{A.1.4})$$

y

$$E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1 - \epsilon} \right) \right) = S(\Delta_{i,j}(\beta_1), s_1 - \epsilon) > S(\Delta_{i,j}(\beta_1), s_1) = E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1} \right) \right) = b. \quad (\text{A.1.5})$$

Definamos las variables aleatorias

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

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

Luego, las sucesiones de variables aleatorias $\{q_1(1/n)\}_{n \geq 1}$ y $\{q_2(1/n)\}_{n \geq 1}$ convergen a $\rho_1 \left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1 + \epsilon} \right)$ y $\rho_1 \left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1 - \epsilon} \right)$ respectivamente.

Entonces por el Teorema de Convergencia Dominada de Lebesgue,

$$\lim_{n \rightarrow \infty} \int_{\Omega} q_1(1/n) dP_{\beta_0}(\omega) = \int_{\Omega} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1 + \epsilon} \right) dP_{\beta_0}(\omega) = E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1 + \epsilon} \right) \right) \quad (\text{A.1.6})$$

y

$$\lim_{n \rightarrow \infty} \int_{\Omega} q_2(1/n) dP_{\beta_0}(\omega) = \int_{\Omega} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1 - \epsilon} \right) dP_{\beta_0}(\omega) = E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}(\beta_1)}{s_1 - \epsilon} \right) \right), \quad (\text{A.1.7})$$

y por la tanto, de (A.1.4) con (A.1.6) y, (A.1.5) con (A.1.7), respectivamente, existe n_0 tal que si $n \geq n_0$,

$$\int_{\Omega} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s_1 + \epsilon} \right) dP_{\beta_0} \leq \int_{\Omega} q_1(1/n) dP_{\beta_0} < b = \int_{\Omega} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)} \right) dP_{\beta_0} \quad (\text{A.1.8})$$

y

$$\int_{\Omega} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s_1 - \epsilon} \right) dP_{\beta_0} \geq \int_{\Omega} q_2(1/n) dP_{\beta_0} > b = \int_{\Omega} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta)} \right) dP_{\beta_0} \quad (\text{A.1.9})$$

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

Luego, sea $\delta = \frac{1}{n_0}$. Por las ecuaciones (A.1.8) y (A.1.9), si $\|\beta - \beta_0\| < \delta$ se satisface

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

y

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

entonces como ρ_1 es una función positiva, se tiene

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

y

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

Por lo tanto, como $\rho_1(|u|)$ es no decreciente y $s(\beta)$ es positiva, se tiene $s_1 - \epsilon < s(\beta) < s_1 + \epsilon$, es decir, s es continua en β_1 cualquiera sea $\beta_1 \in \mathcal{B}$. Luego, s es continua. \square

Demostración Lema 4.3.3:

Sean

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

Entonces, por definición de $s(\beta)$, $h_1 > 0$ y $h_2 < \infty$.

Consideremos la función continua $f(y, \beta, c) = \rho_1\left(\frac{\Phi(B_1, B_2)(y - \mu)}{c}\right) - E_{\beta_0}\left(\rho_1\left(\frac{\Phi(B_1, B_2)(y - \mu)}{c}\right)\right)$ definida sobre $\mathbb{R} \times C$ con $C = \mathcal{B}_0 \times [-d, d] \times [h_1/2, 2h_2]$ compacto. Como $Y = \{Y_{i,j}\}_{(i,j) \in \mathbb{Z}^2}$ es un proceso ergódico, $E(f(Y, \beta, c)) = 0$ y $\sup_{(\beta, c) \in C} |f(Y, \beta, c)| \leq K$ (pues por ser una función continua sobre un compacto, es acotada), del Lema 3 de Muler and Yohai (2002) se tiene que

$$\lim_{N \rightarrow \infty} \sup_{\beta \in \mathcal{B}_0 \times [-d, d], c \in [h_1/2, 2h_2]} \left| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{c}\right) - E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{c}\right)\right) \right| = 0 \quad a.e., \quad (\text{A.1.10})$$

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

Sea $0 \leq \epsilon \leq h_1/2$ y definamos las funciones $g_i : \mathcal{B} \rightarrow \mathbb{R}$ para $i = 1, 2$ como

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

Por la definición de $s(\beta)$ y debido a que ρ_1 satisface **P1**, tenemos que $g_1(\beta) < b$ y $g_2(\beta) > b$ $\forall \beta \in \mathcal{B}$.

Como \mathcal{B}_0 es un conjunto compacto y g_1 y g_2 son continuas (pues $s(\beta)$ y $\varepsilon_{i,j}(\beta)$ son continuas y ρ_1 cumple **P1**), tenemos que

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

Sea $\delta = \min(b - \kappa_1, \kappa_2 - b)$. De la ecuación (A.1.10), existe un N_0 tal que para todo $N \geq N_0$,

$$\sup_{\beta \in \mathcal{B}_0 \times [-d, d], c \in [h_1/2, 2h_2]} \left| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{c}\right) - E_{\beta_0}\left(\rho_1\left(\frac{\varepsilon_{i,j}(\beta)}{c}\right)\right) \right| \leq \frac{\delta}{2} \quad a.e.. \quad (\text{A.1.11})$$

Observemos que $s(\beta) - \epsilon \in [h_1/2, 2h_2]$ pues $h_1/2 = h_1 - \frac{h_1}{2} < s(\beta) - \frac{h_1}{2} < s(\beta) - \epsilon$ y además

$s(\beta) - \epsilon < h_2 - \epsilon < h_2 + \epsilon < h_2 + h_2 = 2h_2$ por la condición sobre ϵ y por las definiciones de h_1 y h_2 respectivamente.

Por lo tanto, de (A.1.11), obtenemos que $\forall N \geq N_0$:

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

es decir,

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

Luego, tomando ínfimo tenemos

$$\inf_{\beta \in \mathcal{B}_0 \times [-d, d]} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) - \epsilon} \right) \geq \inf_{\beta \in \mathcal{B}_0 \times [-d, d]} g_2(\beta) - \frac{\delta}{2} = \kappa_2 - \frac{\delta}{2} \text{ a.e.} \quad (\text{A.1.12})$$

Además, por la definición de δ , se sabe que $\kappa_2 - \frac{\delta}{2} \geq b + \frac{\delta}{2} > b$, y por la ecuación (A.1.12) se tiene que

$$\inf_{\beta \in \mathcal{B}_0 \times [-d, d]} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) - \epsilon} \right) > b = \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{S_N(\varepsilon_N(\beta))} \right) \text{ a.e.} \quad (\text{A.1.13})$$

Análogamente podemos observar que $s(\beta) + \epsilon \in [h_1/2, 2h_2]$ pues $h_1/2 < h_1 < h_1 + \epsilon < s(\beta) + \epsilon$ y además $s(\beta) + \epsilon < h_2 + \epsilon < h_2 + \frac{h_1}{2} < h_2 + h_2 = 2h_2$.

Por lo tanto, de (A.1.11), obtenemos que para todo $N \geq N_1$,

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

es decir,

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

Luego, tomando supremo tenemos

$$\sup_{\beta \in \mathcal{B}_0 \times [-d, d]} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) + \epsilon} \right) \leq \sup_{\beta \in \mathcal{B}_0 \times [-d, d]} g_1(\beta) + \frac{\delta}{2} = \kappa_1 + \frac{\delta}{2} \text{ a.e.} \quad (\text{A.1.14})$$

Además, por la definición de δ , se sabe que $\kappa_1 + \frac{\delta}{2} \geq b - \frac{\delta}{2} < b$, y por la ecuación (A.1.14) se tiene

que

$$\sup_{\beta \in \mathcal{B}_0 \times [-d, d]} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) + \epsilon} \right) < b = \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{S_N(\varepsilon_N(\beta))} \right) \quad a.e.. \quad (\text{A.1.15})$$

Veamos que existe un $(i_0, j_0) \in (W_M \sim T)$ tal que $\left| \frac{\varepsilon_{i_0, j_0}(\beta)}{s(\beta) + \epsilon} \right| < \left| \frac{\varepsilon_{i_0, j_0}(\beta)}{S_N(\varepsilon_N(\beta))} \right|$ a.e., $\forall \beta \in \mathcal{B}_0 \times [-d, d]$.
Supongamos que $\left| \frac{\varepsilon_{i,j}(\beta)}{s(\beta) + \epsilon} \right| \geq \left| \frac{\varepsilon_{i,j}(\beta)}{S_N(\varepsilon_N(\beta))} \right|$, $\forall (i, j) \in (W_M \sim T)$ y $\forall \beta \in \mathcal{B}_0 \times [-d, d]$.

Como $s(\beta) + \epsilon > 0$ y $S_N(\varepsilon_N(\beta)) > 0$ y por la monotonicidad de $\rho_1(|u|)$ obtenemos $\rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s(\beta) + \epsilon} \right) = \rho_1 \left(\frac{|\varepsilon_{i,j}(\beta)|}{s(\beta) + \epsilon} \right) \geq \rho_1 \left(\frac{|\varepsilon_{i,j}(\beta)|}{S_N(\varepsilon_N(\beta))} \right) = \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{S_N(\varepsilon_N(\beta))} \right)$, $\forall (i, j) \in (W_M \sim T)$ y $\forall \beta \in \mathcal{B}_0 \times [-d, d]$, lo cual implica

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

que es absurdo por (A.1.15).

Luego, vale que $\left| \frac{\varepsilon_{i_0, j_0}(\beta)}{s(\beta) + \epsilon} \right| < \left| \frac{\varepsilon_{i_0, j_0}(\beta)}{S_N(\varepsilon_N(\beta))} \right|$ a.e. para algún $(i_0, j_0) \in (W_M \sim T)$. Es decir, $s(\beta) + \epsilon > S_N(\varepsilon_N(\beta))$, a.e. $\forall \beta \in \mathcal{B}_0 \times [-d, d]$.

De la misma manera se demuestra con (A.1.13) que $s(\beta) - \epsilon < S_N(\varepsilon_N(\beta))$, a.e. $\forall \beta \in \mathcal{B}_0 \times [-d, d]$ y para todo $N > N_1$.

Por lo tanto, $|S_N(\varepsilon_N(\beta)) - s(\beta)| \leq \epsilon$, a.e. $\forall \beta \in \mathcal{B}_0 \times [-d, d]$ y $\forall N > \max(N_0, N_1)$, es decir,

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

Luego,

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

como se quería y el lema queda demostrado. □

Demostración Lema 4.3.4:

Dado $\beta = (\phi, \mu)$ con $\phi \in \mathcal{B}_0$, llamemos $\vartheta_{i,j}(\beta) = \varepsilon_{i,j}(\beta) - \varepsilon_{i,j}(\phi, 0)$. De la definición de $\varepsilon_{i,j}(\beta)$ que aparece en la ecuación (A.1.1) se tiene que

$$\vartheta_{i,j}(\beta) = \mu\xi = -\mu \left(1 - \sum_{(k,l) \in T} \phi_{k,l} \right), \quad \forall (i,j) \in (W_M \sim T). \quad (\text{A.1.16})$$

Además, es fácil ver que

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

Usando la compacidad de \mathcal{B}_0 , existe $\delta > 0$ y $K_1 > 0$ tal que para todo $\phi \in \mathcal{B}_0$,

$$\delta \leq 1 - \sum_{(k,l) \in T} \phi_{k,l} \leq K_1. \quad (\text{A.1.17})$$

Luego, de (A.1.16) y usando (A.1.17) se tiene

$$\inf_{\phi \in \mathcal{B}_0} |\vartheta_{i,j}(\beta)| = \inf_{\phi \in \mathcal{B}_0} \left| \mu \left(1 - \sum_{(k,l) \in T} \phi_{k,l} \right) \right| \geq \frac{\delta}{2} |\mu| \quad (\text{A.1.18})$$

y por Lema 4.3.1 (i) obtenemos

$$\sup_{\phi \in \mathcal{B}_0} |\varepsilon_{i,j}(\phi, 0)| \leq W_{i,j}^0 \quad (\text{A.1.19})$$

donde $W^0 = \{W_{i,j}^0\}_{(i,j) \in \mathbb{Z}^2}$ es un proceso estacionario.

Como $\sup \rho_1 > b$ y $\lim_{x \rightarrow \infty} \rho_1(|x|) = \sup \rho_1$, existe $k_0 > 0$ y $\lambda > 1$ tal que para todo x que satisface $|x| \geq k_0$ se cumple que

$$\rho_1(x) \geq \lambda b. \quad (\text{A.1.20})$$

Como $\{W_{i,j}^0\}_{(i,j) \in \mathbb{Z}^2}$ es estrictamente estacionario, para cada (i,j) , las variables $W_{i,j}^0$ poseen la misma distribución por lo que existe un m tal que

$$P(W_{i,j}^0 < m/2) > \frac{1}{\lambda}. \quad (\text{A.1.21})$$

Definimos k por

$$k = \max \left(\frac{m}{s_0 + 1}, k_0 \right) \quad (\text{A.1.22})$$

y sea d constante tal que $d \geq \max(4(s_0 + 1)k/\delta, |\mu_0|)$. Entonces usando (A.1.18) obtenemos que

$$\inf_{\phi \in \mathcal{B}_0, |\mu| > d} |\vartheta_{i,j}(\beta)| \geq \frac{\delta}{2} d \geq 2(s_0 + 1)k. \quad (\text{A.1.23})$$

Como ρ_1 satisface **P1**, se cumple que

$$\inf_{\phi \in \mathcal{B}_0, |\mu| > d} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0 + 1} \right) \geq \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\inf_{\phi \in \mathcal{B}_0, |\mu| > d} \left| \frac{\varepsilon_{i,j}(\beta)}{s_0 + 1} \right| \right) I(A_{i,j}) \quad (\text{A.1.24})$$

donde $A_{i,j} = \{W_{i,j}^0 < m/2\}$ e $I(A_{i,j})$ denota la función indicadora del conjunto $A_{i,j}$. De la ecuación (A.1.19) y la definición de $\vartheta_{i,j}$, se puede escribir

$$|\varepsilon_{i,j}(\beta)| \geq |\vartheta_{i,j}(\beta)| - |\varepsilon_{i,j}(\phi, 0)| \geq |\vartheta_{i,j}(\beta)| - W_{i,j}^0. \quad (\text{A.1.25})$$

Entonces de (A.1.22), (A.1.25) y (A.1.23) obtenemos que

$$A_{i,j} \subset \{W_{i,j}^0 < k \cdot (s_0 + 1)\} \subset \left\{ \inf_{|\mu| > d, \phi \in \mathcal{B}_0} |\varepsilon_{i,j}(\beta)| > k \cdot (s_0 + 1) \right\}. \quad (\text{A.1.26})$$

Como $\rho_1 \geq 0$, y $\rho_1(|u|)$ es no decreciente, de (A.1.26), obtenemos

$$\begin{aligned} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\inf_{\phi \in \mathcal{B}_0, |\mu| > d} \left| \frac{\varepsilon_{i,j}(\beta)}{s_0 + 1} \right| \right) I(A_{i,j}) &\geq \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1(k) I(A_{i,j}) \\ &= \frac{\rho_1(k)}{N} \sum_{(i,j) \in (W_M \sim T)} I(A_{i,j}). \end{aligned} \quad (\text{A.1.27})$$

Como $W_{i,j}^0$ es ergódico y estacionario, por el Teorema Ergódico (Guyon (1995)) y por (A.1.21) se tiene

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} I(A_{i,j}) = E(I(A_{i,j})) = P(A_{i,j}) > \frac{1}{\lambda} \quad (\text{A.1.28})$$

en \mathcal{L}^2 y, por lo tanto, converge *a.e.*. Entonces, de (A.1.24), (A.1.27) se tiene que

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

Tomando límite inferior y de la expresión (A.1.28) se tiene que

$$\liminf_{N \rightarrow \infty} \inf_{\phi \in \mathcal{B}_0, |\mu| > d} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0 + 1} \right) > \frac{\rho_1(k)}{\lambda} \quad a.e.. \quad (\text{A.1.29})$$

Además, de (A.1.20) y (A.1.22) tenemos que

$$\frac{\rho_1(k)}{\lambda} \geq b, \quad (\text{A.1.30})$$

por lo que, de (A.1.29) y (A.1.30):

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

Luego, para $N \gg 0$,

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

Por argumentos similares a los usados en el Lema 4.3.3 se tiene que

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

Por lo tanto,

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

Tomando límite inferior obtenemos

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

Luego, el lema queda demostrado. □

Lema A.1.1. Sea $C = \{f : \mathcal{B} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}\}$, $A \in M^{(M+1) \times (M+1)}(C)$ donde $A = [v_{i,j}(\beta, \sigma)]_{0 \leq i,j \leq M}$ tal que $\forall (i,j) \in (W_M \sim T)$,

$$v_{i,j}(\beta, \sigma) = \mu f_1^{i,j}(\phi) + \sigma f_2^{i,j}(\beta, \sigma) + \sum_{(k,l) \in T} \phi_{k,l} v_{i-k, j-l}(\beta, \sigma) \quad (\text{A.1.31})$$

donde $\forall (i,j) \in (W_M \sim T)$, $f_1^{i,j}$ es un polinomio en ϕ y $f_2^{i,j}$ es una función acotada sobre un compacto.

Entonces, $\forall (i,j) \in (W_M \sim T)$,

$$v_{i,j}(\beta, \sigma) = \mu g_1^{i,j}(\phi) + \sigma g_2^{i,j}(\beta, \sigma) + \sum_{(m,n) \in W_M^c} g_{m,n}^{i,j}(\phi) v_{m,n}(\beta, \sigma) \quad (\text{A.1.32})$$

donde $W_M^c = W_M \setminus (W_M \sim T)$ y tal que $\forall (i,j) \in (W_M \sim T)$ y $\forall (m,n) \in W_M^c$, $g_1^{i,j}$ y $g_{m,n}^{i,j}$ son polinomios en ϕ y $\forall (i,j) \in (W_M \sim T)$, $g_2^{i,j}$ es una función acotada sobre un compacto.

Dem.:

A continuación haremos una demostración por inducción en M .

Supongamos que $M = 2$. Entonces $L = 1$ y $(W_M \sim T) = \{(2, 2), (1, 2), (1, 1), (2, 1)\}$ y la matriz A es

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

donde $v_{i,j}(\beta, \sigma)$ es como en (A.1.31). Por (A.1.31), $v_{1,1}(\beta, \sigma)$ ya satisface (A.1.32). Reemplazando la expresión de $v_{1,1}(\beta, \sigma)$ en $v_{1,2}(\beta, \sigma)$ y $v_{2,1}(\beta, \sigma)$, se obtiene que estos satisfacen (A.1.32). De la misma manera, reemplazando las expresiones (A.1.32) de $v_{1,1}(\beta, \sigma)$, $v_{1,2}(\beta, \sigma)$ y $v_{2,1}(\beta, \sigma)$ en $v_{2,2}(\beta, \sigma)$ se tiene que $v_{2,2}(\beta, \sigma)$ cumple (A.1.32). Luego, el lema vale para $M = 2$.

Supongamos que el lema vale para $M = k$, veamos que se cumple para $M = k + 1$.

Sea A la matriz de tamaño $(M + 1) * (M + 1)$:

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

Consideremos ahora la submatriz de tamaño $M * M$: $A_{1,k} = A[1 : (k + 1), 1 : (k + 1)] = [v_{s,t}^1(\beta, \sigma)]_{0 \leq s,t \leq k}$ donde $v_{s,t}^1(\beta, \sigma) = v_{s+1,t+1}(\beta, \sigma)$. Luego, $v_{i,j}^1(\beta, \sigma)$ satisface las condiciones de (A.1.31) y por hipótesis inductiva para todo

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

$$v_{i-1,j-1}^1(\beta, \sigma) = v_{i,j}(\beta, \sigma) = \mu g_{1,1}^{i,j}(\phi) + \sigma g_{1,2}^{i,j}(\beta, \sigma) + \sum_{(m,n) \in W_{1,M-1}^c} g_{1,m,n}^{i,j}(\phi) v_{m,n}(\beta, \sigma) \quad (\text{A.1.33})$$

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

La idea es probar:

(a) $v_{i,j}(\beta, \sigma)$ se escribe como (A.1.32) para todo

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

(b) $v_{i,j}(\beta, \sigma)$ se escribe como (A.1.32) para todo

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

Probemos (a).

I) Tomemos la submatriz de tamaño $M * M$: $A_{2,k} = A[0 : k, 0 : k] = [\tilde{v}_{i,j}(\beta, \sigma)]_{0 \leq i,j \leq k}$ donde $\tilde{v}_{i,j}(\beta, \sigma) = v_{i,j}(\beta, \sigma)$.

Como la matriz $A_{2,k}$ satisface las condiciones del lema, por hipótesis inductiva se tiene para todo $(i, j) \in (W_{2,M-1} \sim T) = \{(m, n) : L \leq m \leq k, L \leq n \leq k\}$

$$\tilde{v}_{i,j}(\beta, \sigma) = v_{i,j}(\beta, \sigma) = \mu g_{2,1}^{i,j}(\phi) + \sigma g_{2,2}^{i,j}(\beta, \sigma) + \sum_{(m,n) \in W_{2,M-1}^c} g_{2,m,n}^{i,j}(\phi) v_{m,n}(\beta, \sigma) \quad (\text{A.1.34})$$

donde $W_{2,M-1}^c = W_{2,M-1} \setminus (W_{2,M-1} \sim T) \subset W_M^c := W_M \setminus (W_M \sim T)$. En particular, $v_{i,j}(\beta, \sigma)$ se escribe como (A.1.34) para todo $(i, j) \in H_2 := \{(L, n) : L \leq n \leq k\} \cup \{(m, L) : L+1 \leq m \leq k\} \subset H_1$.

II) Resta ver que $v_{L,k+1}(\beta, \sigma)$ y $v_{k+1,L}(\beta, \sigma)$ se escriben como (A.1.32). Por definición (A.1.31),

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

donde $V_{L,k+1} = \{(m, n) : 0 \leq m \leq L, k+1-L \leq n \leq k+1, (m, n) \neq (L, k+1)\}$ y ya que $V_{L,k+1} \cap (W_M \sim T) \subset H_2$, entonces por lo visto en (I), $v_{L,k+1}(\beta, \sigma)$ satisface (A.1.32). Del mismo modo se puede ver que $v_{k+1,L}(\beta, \sigma)$ también lo satisface. Luego, (a) queda demostrado.

Probemos (b). Por lo visto en (A.1.33) para todo $(i, j) \in (W_{1,M-1} \sim T)$,

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

Observemos que $W_{1,M-1}^c \cap W_M^c \subset W_M^c$. Además, por lo probado en (a)

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

Luego, se tiene probado (b) y, por lo tanto, el lema queda demostrado. \square

Demostración Lema 4.3.5:

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

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

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

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

De la definición de $\varepsilon_{i,j}^b(\beta, \sigma)$ (ver 2.1.3), se deduce que $v_{i,j}(\beta, \sigma)$ satisface $\forall (i, j) \in (W_M \sim T)$ la ecuación recursiva:

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

Usando el Lema A.1.1, tiene que $\forall (i, j) \in (W_M \sim T)$, $v_{i,j}(\beta, \sigma)$ se escribe como

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

donde

$\forall (i, j)$, $f_1^{i,j}$ es un polinomio,

$\forall (i, j)$ y $\forall (m, n) \in W_M \setminus (W_M \sim T)$ $f_{m,n}^{i,j}$ es un polinomio y

$\forall (i, j)$, $f_2^{i,j}$ es una función acotada sobre conjuntos compactos.

Como \mathcal{B}_0 es compacto, $\mu \in [-d, d]$ y η es acotada, existen C_1 y C tal que

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

Además, como \mathcal{B}_0 es compacto, existe C_3 constante tal que

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

Por lo tanto, existen $C, D > 0$ constantes tal que

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

Luego, el lema queda demostrado. \square

Demostración Lema 4.3.6:

Para demostrar el lema probaremos que existe $\delta > 0$ tal que

$$\frac{1}{N} \sum_{(i, j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i, j}^b(\beta, \sigma)}{s_0 + \delta} \right) > \frac{1}{N} \sum_{(i, j) \in (W_M \sim T)} \rho_1 \left(\frac{\varepsilon_{i, j}^b(\beta, \hat{\sigma}(\phi))}{S_N(\varepsilon_N^b(\beta, \hat{\sigma}(\phi)))} \right) = b \quad a.e. \quad (\text{A.1.35})$$

pues como $\hat{\sigma}(\phi) \leq \hat{\sigma}_Y$ y como $\hat{\sigma}_Y \rightarrow \sigma_Y$ a.e., entonces $\hat{\sigma}(\phi) \leq \sigma_Y$ para todo $N > N_0$ y se tiene

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

Entonces, $\exists (i_0, j_0) \in (W_M \sim T)$ tal que

$$\frac{|\varepsilon_{i_0, j_0}^b(\beta, \hat{\sigma}(\phi))|}{s_0 + \delta} > \frac{|\varepsilon_{i_0, j_0}^b(\beta, \hat{\sigma}(\phi))|}{S_N(\varepsilon_N^b(\beta, \hat{\sigma}(\phi)))} \quad a.e. \quad \forall \beta \in \mathcal{B}_0 \times [-d, d].$$

Luego,

$$S_N(\varepsilon_N^b(\beta, \hat{\sigma}(\phi))) \geq s_0 + \delta$$

y tomando ínfimo se tiene

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

Por lo tanto,

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

Luego el lema quedaría demostrado.

Para probar (A.1.35) veamos dos hechos:

1) Existe $\delta > 0$ tal que

$$\liminf_{N \rightarrow \infty} \sup_{\beta \in \mathcal{B}_0 \times [-d, d], \sigma \leq \sigma_Y} \frac{1}{N} \sum_{(i, j) \in (W_M \sim T)} E \left(\rho_1 \left(\frac{\varepsilon_{i, j}^b(\beta, \sigma)}{s_0 + \delta} \right) \right) \geq b + \delta \quad a.e.. \quad (\text{A.1.36})$$

2)

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

Veamos 1)

Por la definición de $\hat{\sigma}(\phi)$ (ver 2.4.5), $\hat{\sigma}(\phi) \leq \hat{\sigma}_Y$, donde $\hat{\sigma}_Y$ es un estimador robusto de σ_Y tal que $\lim_{N \rightarrow \infty} \hat{\sigma}_Y = \sigma_Y$ a.e.. Usando el Lema 4.3.5 podemos encontrar constantes $C_1 > 0$ y $C_2 > 0$ tal que

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

Como $\lim_{N \rightarrow \infty} \hat{\sigma}_Y = \sigma_Y$ a.e., con probabilidad 1, $\exists N_0$ tal que para $N > N_0$, $\forall (i, j) \in W_M$, existen constantes \tilde{C}_1 y \tilde{C}_2 tal que

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

Sea $D = \tilde{C}_1 \sigma_Y + \tilde{C}_2$ entonces $\forall N > N_0$, $\forall (i, j) \in (W_M \sim T)$ se tiene que

$$\sup_{\beta \in \mathcal{B}_0 \times [-d, d]} |\varepsilon_{i, j}^b(\beta, \sigma) - Y_{i, j}| \leq D \quad \forall \sigma \leq \sigma_Y, \quad (\text{A.1.37})$$

y en particular,

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

Por otra parte, podemos escribir al proceso $\{Y_{i, j}\}$ como $Y_{i, j} = \mu_0 + \varepsilon_{i, j} + v_{i, j}$, donde $v_{i, j}$ es un proceso estacionario que depende de $\varepsilon_{k, l}$, cuando $(k, l) \preceq (i, j)$ y $(k, l) \neq (i, j)$ ($v_{i, j} = \sum_{(k, l) \in T} \phi_{k, l} \varepsilon_{i-k, j-l}$).

Como el proceso de innovaciones ε satisface **P4**, entonces la distribución de $\varepsilon_{i, j}$ es no acotada ($\forall \delta > 0$, $P(|\varepsilon_{i, j}| > \delta) > 0$). Además, como $Y_{i, j}$ no es un ruido blanco tenemos que $v_{i, j}$ también tiene distribución no acotada pues: supongamos que $v_{i, j}$ tiene distribución acotada, entonces $\exists \delta > 0$ tal que

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

Como $\exists M > 0$ tal que $|Y_{i,j}| < M$ (pues $E(|Y_{i,j}|^2) = \sigma_Y^2 < \infty$), entonces

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

Lo cual es absurdo pues $\varepsilon_{i,j}$ tiene distribución no acotada.

Sea

$$u_{i,j}(\beta, \sigma) = \mu_0 + v_{i,j} + (\varepsilon_{i,j}^b(\beta, \sigma) - Y_{i,j}), \quad \forall (i, j) \in (W_M \sim T). \quad (\text{A.1.38})$$

Podemos escribir

$$\varepsilon_{i,j}^b(\beta, \sigma) = Y_{i,j} + (\varepsilon_{i,j}^b(\beta, \sigma) - Y_{i,j}) = Y_{i,j} + u_{i,j}(\beta, \sigma) - \mu_0 - v_{i,j} = \varepsilon_{i,j} + u_{i,j}(\beta, \sigma). \quad (\text{A.1.39})$$

Observe que (A.1.37) y (A.1.38) implican que $\forall \beta \in \mathcal{B}_0 \times [-d, d]$, $\sigma \leq \sigma_Y$ y $\forall N > N_0$ tenemos

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

Por lo tanto, $\forall (i, j) \in (W_M \sim T)$ y $\forall N > N_0$ se tiene

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

Como la distribución de $v_{i,j}$ es no acotada y estacionaria (todas tienen la misma distribución), tenemos que

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

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

De acuerdo a la definición de s_0 , tenemos que $E_{\beta_0}(\rho_1(\varepsilon_{i,j}/s_0)) = b$.

Como vimos en el Lema 4.3.2, si $S(u, q) = E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j} + u}{q} \right) \right)$, para $q \neq 0$ y $u \neq 0$ se cumple $S(0, q) < S(u, q)$, es decir, $E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}}{q} \right) \right) < E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j} + u}{q} \right) \right)$.

En particular, si $q = s_0 \neq 0$ se tiene

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

Esto implica que

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

Luego,

$$(1 - \gamma)E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}}{s_0} \right) \right) + \gamma \inf_{|u| \geq 1} E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j} + u}{s_0} \right) \right) > (1 - \gamma)b + \gamma b = b. \quad (\text{A.1.40})$$

Sea la función F definida por $F(q) = (1 - \gamma)S(0, q) + \gamma \inf_{|u| \geq 1} S(u, q)$.

F es una función decreciente en $|q|$ (por ser $S(p, q)$ decreciente en $|q|$) y continua (por el Teorema de Convergencia Dominada de Lebesgue). Por (A.1.40), $F(s_0) > b$. Luego, $\exists \delta > 0$ tal que

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

es decir,

$$(1 - \gamma)E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}}{s_0 + \delta} \right) \right) + \gamma \inf_{|u| \geq 1} E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j} + u}{s_0 + \delta} \right) \right) \geq b + \delta. \quad (\text{A.1.41})$$

Sean

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

y

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

Como para cada $(i, j) \in (W_M \sim T)$ y $N > N_0$ se tiene que

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

entonces

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

Luego, sumando sobre $(W_M \sim T)$

$$\begin{aligned}
\gamma_N \inf_{|u| \geq 1} h(u) &\leq \frac{1}{N} \sum_{(i,j) \in W_M \sim T} \inf_{\beta \in \mathcal{B}_0 \times [-d,d], \sigma \leq \sigma_Y} I(A_{i,j}) h(u_{i,j}(\beta, \sigma)) \\
&\leq \inf_{\beta \in \mathcal{B}_0 \times [-d,d], \sigma \leq \sigma_Y} \frac{1}{N} \sum_{(i,j) \in W_M \sim T} I(A_{i,j}) h(u_{i,j}(\beta, \sigma)) \\
&\leq \frac{1}{N} \sum_{(i,j) \in W_M \sim T} I(A_{i,j}) h(u_{i,j}(\beta, \sigma)) \quad \forall \beta \in \mathcal{B}_0 \times [-d,d], \sigma \leq \sigma_Y, N > N_0. \quad (\text{A.1.42})
\end{aligned}$$

Además, como

$$\begin{aligned}
h(u) &= E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j} + u}{s_0 + \delta} \right) \right) \\
&= S(u, s_0 + \delta) \\
&\geq S(0, s_0 + \delta) \\
&= E_{\beta_0} \left(\rho_1 \left(\frac{\varepsilon_{i,j}}{s_0 + \delta} \right) \right) \\
&= h(0) \quad \forall u, \quad (\text{A.1.43})
\end{aligned}$$

entonces por (A.1.42) y (A.1.43) se tiene

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

Por lo tanto,

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

y como $\gamma_N \rightarrow \gamma$ a.e. (por Ley de los Grandes Números para procesos ergódicos Guyon (1995)), por

A.1.41 se tiene

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

y luego 1) queda demostrado.

Probemos ahora 2)

Sea

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

Como se vio en la sección 1.4, tomando las σ -álgebras $\mathcal{F}_{i,j} = \sigma(\{R_{k,l} : (k,l) \in V_{i,j}\})$ se satisface que $\mathcal{F}_{k,l} \subseteq \mathcal{F}_{i,j}$ para cada $(k,l) \preceq (i,j)$. Luego, aplicando la misma idea que para el caso de series de tiempo en Muler et al (2009), resulta que $\{R_{i,j}(\beta, \sigma), \mathcal{F}_{i,j}\}$ es una sucesión diferencia martingala.

Nos pongamos en las condiciones de la Ley de los Grandes Números para diferencias martingalas (Quang and Van Huan (2010)): Sea la sucesión $\{b_{(i,j)}\}$ dada por $b_{(i,j)} = (i - L + 1)(j - L + 1)$. Esta sucesión satisface que $\Delta b_{(i,j)} = 0$, $\forall (i,j) \in \mathbb{Z}^2$ y $b_{(i,j)} \rightarrow \infty$ cuando $(i,j) \rightarrow \infty$ (con ambos órdenes: \preceq y \preceq). Además,

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

Luego, por dicho teorema, se tiene

$$\frac{1}{b_{(M,M)}} \sum_{(0,0) \preceq (i,j) \preceq (M,M)} R_{i,j}(\beta, \sigma) = \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} R_{i,j}(\beta, \sigma) \rightarrow 0 \quad a.e.. \quad (\text{A.1.44})$$

y 2) queda demostrado.

Como $R_{i,j}(\beta, \sigma)$ es continua y usando argumentos de compacidad, $\forall \epsilon > 0$, podemos encontrar $(\beta_l, \sigma_l, \delta_l)$, $1 \leq l \leq m_o$ con $\beta_l \in \mathcal{B}_0 \times [-d, d]$, $\sigma_l \leq \sigma_Y$, tal que si definimos

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

obtenemos que $\mathcal{B}_0 \times [-d, d] \times [0, \sigma_Y] \subset \cup_{l=1}^{m_0} V_l$ (cubrimiento finito de $\mathcal{B}_0 \times [-d, d] \times [0, \sigma_Y]$) y

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

Luego,

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

Tomando límite superior en esta última desigualdad y por A.1.44 se tiene que

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

y como esto vale para todo $\epsilon > 0$, se obtiene

$$\limsup_{N \rightarrow \infty} \sup_{\beta \in \mathcal{B}_0 \times [-d, d], \sigma \leq \sigma_Y} \left| \frac{1}{N} \sum_{(i, j) \in (W_M \sim T)} R_{i, j}(\beta, \sigma) \right| = 0 \quad a.e.. \quad (\text{A.1.45})$$

Por último, por (A.1.36) y (A.1.45) se tiene que a.e.,

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

Luego, tomando límite inferior

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

lo cuál implica que $\forall \beta \in \mathcal{B}_0 \times [-d, d], \sigma \leq \sigma_Y$,

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

que es lo que queríamos probar en (A.1.35) y el lema queda demostrado. \square

Demstración Lema 4.3.7:

Por lo visto en el Lema 4.3.5,

$$\begin{aligned}
|v_{i,j}(\beta, \sigma)| &= |\varepsilon_{i,j}^b(\beta, \hat{\sigma}(\phi)) - Y_{i,j}| \\
&= \left| \mu f_1^{i,j}(\phi) + \hat{\sigma}(\phi) f_2^{i,j}(\beta, \hat{\sigma}(\phi)) - \sum_{(m,n) \in W_M \setminus (W_M \sim T)} f_{m,n}^{i,j}(\phi) Y_{m,n} \right| \\
&\geq |\mu| |f_1^{i,j}(\phi)| - \left| -\hat{\sigma}(\phi) f_2^{i,j}(\beta, \hat{\sigma}(\phi)) + \sum_{(m,n) \in W_M \setminus (W_M \sim T)} f_{m,n}^{i,j}(\phi) Y_{m,n} \right| \\
&\geq |\mu| |f_1^{i,j}(\phi)| - \hat{\sigma}(\phi) |f_2^{i,j}(\beta, \hat{\sigma}(\phi))| - \left| \sum_{(m,n) \in W_M \setminus (W_M \sim T)} f_{m,n}^{i,j}(\phi) Y_{m,n} \right|.
\end{aligned}$$

Luego,

$$|\varepsilon_{i,j}^b(\beta, \hat{\sigma}(\phi))| \geq |\mu| |f_1^{i,j}(\phi)| - \hat{\sigma}(\phi) |f_2^{i,j}(\beta, \hat{\sigma}(\phi))| - \left| \sum_{(m,n) \in W_M \setminus (W_M \sim T)} f_{m,n}^{i,j}(\phi) Y_{m,n} \right| - |Y_{i,j}|. \quad (\text{A.1.46})$$

Como $\forall (m, n) \in W_M \setminus (W_M \sim T)$ y $\forall (i, j) \in (W_M \sim T)$, $f_{m,n}^{i,j}$ son polinomios sobre \mathcal{B}_0 conjunto

compacto,

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

Como $0 < \hat{\sigma}(\phi) \leq \hat{\sigma}_Y$ y $f_2^{i,j}$ son funciones acotadas sobre conjuntos compactos se tiene

$$\sup_{\phi \in \mathcal{B}_0} \hat{\sigma}(\phi) |f_2^{i,j}(\beta, \hat{\sigma}(\phi))| \leq \hat{\sigma}_Y \cdot C,$$

entonces

$$\inf_{\phi \in \mathcal{B}_0} \left(- \left| \sum_{(m,n) \in W_M \setminus (W_M \sim T)} f_{m,n}^{i,j}(\phi) Y_{m,n} \right| - \hat{\sigma}(\phi) |f_2^{i,j}(\beta, \hat{\sigma}(\phi))| \right) \geq -D_1 - \hat{\sigma}_Y \cdot C. \quad (\text{A.1.47})$$

Además, como \mathcal{B}_0 es compacto, $\forall \phi \in \mathcal{B}_0$ existe $\epsilon > 0$ tal que $\epsilon \leq f_1^{i,j}(\phi)$. Luego,

$$\inf_{\phi \in \mathcal{B}_0} |\mu| f_1^{i,j}(\phi) \geq \frac{\epsilon}{2} |\mu|. \quad (\text{A.1.48})$$

Por lo tanto, tomando ínfimo en (A.1.46) y usando las cotas encontradas en (A.1.48) y (A.1.47) tenemos

$$\begin{aligned} \inf_{\phi \in \mathcal{B}_0} |\varepsilon_{i,j}^b(\beta, \hat{\sigma}(\phi))| &\geq \inf_{\phi \in \mathcal{B}_0} |\mu| |f_1^{i,j}(\phi)| - |Y_{i,j}| \\ &+ \inf_{\phi \in \mathcal{B}_0} \left(- \left| \sum_{(m,n) \in W_M \setminus (W_M \sim T)} f_{m,n}^{i,j}(\phi) Y_{m,n} \right| - \hat{\sigma}(\phi) |f_2^{i,j}(\beta, \hat{\sigma}(\phi))| \right) \\ &\geq \frac{\epsilon}{2} |\mu| - |Y_{i,j}| - D_1 - \hat{\sigma}_Y \cdot C. \end{aligned} \quad (\text{A.1.49})$$

Como $\sup \rho_1 > b$ (por hipótesis del Teorema 4.3.2) y $\lim_{n \rightarrow \infty} \rho_1(|x|) = \sup \rho_1$, existe k_0 y $\lambda > 1$ tal que $\forall |x| \geq k_0$,

$$\rho_1(x) \geq \lambda b. \quad (\text{A.1.50})$$

Además, como $\lim_{n \rightarrow \infty} \hat{\sigma}_Y = \sigma_Y$ a.e. entonces existen \tilde{D}_1 y \tilde{C} tal que $D_1 + \hat{\sigma}_Y \cdot C \leq \tilde{D}_1 + \sigma_Y \cdot \tilde{C}$.

Sea k_1 tal que el conjunto definido como

$$C_{i,j} = \{|Y_{i,j}| \leq k_1 - \tilde{D}_1 - \tilde{C} \sigma_Y\}$$

satisface $P(C_{i,j}) \geq \frac{1}{\lambda}$.

Sea $k = \max(k_1/(s_0 + 1), k_0)$ y d constante tal que $d > \frac{4k(s_0+1)}{\epsilon}$.

Entonces, de la definición de k y A.1.49, en $C_{i,j}$

$$\begin{aligned} \inf_{\phi \in \mathcal{B}_0, |\mu| > d} |\varepsilon_{i,j}^b(\beta, \hat{\sigma}(\phi))| &\geq \frac{\epsilon}{2}d - k_1 - \tilde{D}_1 - \tilde{C}\sigma_Y + \tilde{D}_1 + \tilde{C}\sigma_Y \\ &= \frac{\epsilon}{2}d - k_1 \\ &> k(s_0 + 1) > k. \end{aligned} \tag{A.1.51}$$

Para todo $\beta = (\phi, \mu)$ tal que $|\mu| > d$, $\phi \in \mathcal{B}_0$ se tiene

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

Como ρ_1 satisface **P1**, $\forall |\mu| > d$, $\phi \in \mathcal{B}_0$ y $\forall (i, j) \in (W_M \sim T)$ tenemos

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

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

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

Luego, tomando ínfimo:

$$\inf_{|\mu| > d, \phi \in \mathcal{B}_0} \frac{1}{N} \sum_{(i,j) \in W_M \sim T} \rho_1 \left(\frac{\varepsilon_{i,j}^b(\beta, \hat{\sigma}(\phi))}{s_0 + 1} \right) \geq \frac{1}{N} \sum_{(i,j) \in W_M \sim T} \rho_1 \left(\inf_{|\mu| > d, \phi \in \mathcal{B}_0} \left| \frac{\varepsilon_{i,j}^b(\beta, \hat{\sigma}(\phi))}{s_0 + 1} \right| \right) I(C_{i,j}). \tag{A.1.52}$$

Además, por (A.1.51) y porque ρ_1 satisface **P1** se tiene

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

entonces, sumando sobre $(W_M \sim T)$:

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

y tomando límite inferior tenemos

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_1 \left(\inf_{\phi \in \mathcal{B}_0, |\mu| > d} |\varepsilon_{i,j}^b(\beta, \hat{\sigma}(\phi))| \right) I(C_{i,j}) \geq \liminf_{N \rightarrow \infty} \frac{\rho_1(k)}{N} \sum_{(i,j) \in (W_M \sim T)} I(C_{i,j}). \quad (\text{A.1.53})$$

Por la ecuación (A.1.50), el hecho de que $\{I(C_{i,j})\}$ es estacionario y ergódico y $E(I(C_{i,j})) = P(C_{i,j}) \geq 1/\lambda$, se tiene por el Teorema Ergódico (1.2.1) que

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

Entonces, por (A.1.52) y (A.1.53) se tiene que

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

Luego, por esto último, para N suficientemente grande y $\forall |\mu| > d, \phi \in \mathcal{B}_0$,

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

entonces, por argumentos similares a los utilizados en el Lema (4.3.3), existen $(i_0, j_0) \in (W_M \sim T)$ tal que $\frac{|\varepsilon_{i_0, j_0}^b(\beta, \hat{\sigma}(\phi))|}{s_0 + 1} > \frac{|\varepsilon_{i_0, j_0}^b(\beta, \hat{\sigma}(\phi))|}{S_N(\varepsilon_N^b(\beta, \hat{\sigma}(\phi)))} \quad a.e.$, por lo que para $N \supset 0$ y $\forall |\mu| > d, \phi \in \mathcal{B}_0$ se tiene que

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

Por lo tanto,

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

y el lema queda demostrado. □

Demostración Lema 4.3.8:

Probemos (i):

Sea $\beta = (\phi, \mu) \neq \beta_0 = (\phi_0, \mu_0)$. Expresemos como en el Lema 4.3.2 $\varepsilon_{i,j}(\beta)$ de la siguiente forma: $\varepsilon_{i,j}(\beta) = \varepsilon_{i,j} + \Delta_{i,j}(\beta)$ donde $\Delta_{i,j}(\beta) = \sum_{(k,l) \in I \setminus \{(0,0)\}} w_{k,l} \varepsilon_{i-k, j-l} + \left(1 - \sum_{(k,l) \in T} \phi_{k,l}\right) \cdot (\mu_0 - \mu)$.

Entonces,

$$\begin{aligned} m(\beta) &= E_{\beta_0} \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \right) \\ &= E_{\beta_0} \left(\rho_2 \left(\frac{\varepsilon_{i,j} + \Delta_{i,j}(\beta)}{s_0} \right) \right). \end{aligned}$$

Además, como $\tilde{S}(p, q) := E_{\beta_0} \left(\rho_2 \left(\frac{\varepsilon_{i,j} + p}{q} \right) \right)$ es decreciente en $|q|$ y $\tilde{S}(0, q) < \tilde{S}(p, q)$ para todo $p \neq 0, q \neq 0$, entonces

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

La igualdad vale si y sólo si $\Delta_{i,j}(\beta) = 0$ a.e.. Por la identificabilidad del modelo AR esto ocurre si y sólo si $\beta = \beta_0$.

Luego, si $\beta \neq \beta_0$ entonces $m(\beta) > m(\beta_0)$. Por lo tanto,

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

Veamos ahora (ii):

La continuidad de $m(\beta)$ es inmediata ya que como $\varepsilon_{i,j}(\beta)$ y ρ_2 son continuas y además ρ_2 es acotada, por el Teorema de Convergencia Dominada de Lebesgue, resulta m continua. □

Demostración Lema 4.3.9:

Por el Teorema de Convergencia Dominada, como ρ_2 es continua y acotada y $\varepsilon_{i,j}(\beta)$ es continua, se tiene que

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

es una función continua con respecto a las dos variables.

Entonces, dado $\epsilon > 0$ y $\beta \in \mathcal{B}_0 \times [-d, d]$, por la continuidad de $M(\beta, v)$ en $v = s_0$ se tiene que existe $0 < \delta(\beta) < s_0$ tal que si $|v - s_0| < \delta(\beta)$ entonces $|M(\beta, v) - M(\beta, s_0)| < \epsilon/2$ para cada $\beta \in \mathcal{B}_0 \times [-d, d]$. Por la compacidad de $\mathcal{B}_0 \times [-d, d]$ se tiene que $\exists \delta > 0$ tal que $|M(\beta, v) - M(\beta, s_0)| \leq \epsilon/2, \forall \beta \in \mathcal{B}_0 \times [-d, d]$ y $\forall v \in [s_0 - \delta, s_0 + \delta]$, y entonces

$$\sup_{\beta \in \mathcal{B}_0 \times [-d, d], v \in [s_0 - \delta, s_0 + \delta]} |M(\beta, v) - M(\beta, s_0)| \leq \epsilon/2. \quad (\text{A.1.54})$$

Consideremos la función $f(y, \beta, v) = \left| \rho_2 \left(\frac{\Phi(B_1, B_2)(y - \mu)}{v} \right) - E_{\beta_0} \left(\rho_2 \left(\frac{\Phi(B_1, B_2)(y - \mu)}{v} \right) \right) \right|$ continua definida en $\mathbb{R} \times C_0$ con $C_0 = \{(\beta, v) : \beta \in \mathcal{B}_0 \times [-d, d], v \in [s_0 - \delta, s_0 + \delta]\}$ compacto. Como $\{Y_{i,j}\}$ es un proceso ergódico, $E_{\beta_0}(f(Y, \beta, v)) = 0$ y $\sup_{(\beta, v) \in C_0} |f(Y, \beta, v)| \leq K$ con K constante, del Lema 3 de Muler and Yohai (2002) se tiene que

$$\lim_{N \rightarrow \infty} \sup_{(\beta, v) \in C_0} \left| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{v} \right) - E_{\beta_0} \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{v} \right) \right) \right| = 0 \quad a.e.. \quad (\text{A.1.55})$$

Por el Teorema 4.3.2, $\lim_{N \rightarrow \infty} s_N^* = s_0$ a.e.. Entonces con probabilidad 1, existe N_0 tal que $\forall N > N_0, s_N^* \in [s_0 - \delta, s_0 + \delta]$ y

$$\sup_{\beta \in \mathcal{B}_0 \times [-d, d]} \left| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*} \right) - E_{\beta_0} \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*} \right) \right) \right| < \epsilon/2 \quad a.e.. \quad (\text{A.1.56})$$

De (A.1.54) y (A.1.56) tenemos que $\forall N > N_0$,

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

Entonces, tomando supremo:

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

y el lema queda demostrado. \square

La demostración del siguiente lema es análoga a la prueba del Lema 4.3.4.

Demostración Lema 4.3.10:

Dado $\beta = (\phi, \mu)$ con $\phi \in \mathcal{B}_0$, sea $\vartheta_{i,j}(\beta) = \varepsilon_{i,j}(\beta) - \varepsilon_{i,j}(\phi, 0)$. De la definición de $\varepsilon_{i,j}(\beta)$ que aparece en la ecuación (A.1.1) se tiene que

$$\vartheta_{i,j}(\beta) = \mu\xi = -\mu \left(1 - \sum_{(k,l) \in T} \phi_{k,l} \right), \quad \forall (i,j) \in (W_M \sim T). \quad (\text{A.1.57})$$

Además, es fácil ver que

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

Usando la compacidad de \mathcal{B}_0 , existe $\tilde{\delta} > 0$ y $K_1 > 0$ tal que para todo $\phi \in \mathcal{B}_0$,

$$\tilde{\delta} \leq 1 - \sum_{(k,l) \in T} \phi_{k,l} \leq K_1. \quad (\text{A.1.58})$$

Luego, de (A.1.57) y usando (A.1.58) se tiene

$$\inf_{\phi \in \mathcal{B}_0} |\vartheta_{i,j}(\beta)| = \inf_{\phi \in \mathcal{B}_0} \left| \mu \left(1 - \sum_{(k,l) \in T} \phi_{k,l} \right) \right| \geq \frac{\tilde{\delta}}{2} |\mu| \quad (\text{A.1.59})$$

y por Lema 4.3.1 (i) obtenemos

$$\sup_{\phi \in \mathcal{B}_0} |\varepsilon_{i,j}(\phi, 0)| \leq W_{i,j}^0, \quad (\text{A.1.60})$$

donde $W^0 = \{W_{i,j}^0\}$ es un proceso estacionario.

Como la innovación ε satisface **P3** y ρ_2 satisface **P1**, entonces $m(\beta_0) = E(\rho_2(\varepsilon_{i,j}/s_0)) < \sup \rho_2$. Entonces, $\exists \delta > 0$ tal que $\sup \rho_2 > m(\beta_0) + \delta$. Además, como $\lim_{x \rightarrow \infty} \rho_2(|x|) = \sup \rho_2$, existe $k_0 > 0$ y $\lambda > 1$ tal que $\forall x$ que satisface $|x| \geq k_0$ se cumple que

$$\rho_2(x) \geq \lambda(m(\beta_0) + \delta). \quad (\text{A.1.61})$$

Como $\{W_{i,j}^0\}$ es estrictamente estacionario, para cada (i,j) , las variables $W_{i,j}^0$ poseen la misma distribución, por lo que existe un valor m tal que

$$P(W_{i,j}^0 < m/2) > \frac{1}{\lambda}. \quad (\text{A.1.62})$$

Definimos k por

$$k = \max\left(\frac{m}{s_N^*}, k_0\right) \quad (\text{A.1.63})$$

y sea d constante tal que $d \geq \max(4s_N^*k/\tilde{\delta}, |\mu_0|)$. Entonces usando (A.1.59) obtenemos que

$$\inf_{\phi \in \mathcal{B}_0, |\mu| > d} |\vartheta_{i,j}(\beta)| \geq \frac{\tilde{\delta}}{2}d \geq 2s_N^*k. \quad (\text{A.1.64})$$

Como ρ_2 satisface **P1**, se cumple que

$$\inf_{\phi \in \mathcal{B}_0, |\mu| > d} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_2\left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*}\right) \geq \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_2\left(\inf_{\phi \in \mathcal{B}_0, |\mu| > d} \left|\frac{\varepsilon_{i,j}(\beta)}{s_N^*}\right|\right) I(A_{i,j}) \quad (\text{A.1.65})$$

donde $A_{i,j} = \{W_{i,j}^0 < m/2\}$ e $I(A_{i,j})$ denota la función indicadora del conjunto $A_{i,j}$. De la ecuación (A.1.60) y la definición de $\vartheta_{i,j}(\beta)$, se puede escribir

$$|\varepsilon_{i,j}(\beta)| \geq |\vartheta_{i,j}(\beta)| - |\varepsilon_{i,j}(\phi, 0)| \geq |\vartheta_{i,j}(\beta)| - W_{i,j}^0. \quad (\text{A.1.66})$$

Entonces de (A.1.63), (A.1.66) y (A.1.64) obtenemos que

$$A_{i,j} \subset \{W_{i,j}^0 < k \cdot s_N^*\} \subset \left\{ \inf_{|\mu| > d, \phi \in \mathcal{B}_0} |\varepsilon_{i,j}(\beta)| > k \cdot s_N^* \right\}. \quad (\text{A.1.67})$$

Como $\rho_2 \geq 0$, y $\rho_2(|u|)$ es no decreciente, de (A.1.67) obtenemos

$$\begin{aligned} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_2\left(\inf_{\phi \in \mathcal{B}_0, |\mu| > d} \left|\frac{\varepsilon_{i,j}(\beta)}{s_N^*}\right|\right) I(A_{i,j}) &\geq \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_2(k) I(A_{i,j}) \\ &= \frac{\rho_2(k)}{N} \sum_{(i,j) \in (W_M \sim T)} I(A_{i,j}). \end{aligned} \quad (\text{A.1.68})$$

Como $W_{i,j}^0$ es ergódico y estacionario, por el Teorema Ergódico (Guyon (1995)) y por (A.1.62) se tiene

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{(i,j) \in W_M \sim T} I(A_{i,j}) = E(I(A_{i,j})) = P(A_{i,j}) > \frac{1}{\lambda} \quad (\text{A.1.69})$$

en \mathcal{L}^2 y, por lo tanto, converge *a.e.*. Entonces, de (A.1.65), (A.1.68) se tiene

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

Tomando límite inferior y de la expresión (A.1.69)

$$\liminf_{N \rightarrow \infty} \inf_{\phi \in \mathcal{B}_0, |\mu| > d} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*} \right) > \frac{\rho_2(k)}{\lambda} \quad a.e.. \quad (\text{A.1.70})$$

Además, de (A.1.61) y (A.1.63) tenemos que

$$\frac{\rho_2(k)}{\lambda} \geq m(\beta_0) + \delta, \quad (\text{A.1.71})$$

por lo que (A.1.70) y (A.1.71):

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

Es decir,

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

y el lema queda demostrado. □

Demostración Lema 4.3.11:

La demostración de este lema es similar a la del lema 4.3.6.

Por el Lema 4.3.5 podemos encontrar constantes $C_1 > 0$ y $C_2 > 0$ tal que $\forall d > 0$ y $\forall \tilde{\sigma} > 0$,

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

Dado $\alpha > 0$ y llamando $D = C_1(s_0 + \alpha) + C_2$, $\forall \sigma \in [0, s_0 + \alpha]$ se tiene

$$\sup_{\beta \in \mathcal{B}_0 \times [-d, d]} |\varepsilon_{i,j}^b(\beta, \sigma) - Y_{i,j}| \leq D. \quad (\text{A.1.72})$$

Por otra parte, podemos escribir al proceso $\{Y_{i,j}\}$ como $Y_{i,j} = \mu_0 + \varepsilon_{i,j} + v_{i,j}$, donde $v_{i,j}$ es un proceso estacionario que depende de $\varepsilon_{k,l}$, $(k, l) \prec (i, j)$ ($v_{i,j} = \sum_{(k,l) \in T} \phi_{k,l} \varepsilon_{i-k, j-l}$).

Como $Y_{i,j}$ no es un ruido blanco y la distribución de $\varepsilon_{i,j}$ es no acotada, tenemos que $v_{i,j}$ también tiene distribución no acotada.

Sea

$$u_{i,j}(\beta, \sigma) = \mu_0 + v_{i,j} + (\varepsilon_{i,j}^b(\beta, \sigma) - Y_{i,j}) \forall (i, j) \in (W_M \sim T). \quad (\text{A.1.73})$$

Podemos escribir

$$\varepsilon_{i,j}^b(\beta, \sigma) = Y_{i,j} + (\varepsilon_{i,j}^b(\beta, \sigma) - Y_{i,j}) = Y_{i,j} + u_{i,j}(\beta, \sigma) - \mu_0 - v_{i,j} = \varepsilon_{i,j} + u_{i,j}(\beta, \sigma). \quad (\text{A.1.74})$$

Por lo tanto, de (A.1.72) y (A.1.73), $\forall (i, j) \in (W_M \sim T)$ tenemos

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

Como la distribución de $v_{i,j}$ es no acotada y $\{v_{i,j}\}$ es estacionario (todas tienen la misma distribución), tenemos que

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

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

De acuerdo a la definición de $m(\beta)$, tenemos que $m(\beta_0) = E_{\beta_0}(\rho_2(\varepsilon_{i,j}/s_0))$. Como vimos en el Lema 4.3.2, para $q \neq 0$ y $u \neq 0$ se cumple $E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}}{q}\right)\right) < E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}+u}{q}\right)\right)$.

En particular, si $q = s_0 \neq 0$ se tiene

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

Esto implica que,

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

Luego,

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

Con argumentos similares a los usados en el Lema 4.3.6, podemos encontrar un $\delta > 0$ tal que

$$(1 - \gamma)E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}}{s_0 + \delta}\right)\right) + \gamma \inf_{|u| \geq 1} E_{\beta_0}\left(\rho_2\left(\frac{\varepsilon_{i,j}+u}{s_0 + \delta}\right)\right) \geq m(\beta_0) + \delta. \quad (\text{A.1.75})$$

Sean

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

y

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

De manera análoga a la usada en el lema 4.3.6, para todo $\beta \in \mathcal{B}_0 \times [-d, d]$, $0 \leq \sigma \leq s_0 + \alpha$,

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

Por lo tanto,

$$\gamma_N \inf_{|u| \geq 1} h(u) + (1 - \gamma_N)h(0) \leq \inf_{\beta \in \mathcal{B}_0 \times [-d, d], \sigma \leq s_0 + \alpha} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} h(u_{i,j}(\beta, \sigma)). \quad (\text{A.1.76})$$

Debido a que $\gamma_N \rightarrow \gamma$ *a.e.* (por Ley de los Grandes Números para procesos ergódicos (Guyon (1995))), y por (A.1.76) y (A.1.75) y tomando límite inferior se tiene

$$\begin{aligned} \liminf_{N \rightarrow \infty} \inf_{\beta \in \mathcal{B}_0 \times [-d, d], \sigma \leq s_0 + \alpha} \frac{1}{N} \sum_{(i,j) \in W_M \sim T} h(u_{i,j}(\beta, \sigma)) &\geq \gamma \inf_{|u| \geq 1} h(u) + (1 - \gamma)h(0) \\ &\geq m(\beta_0) + \delta \quad \text{a.e.} \end{aligned} \quad (\text{A.1.77})$$

Por otra parte, sea

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

Tomando las σ -álgebras $\mathcal{F}_{i,j} = \sigma(\{R_{k,l} : (k, l) \in V_{i,j}\})$ resulta que $\{R_{i,j}(\beta, \sigma), \mathcal{F}_{i,j}\}$ es una sucesión diferencia martingala. Luego, por la Ley de los Grandes Números para Martingalas 1.4.1 se satisface que

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

Por la compacidad de $\mathcal{B}_0 \times [-d, d] \times [0, s_0 + \alpha]$, se obtiene

$$\limsup_{N \rightarrow \infty} \sup_{\beta \in \mathcal{B}_0 \times [-d, d], \sigma \leq s_0 + \alpha} \left| \frac{1}{N} \sum_{(i,j) \in W_M \sim T} [R_{i,j}(\beta, \sigma)] \right| = 0 \quad \text{a.e.} \quad (\text{A.1.78})$$

Por último, por (A.1.77) y (A.1.78) se tiene que para $N \gg 0$, a.e.

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

Luego, tomando límite inferior

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

es decir, para $N > N_1$, $\forall \sigma \leq s_0 + \alpha$ se tiene

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

Como por el Teorema 4.3.2 $s_N^* \rightarrow s_0$, $\exists N_2$ tal que $\forall N > N_2$, $s_N^* < s_0 + \min(\alpha, \delta)$. Luego, $0 \leq s_N^* \leq s_0 + \alpha$ y debido a que ρ_2 satisface **P1** se tiene que

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

Por lo tanto,

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

Luego el lema queda demostrado. □

Demostración Lema 4.3.12:

La demostración de este lema es similar a la del lema 4.3.7.

De la misma manera en que se procede en el Lema 4.3.7, existen constantes positivas ϵ , D_1 y C tal que para algún $\alpha > 0$, si $0 < \sigma \leq s_0 + \alpha$ se tiene

$$\inf_{\phi \in \mathcal{B}_0} |\varepsilon_{i,j}^b(\beta, \sigma)| \geq \frac{\epsilon}{2} |\mu| - D_1 - (s_0 + \alpha) \cdot C - |Y_{i,j}|. \quad (\text{A.1.79})$$

Como $\sup \rho_2 > m(\beta_0)$, existe $\delta > 0$ tal que $\sup \rho_2 > m(\beta_0) + \delta$. Luego, como $\lim_{n \rightarrow \infty} \rho_2(|x|) = \sup \rho_2$, existe k_0 y $\lambda > 1$ tal que $\forall |x| \geq k_0$,

$$\rho_2(x) \geq \lambda(m(\beta_0) + \delta). \quad (\text{A.1.80})$$

Sea k_1 tal que el conjunto

$$C_{i,j} = \{|Y_{i,j}| \leq k_1 - D_1 - C(s_0 + \alpha)\}$$

satisface $P(C_{i,j}) \geq \frac{1}{\lambda}$.

Sea

$$k = \max(k_1/s_N^*, k_0)$$

y sea d constante tal que

$$d > \max\left(\frac{4ks_N^*}{\epsilon}, |\mu_0|\right).$$

Entonces, de la definición de k y A.1.79, en $C_{i,j}$ sucede que

$$\begin{aligned} \inf_{\phi \in \mathcal{B}_0, |\mu| > d} |\varepsilon_{i,j}^b(\beta, \sigma)| &\geq \frac{\epsilon}{2} d - D_1 - C(s_0 + \alpha) - k_1 + D_1 + C(s_0 + \alpha) \\ &= \frac{\epsilon}{2} d - k_1 \\ &> k \cdot s_N^*. \end{aligned} \quad (\text{A.1.81})$$

Para todo $\beta = (\phi, \mu)$ tal que $|\mu| > d$, $\phi \in \mathcal{B}_0$, $0 \leq \sigma \leq s_0 + \alpha$ se tiene

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

Como ρ_2 satisface **P1**, $\forall |\mu| > d, \phi \in \mathcal{B}_0$,

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

Entonces, $\forall |\mu| > d$ y $\phi \in \mathcal{B}_0$, sumando sobre $(W_M \sim T)$:

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

luego, tomado ínfimo se tiene

$$\inf_{|\mu| > d, \phi \in \mathcal{B}_0} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_2 \left(\frac{\varepsilon_{i,j}^b(\beta, \sigma)}{s_N^*} \right) \geq \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_2 \left(\inf_{|\mu| > d, \phi \in \mathcal{B}_0} \left| \frac{\varepsilon_{i,j}^b(\beta, \sigma)}{s_N^*} \right| \right) I(C_{i,j}) \quad (\text{A.1.82})$$

para $\sigma \leq s_0 + \alpha$. Además, por la ecuación (A.1.81) y debido a que ρ_2 satisface **P1** se tiene

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

Entonces, sumando sobre $(W_M \sim T)$:

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

y tomando límite inferior:

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \rho_2 \left(\inf_{\phi \in \mathcal{B}_0, |\mu| > d} \left| \frac{\varepsilon_{i,j}^b(\beta, \sigma)}{s_N^*} \right| \right) I(C_{i,j}) \geq \liminf_{N \rightarrow \infty} \frac{\rho_2(k)}{N} \sum_{(i,j) \in (W_M \sim T)} I(C_{i,j}). \quad (\text{A.1.83})$$

Debido a (A.1.80), a la definición de k y como $\{I(C_{i,j})\}$ es un proceso estacionario y ergódico con $E(I(C_{i,j})) = P(C_{i,j}) \geq 1/\lambda$, se tiene por el Teorema Ergódico (1.2.1) que

$$\begin{aligned} \liminf_{N \rightarrow \infty} \frac{\rho_2(k)}{N} \sum_{(i,j) \in (W_M \sim T)} I(C_{i,j}) &= \rho_2(k) P(C_{i,j}) \\ &\geq \lambda(m(\beta_0) + \delta) \frac{1}{\lambda} \\ &= m(\beta_0) + \delta \quad a.e.. \end{aligned} \quad (\text{A.1.84})$$

Entonces, por (A.1.82), (A.1.83) y (A.1.84) se tiene que para $0 \leq \sigma \leq s_0 + \alpha$:

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

Además, como $s_N^* \rightarrow s_0$ a.e., para $N \gg 0$, $s_N^* \leq s_0 + \alpha$ y

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

es decir,

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

luego el lema queda demostrado. □

Demostración Lema 4.4.1:

Por lo visto en (4.2.15)

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

Como $E \left(\frac{1}{s_0^2} \psi_2^2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \right) < \infty$ pues ψ_2 es acotada y $E \left[\nabla (\varepsilon_{i,j}(\beta)) \nabla (\varepsilon_{i,j}(\beta))^t \right] < \infty$ pues $E(Y^2) < \infty$, entonces $V_0 < \infty$.

A continuación probaremos que dado un vector columna $\mathbf{c} \neq 0$ en $\mathbb{R}^{(L+1)^2}$ tenemos que

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

junto con $\mathcal{F}_{i,j} = \sigma(\{Z_{s,t} : (s,t) \prec (i,j)\})$ es una sucesión diferencia martingala estacionaria:

1) Debido a que ψ_2 es acotada, $E(Y^2) < \infty$ y (4.2.5) se tiene:

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

2) Por la observación (1.4.1), basta ver que $E(Z_{i,j}|Z_{s,t}) = 0$ si $(s,t) \prec (i,j)$. Debido a (4.2.5), a que $\nabla(\varepsilon_{i,j}(\beta_0))$ es función de $Z_{s,t}$, a que $\psi_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right)$ es independiente de $Z_{s,t}$ y a (4.2.12):

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

Por lo tanto, $\{Z_{i,j}, \mathcal{F}_{i,j}\}$ es una sucesión diferencia martingala. Veamos que es estacionaria.

a) Debido a (4.2.13):

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

b)

$$\begin{aligned}
Cov(Z_{i,j}, Z_{i+l,j+m}) &= E(Z_{i,j} \cdot Z_{i+l,j+m}) \\
&= E \left(\sum_{k=1}^{(L+1)^2} c_k \nabla_k \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \cdot \sum_{r=1}^{(L+1)^2} c_r \nabla_r \left(\rho_2 \left(\frac{\varepsilon_{i+l,j+m}(\beta_0)}{s_0} \right) \right) \right) \\
&= \sum_{k=1}^{(L+1)^2} \sum_{r=1}^{(L+1)^2} c_k c_r E \left(\nabla_k \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \cdot \nabla_r \left(\rho_2 \left(\frac{\varepsilon_{i+l,j+m}(\beta_0)}{s_0} \right) \right) \right) \\
&= \sum_{k=1}^{(L+1)^2} \sum_{r=1}^{(L+1)^2} \frac{c_k c_r}{s_0^2} E \left(\psi_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \psi_2 \left(\frac{\varepsilon_{i+l,j+m}(\beta_0)}{s_0} \right) \nabla_k (\varepsilon_{i,j}(\beta_0)) \nabla_r (\varepsilon_{i+l,j+m}(\beta_0)) \right) \\
&= \sum_{k=1}^{(L+1)^2} A_k. \tag{A.1.85}
\end{aligned}$$

▪ Si $l \neq 0$ o $m \neq 0 \Rightarrow A_k = 0 \forall k = 1, \dots, (L+1)^2$ por (4.2.12).

▪ Si $l = 0$ y $m = 0$, (A.1.85) es igual a

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

lo cuál es independiente de i, j pues $Y_{i,j}$ es estacionaria.

Además

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

Luego, por el Teorema Central del Límite para Martingalas (Teorema 1.4.2) se tiene:

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

es decir,

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

lo cual, por el Teorema 29.4 de Billingsley (2013), implica que

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

□

Demostración Lema 4.4.2:

Por lo probado en la propiedad 1, (4.2.5) podemos escribir

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

Definimos para $0 \leq v \leq 1$, $1 \leq k \leq (L+1)^2$ a las funciones $A_{N,k}(v)$ como,

$$A_{N,k}(v) = \frac{1}{\sqrt{N}} \sum_{(i,j) \in (W_M \sim T)} \psi_2\left(\frac{\varepsilon_{i,j}(\beta_0)}{(0,5+v)s_0}\right) \cdot \nabla_k(\varepsilon_{i,j}(\beta_0)).$$

Como del Teorema 4.3.2 $\lim_{N \rightarrow \infty} s_N^* = s_0$ a.e., más aún, la convergencia es en probabilidad. Para probar este lema es suficiente probar que $A_{N,k}(v)$, $1 \leq k \leq (L+1)^2$, son acotadas en probabilidad (tight).

Usando el Teorema 12.3 de Billingsley (2013) es suficiente probar las siguientes 2 condiciones,

- (i) $A_{N,k}(0)$ es acotada en probabilidad.
- (ii) Para cualquier $0 \leq v_1 \leq v_2$ y cualquier $\lambda > 0$ tenemos que existe una constante c_1 tal que

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

Probemos (ii). Podemos escribir para $1 \leq k \leq (L+1)^2$,

$$G(a, v) = \psi_2 \left(\frac{a}{(0,5 + v)s_0} \right).$$

Entonces,

$$\begin{aligned} & E((A_{N,k}(v_2) - A_{N,k}(v_1))^2) \tag{A.1.86} \\ &= \frac{1}{N} E \left(\left(\sum_{(i,j) \in (W_M \sim T)} (G(\varepsilon_{i,j}, v_2) - G(\varepsilon_{i,j}, v_1)) \nabla_k (\varepsilon_{i,j}(\beta_0)) \right)^2 \right) \\ &= \frac{1}{N} \left(\sum_{(i,j) \in (W_M \sim T)} E(B_{i,j}^2 C_{i,j}^2) + \sum_{(i,j) \in (W_M \sim T)} \sum_{(l,m) \in (W_M \sim T) | (l,m) \neq (i,j)} E(B_{i,j} C_{l,m} B_{l,m} C_{i,j}) \right) \end{aligned}$$

donde

$$B_{i,j} = \psi_2 \left(\frac{\varepsilon_{i,j}}{(0,5 + v_2)s_0} \right) - \psi_2 \left(\frac{\varepsilon_{i,j}}{(0,5 + v_1)s_0} \right) \tag{A.1.87}$$

y

$$C_{i,j} = \nabla_k (\varepsilon_{i,j}(\beta_0)). \tag{A.1.88}$$

Veamos que ocurre con el segundo sumando:

Sea $\tilde{\mathbf{Y}}_{i,j} = (Y_{i-1,j}, Y_{i-1,j-1}, Y_{i,j-1}, Y_{i-2,j}, Y_{i-2,j-1}, Y_{i-2,j-2}, Y_{i-1,j-2}, Y_{i,j-2}, \dots)$ el vector con los pasados de $Y_{i,j}$. Si $l \leq i$ y $m \leq j$ pero $(l, m) \neq (i, j)$,

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

Como $B_{l,m}$ depende de $Y_{l,m} \in \tilde{\mathbf{Y}}_{i,j}$:

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

Como $C_{i,j} \in \tilde{\mathbf{Y}}_{i,j}$:

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

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

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

Como $B_{i,j}$ depende de $\varepsilon_{i,j}$, $B_{i,j}$ no depende de $\tilde{\mathbf{Y}}_{i,j}$ entonces

$$E(B_{i,j}|\tilde{\mathbf{Y}}_{i,j}) = E(B_{i,j}) = 0. \quad (\text{A.1.89})$$

Veamos ahora como queda el primer sumando:

Como $B_{i,j}$ es independiente de $C_{i,j}$,

$$E(B_{i,j}^2 C_{i,j}^2) = E(B_{i,j}^2)E(C_{i,j}^2). \quad (\text{A.1.90})$$

Por (A.1.86), (A.1.87), (A.1.88), (A.1.89) y (A.1.90) se obtiene

$$E((A_{N,k}(v_2) - A_{N,k}(v_1))^2) \quad (\text{A.1.91})$$

$$\begin{aligned} &= \frac{1}{N} \left(\sum_{(i,j) \in (W_n \sim T)} E(B_{i,j}^2)E(C_{i,j}^2) \right) \\ &= E \left(\psi_2 \left(\frac{\varepsilon_{i,j}}{(0,5 + v_2)s_0} \right) - \psi_2 \left(\frac{\varepsilon_{i,j}}{(0,5 + v_1)s_0} \right) \right)^2 \cdot E(\nabla_k(\varepsilon_{i,j}(\beta_0)))^2. \end{aligned}$$

Sea $v_1 < v < v_2$. Entonces, usando el Teorema del Valor Medio obtenemos que

$$\psi_2 \left(\frac{\varepsilon_{i,j}}{(0,5 + v_2)s_0} \right) - \psi_2 \left(\frac{\varepsilon_{i,j}}{(0,5 + v_1)s_0} \right) = \frac{(v_2 - v_1)}{s_0(0,5 + v)^2} \varepsilon_{i,j} \psi_2' \left(\frac{\varepsilon_{i,j}}{(0,5 + v)s_0} \right),$$

entonces,

$$E \left(\psi_2 \left(\frac{\varepsilon_{i,j}}{(0,5 + v_2)s_0} \right) - \psi_2 \left(\frac{\varepsilon_{i,j}}{(0,5 + v_1)s_0} \right) \right)^2 = \frac{(v_2 - v_1)^2}{s_0^2(0,5 + v)^4} E \left(\varepsilon_{i,j} \psi_2' \left(\frac{\varepsilon_{i,j}}{(0,5 + v)s_0} \right) \right)^2.$$

Luego, como ψ_2' es acotada, $\varepsilon_{i,j}$ tiene segundo momento finito y $s_0 > 0$, podemos concluir que existe $k_0 > 0$ tal que

$$E \left(\psi_2 \left(\frac{\varepsilon_{i,j}}{(0,5 + v_2)s_0} \right) - \psi_2 \left(\frac{\varepsilon_{i,j}}{(0,5 + v_1)s_0} \right) \right)^2 \leq k_0(v_2 - v_1)^2. \quad (\text{A.1.92})$$

Además, como $E(Y_{i,j}^2) < \infty$ (pues Y tiene segundo momento finito) resulta que $E(\nabla_k(\varepsilon_{i,j}(\beta_0)))^2 < \infty$ (ver ecuaciones (2.2.2) y (2.2.3) para conocer las entradas del vector $\nabla_k(\varepsilon_{i,j}(\beta_0))$).

Entonces, de (A.1.91) y (A.1.92), existe $c_1 > 0$ tal que

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

Así, (ii) sigue de la desigualdad de Chebyshev.

Veamos (i). Calculemos $E(A_{N,k}(0)^2)$.

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

donde

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

y

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

Veamos que ocurre con el segundo sumando:

De la misma manera en que se probó en la parte (ii), tomemos

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

el vector con los pasados de $Y_{i,j}$. Si $l \leq i$ y $m \leq j$ pero $(l, m) \neq (i, j)$,

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

y como $\tilde{B}_{i,j}$ depende de $\varepsilon_{i,j}$, $\tilde{B}_{i,j}$ no depende de $\tilde{Y}_{i,j}$ entonces

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

Además, como ψ_2 es impar y la distribución de $\varepsilon_{i,j}$ es simétrica, se tiene que $E(\tilde{B}_{i,j}) = 0$. Luego, el segundo sumando se hace cero.

Veamos ahora como queda el primer sumando:

Como $\tilde{B}_{i,j}$ es independiente de $\tilde{C}_{i,j}$,

$$E(\tilde{B}_{i,j}^2 \tilde{C}_{i,j}^2) = E(\tilde{B}_{i,j}^2) E(\tilde{C}_{i,j}^2). \quad (\text{A.1.93})$$

Debido a que la función ψ_2 es acotada, existe una constante $M > 0$ tal que $E(\tilde{B}_{i,j}^2) < M \forall (i, j)$. Además, de la expresión de $\nabla_k (\varepsilon_{i,j}(\beta_0))^2$, se tiene que $E(\tilde{C}_{i,j}^2) \leq \kappa$ donde κ es tal que $\kappa = \max(\sigma_Y^2 + \mu_0^2, \xi_0^2) = \max(\sigma_Y^2 + \mu_0^2, (-1 + \sum_{(k,l) \in T} \phi_{k,l}^0)^2)$. Luego, de la ecuación (A.1.93) se tiene que

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

Por lo que

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

Sea $\epsilon > 0$, entonces existe $\delta = \left(\frac{M \cdot \kappa}{\epsilon}\right)^{1/2}$. Luego, por Chebyshev se tiene que

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

Por lo tanto, $A_{N,k}(0)$ es acotada en probabilidad y el lema queda demostrado.

□

Demostración Lema 4.4.3:

La prueba de (i) es similar a la del Lemma 4.3.9. Veamos como sería:

Por el Teorema de Convergencia Dominada, como $\psi'_2 = \rho''_2$ es continua y acotada y $\varepsilon_{i,j}(\beta)$ es continua, se tiene que

$$\tilde{M}(\beta, v) = E_{\beta_0} \left(\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{v} \right) \right) \right)$$

es una función continua con respecto a las dos variables.

Entonces, dado $\epsilon > 0$ y $\beta \in \mathcal{B}_0 \times [-d, d]$, por la continuidad de $\tilde{M}(\beta, v)$ en $v = s_0$ se tiene que existe $0 < \delta(\beta) < s_0$ tal que si $|v - s_0| < \delta(\beta)$ entonces $|\tilde{M}(\beta, v) - \tilde{M}(\beta, s_0)| < \epsilon/2$ para cada $\beta \in \mathcal{B}_0 \times [-d, d]$. Por la compacidad de $\mathcal{B}_0 \times [-d, d]$ se tiene que $\exists \delta > 0$ tal que $|\tilde{M}(\beta, v) - \tilde{M}(\beta, s_0)| \leq \epsilon/2, \forall \beta \in \mathcal{B}_0 \times [-d, d]$ y $\forall v \in [s_0 - \delta, s_0 + \delta]$, y entonces

$$\sup_{\beta \in \mathcal{B}_0 \times [-d, d], v \in [s_0 - \delta, s_0 + \delta]} |\tilde{M}(\beta, v) - \tilde{M}(\beta, s_0)| \leq \epsilon/2. \quad (\text{A.1.94})$$

Consideremos la función

$$f(y, \beta, v) = \sum_{k=1}^{(L+1)^2} \sum_{l=1}^{(L+1)^2} \left| \nabla_{k,l}^2 \left(\rho_2 \left(\frac{\Phi(B_1, B_2)(y - \mu)}{v} \right) \right) - E_{\beta_0} \left(\nabla_{k,l}^2 \left(\rho_2 \left(\frac{\Phi(B_1, B_2)(y - \mu)}{v} \right) \right) \right) \right|$$

continua definida en $\mathbb{R} \times C_0$ con $C_0 = \{(\beta, v) : \beta \in \mathcal{B}_0 \times [-d, d], v \in [s_0 - \delta, s_0 + \delta]\}$ compacto. Como $\{Y_{i,j}\}$ es un proceso ergódico, $E_{\beta_0}(f(Y, \beta, v)) = 0$ y $\sup_{(\beta, v) \in C_0} |f(Y, \beta, v)| \leq K$ con K constante, del Lema 3 de Muler and Yohai (2002) se tiene que

$$\lim_{N \rightarrow \infty} \sup_{(\beta, v) \in C_0} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} f(Y_{i,j}, \beta, v) = 0 \quad a.e..$$

Luego,

$$\begin{aligned} & \lim_{N \rightarrow \infty} \sup_{(\beta, v) \in C_0} \left\| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{v} \right) \right) - E_{\beta_0} \left(\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{v} \right) \right) \right) \right\| \\ & \leq \lim_{N \rightarrow \infty} \sup_{(\beta, v) \in C_0} \left(\sum_{k=1}^{(L+1)^2} \sum_{l=1}^{(L+1)^2} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \left| \nabla_{k,l}^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{v} \right) \right) - E_{\beta_0} \left(\nabla_{k,l}^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{v} \right) \right) \right) \right|^2 \right)^{1/2} \\ & = \lim_{N \rightarrow \infty} \sup_{(\beta, v) \in C_0} \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} f(Y_{i,j}, \beta, v) = 0 \quad a.e. \end{aligned}$$

Por el Teorema 4.3.2, $\lim_{N \rightarrow \infty} s_N^* = s_0$ a.e.. Entonces con probabilidad 1, existe N_0 tal que

$\forall N > N_0, s_N^* \in [s_0 - \delta, s_0 + \delta]$ y

$$\sup_{\beta \in \mathcal{B}_0 \times [-d, d]} \left\| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{v} \right) \right) - E_{\beta_0} \left(\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{v} \right) \right) \right) \right\| < \epsilon/2 \quad a.e.. \quad (\text{A.1.95})$$

De (A.1.94) y (A.1.95) tenemos que $\forall N > N_0$,

$$\begin{aligned} & \left\| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*} \right) \right) - E_{\beta_0} \left(\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \right) \right) \right\| \\ & \leq \left\| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*} \right) \right) - E_{\beta_0} \left(\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*} \right) \right) \right) \right\| + \\ & + \left\| E_{\beta_0} \left(\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*} \right) \right) \right) - E_{\beta_0} \left(\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \right) \right) \right\| \\ & < \epsilon/2 + \epsilon/2, \quad \forall \beta \in \mathcal{B}_0 \times [-d, d] \quad a.e.. \end{aligned}$$

Entonces, tomando supremo:

$$\sup_{\beta \in \mathcal{B}_0 \times [-d, d]} \left\| \frac{1}{N} \sum_{(i,j) \in (W_M \sim T)} \nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_N^*} \right) \right) - E \left[\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta)}{s_0} \right) \right) \right] \right\| < \epsilon \quad a.e., \quad \forall N > N_0$$

y la parte (i) del lema queda demostrada.

Ahora probaremos (ii). De la propiedad 2 del Capítulo 2, se tiene (4.2.8). Además, del hecho de que $\nabla(\varepsilon_{i,j}(\beta_0))$ y $\nabla^2(\varepsilon_{i,j}(\beta_0))$ dependen de $\tilde{\mathbf{Y}}_{i,j}$ se obtiene que

$$\begin{aligned} E \left(\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \right) &= \frac{1}{s_0^2} E \left(\psi_2' \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \cdot E \left(\nabla(\varepsilon_{i,j}(\beta_0)) \nabla(\varepsilon_{i,j}(\beta_0))^t \right) \\ &+ \frac{1}{s_0} E \left(\psi_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \cdot E \left(\nabla^2(\varepsilon_{i,j}(\beta_0)) \right). \end{aligned}$$

Por la propiedad 3-(a), se tiene que $E(\psi_2(\varepsilon_{i,j}(\beta_0)/s_0)) = 0$. Por lo tanto

$$E \left(\nabla^2 \left(\rho_2 \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \right) = \frac{1}{s_0^2} E \left(\psi_2' \left(\frac{\varepsilon_{i,j}(\beta_0)}{s_0} \right) \right) \cdot E \left(\nabla(\varepsilon_{i,j}(\beta_0)) \nabla(\varepsilon_{i,j}(\beta_0))^t \right) \quad (\text{A.1.96})$$

e (ii) queda demostrado. \square

A.2. Rutinas en R

A.2.1. Experimentos

Simulación de un proceso AR-2D

En el siguiente código de R se simula un proceso AR-2D no contaminado como el que sigue:

$$Y_{i,j} = \phi_1^0 Y_{i-1,j} + \phi_2^0 Y_{i,j-1} + \phi_3^0 Y_{i-1,j-1} + \varepsilon_{i,j} \quad (\text{A.2.1})$$

donde $\varepsilon_{i,j}$ son v.a.i.i.d. con distribución simétrica.

```
ARprocess <- function(N, phi0) {
  # La función simula un proceso autorregresivo bidimensional
  #
  # Argumentos:
  #   N: Número natural.
  #   phi0: Vector en R^3.
  #
  # Salida:
  #   Una matriz N*N que contiene valores simulados del proceso AR-2D.
  matrix.normal <- matrix(rnorm(N*N, mean = 0, sd = 1), nrow = N, ncol = N)
  matrix.ar <- matrix(0, nrow = N, ncol = N)
  matrix.ar[1, ] <- matrix.normal[1, ]
  matrix.ar[, 1] <- matrix.normal[, 1]
  for (i in 2:N) {
    for (j in 2:N) {
      matrix.ar[i, j] = matrix.normal[i, j] + (phi0[1] * matrix.ar[i-1, j])
        + (phi0[2] * matrix.ar[i, j-1]) + (phi0[3] * matrix.ar[i-1, j-1])
    }
  }
  return(matrix.ar)
}
```

Contaminación aditiva

El código R genera una matriz llamada “matrix.cont.”^a partir de una matriz inicial “y” y otra matriz “nu” de la misma dimensión. Los valores de “nu” siguen una distribución $N(\mu, \sigma^2)$. El *prob*,100% de los valores “y” son elegidos al azar. Si $y_{i,j}$ es seleccionado, entonces es reemplazado por $y_{i,j} + \nu_{i,j}$. Este procedimiento es llamado contaminación aditiva de *prob*,100%.

```

AditCont <- function(y, prob, mu, sigm) {
  # Esta función contamina aditivamente una matriz.
  #
  # Argumentos:
  #   y: Matriz de datos.
  #   prob: Numero entre 0 y 1.
  #   mu: Un número real.
  #   sigm: Un número real positivo.
  #
  # Salida:
  #   Matriz con datos contaminados.
  N <- dim(y)[1]
  matrix.cont <- y
  matrix.bin <- matrix(rbinom(N*N, 1, prob), nrow = N, ncol = N)
  matrix.normal <- matrix(rnorm(N*N, mean = mu, sd = sigm), nrow = N, ncol = N)
  for (i in 1:N) {
    for (j in 1:N) {
      if (matrix.bin[i, j] >= 1) {
        matrix.cont[i, j] = matrix.cont[i, j] + matrix.normal[i, j]
      }
    }
  }
  return(matrix.cont)
}

```

Contaminación por t-Student

El código R genera una matriz llamada “matrix.cont.”^a partir de una matriz inicial “*y*” y otra matriz “matrix.tst” de la misma dimensión. Los valores de “matrix.tst” siguen una distribución t-Student de *dfr* grados de libertad. El *prob*, 100% de los valores de “*y*” son elegidos aleatoriamente. Si $y_{i,j}$ es elegido, entonces es reemplazado por la entrada (i, j) de la matriz “matrix.sts”. Este procedimiento es llamado contaminación de reemplazo por t-Student de *prob*, 100%.

```

TStudCont <- function(y, prob, dfr){
  # Esta función contamina una matriz con contaminación de reemplazo.
  #
  # Argumentos:
  #   y: Matriz de datos.
  #   prob: Numero entre 0 y 1.
  #   dfr: Un número positivo.

```

```

#
# Salida:
#   Matriz con datos contaminados.
N <- dim(y)[1]
matrix.cont <- y
matrix.bin <- matrix(rbinom(N*N, 1, prob),nrow = N,ncol = N)
matrix.tst <- matrix(rt(N*N, df=dfr), nrow = N, ncol = N)
for (i in 1:N) {
  for (j in 1:N) {
    if (matrix.bin[i, j] >= 1) {
      matrix.cont[i, j] = matrix.cont[i, j] + matrix.tst[i, j]
    }
  }
}
return(matrix.cont)
}

```

Contaminación por otro proceso autorregresivo

El código R genera una matriz llamada “matrix.cont.” a partir de una matriz inicial “y” y otra matriz “matrix.ar.2” de la misma dimensión. Los valores de “matrix.ar.2” son generados con un proceso AR-2D con un vector de parámetros $\phi \in \mathbb{R}^3$. El *prob*, 100% de los valores de “y” son elegidos al azar. Si $y_{i,j}$ es seleccionado, entonces es reemplazado por la entrada (i, j) de la matriz “matrix.ar.2”. Este procedimiento es llamado contaminación de reemplazo por otro proceso AR-2D al *prob*, 100%.

```

ReplARCont <- function(y, prob, sigm, phi2){
  # Esta función contamina una matriz con contaminación de reemplazo.
  # Argumentos:
  #   y: Matriz de datos.
  #   prob: Numero entre 0 y 1.
  #   sigm: Un número real positivo.
  #   phi2: Vector de parámetros.
  # Salida:
  #   Matriz con datos contaminados.
N <- dim(y)[1]
matrix.cont <- y
matrix.bin <- matrix(rbinom(N*N, 1, prob), nrow = N, ncol = N)
matrix.normal <- matrix(rnorm(N*N, mean=0, sd=sigm), nrow = N, ncol = N)
matrix.ar.2 <- matrix(0, nrow = N, ncol = N)
matrix.ar.2[1, ] <- matrix.normal[1, ]

```

```

matrix.ar.2[, 1] <- matrix.normal[, 1]
for (i in 2:N) {
  for (j in 2:N) {
    matrix.ar.2[i, j] = matrix.normal[i, j] + (phi2[1] * matrix.ar.2[i-1, j])
    + (phi2[3] * matrix.ar.2[i, j-1]) + (phi2[3] * matrix.ar.2[i-1, j-1])
  }
}
for (i in 1:N) {
  for (j in 1:N) {
    if (matrix.bin[i, j] >= 1) {
      matrix.cont[i, j] = matrix.ar.2[i, j]
    }
  }
}
return(matrix.cont)
}

```

Contaminación por ruido blanco

El código R genera una matriz llamada “matrix.cont.” a partir de una matriz inicial “y” y otra matriz “matrix.normal” de la misma dimensión. Los valores de “matrix.normal” siguen una distribución normal $N(\mu, \sigma^2)$. El *prob*, 100% de los valores de “y” son elegidos al azar. Si $y_{i,j}$ es seleccionado, entonces es reemplazado por la entrada (i, j) de la matriz “matrix.normal”. Este procedimiento es llamado contaminación por ruido blanco al *prob*, 100%.

```

ReplWNCCont <- function(y, prob, mu, sigm){
  # Esta función contamina una matriz con contaminación de reemplazo.
  #
  # Argumentos:
  # y: Matriz de datos.
  # prob: Numero entre 0 y 1.
  # mu: Un número real.
  # sigm: Un número real positivo.
  #
  # Salida:
  # Matriz con los datos contaminados.
  N <- dim(y)[1]
  matrix.cont <- y
  matrix.bin <- matrix(rbinom(N*N, 1, prob), nrow = N, ncol = N)
  matrix.normal <- matrix(rnorm(N*N, mean = mu, sd = sigm), nrow = N, ncol = N)

```



```

for (i in 1:N) {
  for (j in 1:N) {
    if (matrix.bin[i, j] >= 1) {
      matrix.cont[i, j] = matrix.normal[i, j]
    }
  }
}
return(matrix.cont)
}

```

Estimación BMM 2D

En el próximo código, se calcula la estimación BMM 2D de los parámetros del modelo (A.2.1):

```

library(minpack.lm)
library(nleqslv)
source("contaminations.R")
LSFunction <- function(x) {
  # Esta función estima los parámetros de un modelo AR-2D de orden 1 con tres parámetros
  #por el método de Mínimos Cuadrados.
  #
  # Argumentos:
  # x: Matriz de datos.
  #
  # Salida:
  # Vector de parámetros estimados.
  n = dim(x)[1]
  m = dim(x)[2]
  x.cuadrada = x ^ 2
  a1 = x.cuadrada[1:(n-1), 2:m]
  a2 = x.cuadrada[2:n, 1:(m-1)]
  a3 = x.cuadrada[1:(n-1), 1:(m-1)]
  b1 = x[2:n, 2:m] * x[1:(n-1), 2:m]
  b2 = x[2:n, 1:(m-1)] * x[1:(n-1), 2:m]
  b3 = x[1:(n-1), 1:(m-1)] * x[1:(n-1), 2:m]
  c1 = x[2:n, 2:m] * x[2:n, 1:(m-1)]
  c2 = x[1:(n-1), 1:(m-1)] * x[2:n, 1:(m-1)]
  d = x[2:n, 2:m] * x[1:(n-1), 1:(m-1)]
  aux1 <- c(sum(a1), sum(b2), sum(b3), sum(b2), sum(a2), sum(c2), sum(b3),
            sum(c2), sum(a3))

```

```
A <- matrix(aux1, nrow = 3, ncol = 3)
aux2 <- c(sum(b1), sum(c1), sum(d))
B <- matrix(aux2, nrow = 3, ncol = 1)
ls <- solve(A, B)
return(ls)
}
Rho2Function <- function(a) {
  # Este proceso define la función rho2.
  #
  # Argumentos:
  #   a: Número real.
  #
  # Salida:
  #   Valor que asume la función rho2 en a.
  abs.a = abs(a)
  if (abs.a <= 2) {
    v = 0.5 * (a * a)
  } else if ((2 < abs.a) && (abs.a <= 3)) {
    a2 = a * a
    a4 = a2 * a2
    a6 = a2 * a4
    a8 = a4 * a4
    v = 0.002 * a8 - 0.052 * a6 + 0.432 * a4 - 0.972 * a2 + 1.792
  } else {
    v = 3.25
  }
  return(v)
}
Rho1Function <- function(c) {
  # Este proceso define la función rho1.
  #
  # Argumentos:
  #   c: Número real.
  #
  # Salida:
  #   Valor que asume la función rho1 en c.
  a = (c / 0.405)
  abs.a = abs(a)
  if (abs.a <= 2) {
```

```
v = 0.5 * (a * a)
} else if ((2 < abs.a) && (abs.a <= 3)) {
  a2 = a * a
  a4 = a2 * a2
  a6 = a2 * a4
  a8 = a4 * a4
  v = 0.002 * a8 - 0.052 * a6 + 0.432 * a4 - 0.972 * a2 + 1.792
} else {
  v = 3.25
}
return(v)
}
EtaFunction <- function(y) {
  # Este proceso define la función eta.
  #
  # Argumentos:
  # y: Número real.
  #
  # Salida:
  # Valor que asume la función eta en y.
  abs.y = abs(y)
  if (abs.y <= 2) {
    v = y
  } else {
    if ((2 < abs.y) && (abs.y <= 3)) {
      y2 = y * y
      y3 = y2 * y
      y5 = y3 * y2
      y7 = y5 * y2
      v = 0.016 * (y7) - 0.312 * (y5) + 1.728 * (y3) - 1.944 * (y)
    } else {
      v = 0
    }
  }
  return(v)
}
#
kConstantB <- 3.25 / 2
#
```

```

PhiFunction <- function(s, u) {
  # Función para obtener una estimación de la escala sigma.
  #
  # Argumento:
  #   s: Número real.
  #   u: Vector en  $R^{\{(n-1)*(m-1)\}}$ .
  #
  # Salida:
  # Valor que asume la función PhiFunction en (s, u).
  n = dim(u)[1] + 1
  m = dim(u)[2] + 1
  u <- u / s
  r <- sapply(u, Rho1Function)
  p <- (sum(r) / ((n - 1) * (m - 1))) - kConstantB
  return(p)
}

SigmaFunction <- function(y) {
  # Función que estima la escala.
  #
  # Argumento:
  #   y: Matriz de datos.
  #
  # Salida:
  # Valor que asume la función definida en y.
  sigma <- (median(abs(y)) / 0.6745) ^ 2
  return(sigma)
}

ResiduesAR <- function(x, epsilon, beta) {
  # Función que calcula los residuos de el modelo AR-2D de tres parámetros.
  #
  # Argumento:
  #   x: Matriz de datos n*n generada por un proceso AR-2D con parámetros beta.
  #   epsilon: Matriz n*n con distribución normal.
  #   beta: Vector en  $R^3$ .
  #
  # Salida:
  # Matriz (n-1)*(n-1) con los residuos del modelo AR-2D.
  n = dim(x)[1]
  m = dim(x)[2]

```

```

res <- matrix(0, n, m)
res[1, ] <- epsilon[1, ]
res[, 1] <- epsilon[, 1]
for (i in 2:n) {
  for (j in 2:m) {
    res[i, j] = x[i, j] - ((beta[1] * x[i-1, j]) + (beta[2] * x[i, j-1])
      + (beta[3] * x[i-1, j-1]))
  }
}
resf <- res[2:n, 2:m]
return(resf)
}

ResiduosBIPAR <- function(x, epsilon, beta) {
  # Función que calcula los residuos del modelo BIP-AR-2D de tres parámetros.
  #
  # Argumento:
  # x: Matriz de datos n*n generadapor un proceso AR-2D con parámetros beta.
  # epsilon: Matriz n*n con distribución normal.
  # beta: Vector en R^3.
  #
  # Salida:
  # Matriz (n-1)*(n-1) con los residuos del modelo BIP-AR-2D.
  n = dim(x)[1]
  m = dim(x)[2]
  res <- matrix(0, n, m)
  res[1, ] <- epsilon[1, ]
  res[, 1] <- epsilon[, 1]
  scale.value = SigmaFunction(epsilon)
  for (i in 2:n){
    for (j in 2:m){
      res[i, j] = x[i, j] - ((beta[1] * x[i-1, j]) + (beta[2] * x[i, j-1])
        + (beta[3] * x[i-1, j-1])) - (scale.value * ((beta[1] * EtaFunction(res[i-1, j]
          / scale.value)) + (beta[2] * EtaFunction(res[i, j-1] / scale.value))
            + (beta[3] * EtaFunction(res[i-1, j-1] / scale.value))))
        + (beta[1] * res[i-1, j]) + (beta[2] * res[i, j-1]) + (beta[3] * res[i-1, j-1])
    }
  }
  resf <- res[2:n, 2:m]
  return(resf)
}

```

```

}
#
BetaSEstimationAR <- function(x, epsilon, beta) {
  # Esta función objetivo es usada para obtener, en la primera etapa, una S-estimación
  # de los parámetros con los residuos del modelo AR-2D puro:
  #
  # Argumento:
  #   x: Matriz de datos n*n generada por un proceso AR-2D con parámetros beta.
  #   epsilon: Matriz n*n con distribución normal.
  #   beta: Vector en R^3.
  #
  # Salida:
  #   Vector en R^{(n-1)*(n-1)}.
  n = dim(x)[1]
  m = dim(x)[2]
  k <- 1.483 * median(abs(epsilon))
  residue.AR.output = ResiduesAR(x, epsilon, beta)
  s <- uniroot(PhiFunction, c(k/10, k*10), tol = 0.0001, u = residue.AR.output)[[1]]
  rho1.partial <- sapply(residue.AR.output / s, Rho1Function)
  rho1.partial <- sapply(rho1.partial, sqrt)
  r <- rho1.partial * s / (sqrt((n - 1) * (m - 1) * kConstantB))
  return(r)
}
#
BetaSEstimationBIPAR <- function(x, epsilon, beta) {
  # Esta función objetivo es usada para obtener, en la primera etapa, una S-estimación
  # de los parámetros con los residuos del modelo BIP-AR 2D.
  #
  # Argumento:
  #   x: Matriz de datos n*n generada por un proceso AR-2D con parámetros beta.
  #   epsilon: Matriz n*n con distribución normal.
  #   beta: Vector en R^3.
  #
  # Salida:
  #   Vector en R^{(n-1)*(n-1)}.
  n = dim(x)[1]
  m = dim(x)[2]
  k <- 1.483 * median(abs(epsilon))
  residue.BIPAR.output = ResiduesBIPAR(x, epsilon, beta)

```

```

sB <- uniroot(PhiFunction, c(k/10, k*10), tol = 0.0001, u = residue.BIPAR.output)[[1]]
rho1.partial.BIP <- sapply(residue.BIPAR.output/sB, Rho1Function)
rho1.partial.BIP <- sapply(rho1.partial.BIP, sqrt)
r <- rho1.partial.BIP * sB / (sqrt((n - 1) * (m - 1) * kConstantB))
return(r)
}
#
BetaMEstimationAR <- function(x, epsilon, beta, s) {
  # Esta función objetivo es usada para obtener, en la segunda etapa, una M-estimación
  # de los parámetros con los residuos del modelo Ar-2D puro.
  #
  # Argumento:
  # x: Matriz de datos n*n generada por un proceso AR-2D con parámetros beta.
  # epsilon: Matriz n*n con distribución normal.
  # beta: Vector en R^3.
  # s: Número real positivo.
  #
  # Salida:
  # Vector en R^{(n-1)*(n-1)}.
  n = dim(x)[1]
  m = dim(x)[2]
  rho2.partial <- sapply(ResiduesAR(x, epsilon, beta)/s, Rho2Function)
  / ((n - 1) * (m - 1))
  r <- sapply(rho2.partial, sqrt)
  return(r)
}
#
BetaMEstimationBIPAR <- function(x, epsilon, beta, s) {
  # Esta función objetivo es usada para obtener, en la segunda etapa, una M-estimación
  # de los parámetros con los residuos del modelo BIP-AR 2D.
  #
  # Argumento:
  # x: Matriz de datos n*n generada por un proceso AR-2D con parámetros beta.
  # epsilon: Matriz n*n con distribución normal.
  # beta: Vector en R^3.
  # s: Número real positivo.
  #
  # Salida:
  # Vector en R^{(n-1)*(n-1)}.

```

```

n = dim(x)[1]
m = dim(x)[2]
rho2.partial.BIP <- sapply(ResiduesBIPAR(x, epsilon, beta)/s, Rho2Function)
                        / ((n - 1) * (m - 1))
r <- sapply(rho2.partial.BIP, sqrt)
return(r)
}
#####
##Rutina para calcular la BMM-estimación
#
# N: Un número real positivo. El valor N*N es el tamaño de la ventana.
N = 8
# phi0: Vector de parámetros del proceso observado.
phi0 = c(0.15,0.17,0.2)
# Una matriz "Y" de tamaño N*N es generada con el proceso AR-2D observado:
Y = ARprocess(N, phi0)
# prob: Un número entre 0 y 1.
prob = 0.1
# mu: Un número real. Es la media del proceso contaminante.
mu = 0
# sigm: Un número positivo. Es el desvío del proceso contaminante.
sigm = sqrt(50)
# Definición de innovaciones correspondiente a "Y":
matrix.innovations <- matrix(0, nrow = N, ncol = N)
matrix.innovations[1, ] <- Y[1, ]
matrix.innovations[, 1] <- Y[, 1]
for (i in 2:N) {
  for (j in 2:N) {
    matrix.innovations[i, j] = Y[i, j] - (phi0[1] * Y[i-1, j])
    - (phi0[2] * Y[i, j-1]) - (phi0[3] * Y[i-1, j-1])
  }
}
# La matriz "Y" es contaminada aditivamente:
Z.contaminated = AditCont(Y, prob, mu, sigm)
#
BMMEstimation <- function(Z) {
  # Función que obtiene la estimación BMM de los parámetros de "Z".
  #
  # Argumento:

```



```

# Z: Matriz de datos N*N generada por un proceso AR-2D con parámetros beta.
#
# Salida:
# Vector en R^3 que contiene los parámetros estimados por el método BMM.
#
#Desvío estimado:
k1 <- 1.483 * median(abs(matrix.innovations))
#Parámetros iniciales obtenidos por la función LSFunction
par = LSFunction(Z)
#####First Step#####
# Calculo de la S-estimación de los parámetros:
nls.out <- nls.lm(par, lower = NULL, upper = NULL, fn = BetaSEstimationAR,
                 x = Z, epsilon = matrix.innovations, jac = NULL)
nls.out.B <- nls.lm(par, lower = NULL, upper = NULL, fn = BetaSEstimationBIPAR,
                  x = Z, epsilon = matrix.innovations, jac = NULL)
beta.s <- nls.out[[1]]
beta.sB <- nls.out.B[[1]]
MSE <- sum((phi0 - beta.s) ^ 2) / length(phi0)
MSE.B <- sum((phi0 - beta.sB) ^ 2) / length(phi0)
if (MSE >= MSE.B) {
  beta.s.M <- beta.sB
} else {
  beta.s.M <- beta.s
}
# Calculo de la M-estimación de la escala:
s1 <- uniroot(PhiFunction, c(k1/10, k1*10), tol = 0.0001,
             u = ResiduesAR(Z, matrix.innovations, beta.s))[[1]]
s2 <- uniroot(PhiFunction, c(k1/10, k1*10), tol = 0.0001,
             u = ResiduesBIPAR(Z, matrix.innovations, beta.sB))[[1]]
# Calculo de la M-estimación final de la escala:
sigma <- min(s1, s2)
#####Second Step#####
# Calculo de la M-estimación de los parámetros:
nls.out2 <- nls.lm(beta.s.M, lower = NULL, upper = NULL, fn = BetaMEstimationAR,
                  x = Z, epsilon = matrix.innovations, s = sigma)
nls.out2.B <- nls.lm(beta.s.M, lower = NULL, upper = NULL, fn = BetaMEstimationBIPAR,
                   x = Z, epsilon = matrix.innovations, s = sigma)
beta.BMM <- numeric()
sum1 <- sum((BetaMEstimationAR(Z, matrix.innovations, nls.out2[[1]], sigma))^2)

```

```

sum2 <- sum((BetaMEstimationBIPAR(Z, matrix.innovations, nls.out2.B[[1]], sigma))^2)
if (sum1 <= sum2) {
  beta.BMM = nls.out2[[1]]
} else {
  beta.BMM = nls.out2.B[[1]]
}
return(beta.BMM)
}
BMMEstimation(Y)

```

A.2.2. Aplicación

El código R transforma la imagen y (contaminada o no contaminada) en una imagen restaurada BMM-2D, llamada $w1$, usando el Algoritmo 1.

```

library(minpack.lm)
library(readbitmap)
library(ripa)
source("BMM_estimate.R")

#Tamaño de la ventana deslizante:
k = 57

#Lectura de la imagen:
y <- read.bitmap("lenna.bmp")
#Selección de la banda:
y = y[, , 1]
#Definimos el número de ventanas de tamaño k que entran en la imagen:
N <- dim(y)[1]
M <- dim(y)[2]
nfil = N
ncol = M
nvent = (nfil - 1) / (k - 1)
mvent = (ncol - 1) / (k - 1)
nvent = as.integer(nvent)
mvent = as.integer(mvent)
N = nvent * (k - 1) + 1
M = mvent * (k - 1) + 1
nfil <- N
ncol <- M
#La imagen original es recortada tal que todos los bloques de tamaño k entren exactamente:

```

```

y <- y[1:N, 1:M]
#
ResiduesAR2 <- function(x, beta) {
  # Función que calcula los residuos del modelo AR-2D de tres parámetros.
  #
  # Argumento:
  #   x: Matriz de datos n*n generada por un proceso AR-2D con parámetros beta.
  #   beta: Vector en R^3.
  #
  # Salida:
  #   Matriz (n-1)*(n-1) con los residuos del modelo AR-2D.
  n = dim(x)[1]
  m = dim(x)[2]
  res <- matrix(0, n, m)
  for (i in 2:n) {
    for (j in 2:m) {
      res[i, j] = x[i, j] - ((beta[1] * x[i-1, j]) + (beta[2] * x[i, j-1])
        + (beta[3] * x[i-1, j-1]))
    }
  }
  resf <- res[2:n, 2:m]
  return(resf)
}
#
ResiduesBIPAR2 <- function(x, beta) {
  # Función que calcula los residuos del modelo BIP-AR 2D de tres parámetros.
  #
  # Argumento:
  #   x: Matriz de datos n*n generada por un proceso AR-2D con parámetros beta.
  #   beta: Vector en R^3.
  #
  # Salida:
  #   Matriz (n-1)*(n-1) con los residuos del modelo BIP-AR 2D.
  n = dim(x)[1]
  m = dim(x)[2]
  res <- matrix(0, n, m)
  scale.value = SigmaFunction(ResiduesAR2(x, beta))
  for (i in 2:n){
    for (j in 2:m){

```

```

    res[i, j] = x[i, j] - ((beta[1] * x[i-1, j]) + (beta[2] * x[i, j-1])
      + (beta[3] * x[i-1, j-1])) - (scale.value
      * ((beta[1] * EtaFunction(res[i-1, j] / scale.value))
      + (beta[2] * EtaFunction(res[i, j-1] / scale.value))
      + (beta[3] * EtaFunction(res[i-1, j-1] / scale.value))))
    + (beta[1] * res[i-1, j]) + (beta[2] * res[i, j-1]) + (beta[3] * res[i-1, j-1])
  }
}
resf <- res[2:n, 2:m]
return(resf)
}
#
BetaSEstimationAR2 <- function(x, beta) {
  # Esta función objetivo es usada para obtener, en la primera etapa, una S-estimación.
  # de los parámetros con los residuos del modelo AR-2D puro:
  #
  # Argumento:
  # x: Matriz de datos n*n generada por un proceso AR-2D con parámetros beta.
  # beta: Vector en R^3.
  #
  # Salida:
  # Vector en R^{(n-1)*(n-1)}.
  n = dim(x)[1]
  m = dim(x)[2]
  residue.AR.output = ResiduesAR2(x, beta)
  s <- uniroot(PhiFunction, c(1/(10^15), 10*(10^15)), tol = 0.01,
    u = residue.AR.output)[[1]]
  rho1.partial <- sapply(residue.AR.output / s, Rho1Function)
  rho1.partial <- sapply(rho1.partial, sqrt)
  r <- rho1.partial * s / (sqrt((n - 1) * (m - 1) * kConstantB))
  return(r)
}
#
BetaSEstimationBIPAR2 <- function(x, beta) {
  # Esta función es usada para obtener, en la primera etapa, una S-estimación
  # de los parámetros con los residuos del modelo BIP-AR 2D:
  #
  # Argumento:
  # x: Matriz de datos n*n generada por un proceso AR-2D con parámetros beta.

```

```

#   beta: Vector en R^3.
#
# Salida:
#   Vector en R^{(n-1)*(n-1)}.
n = dim(x)[1]
m = dim(x)[2]
residue.BIPAR.output = ResiduesBIPAR2(x, beta)
sB <- uniroot(PhiFunction, c(1/(10^15), 10*(10^15)), tol = 0.01,
             u = residue.BIPAR.output)[[1]]
rho1.partial.BIP <- sapply(residue.BIPAR.output/sB, Rho1Function)
rho1.partial.BIP <- sapply(rho1.partial.BIP, sqrt)
r <- rho1.partial.BIP * sB / (sqrt((n - 1) * (m - 1) * kConstantB))
return(r)
}
#
BetaMEstimationAR2 <- function(x, beta, s) {
  # Esta función objetivo es usada para obtener, en la primera etapa, una M-estimación
  # de los parámetros con los residuos del modelo AR-2D puro:
  #
  # Argumento:
  #   x: Matriz de datos n*n generada por un proceso AR-2D con parámetros beta.
  #   beta: Vector en R^3.
  #   s: Número real positivo.
  #
  # Salida:
  #   Vector en R^{(n-1)*(n-1)}.
n = dim(x)[1]
m = dim(x)[2]
rho2.partial <- sapply(ResiduesAR2(x, beta)/s, Rho2Function) / ((n - 1) * (m - 1))
r <- sapply(rho2.partial, sqrt)
return(r)
}
#
BetaMEstimationBIPAR2 <- function(x, beta, s) {
  # Esta función objetivo es usada para obtener, en la primera etapa, una M-estimación
  # de los parámetros con los residuos del modelo BIP-AR 2D:
  #
  # Argumento:
  #   x: Matriz de datos n*n generada por un proceso AR-2D con parámetros beta.

```

```

# beta: Vector en R^3.
# s: Número real positivo.
#
# Salida:
# Vector en R^{(n-1)*(n-1)}.
n = dim(x)[1]
m = dim(x)[2]
rho2.partial.BIP <- sapply(ResiduesBIPAR2(x, beta)/s, Rho2Function)
                        / ((n - 1) * (m - 1))
r <- sapply(rho2.partial.BIP, sqrt)
return(r)
}
#
# Comienza la reconstrucción de la imagen por el método BMM.
w1 = matrix(0, N, M)
for (ii in 1:nvent) {
  for (jj in 1:mvent) {
    z1ij = y[((ii-1)*(k-1)+1):(ii*(k-1)+1), ((jj-1)*(k-1)+1):(jj*(k-1)+1)]
    mediaij = mean(z1ij)
    z1ij = z1ij - mediaij
    w1ij = z1ij
    w1ijLS = z1ij
    w1ij.1 = z1ij
    w1ij.1B = z1ij
    par <- LSFunction(w1ijLS)
    nls.out <- nls.lm(par, lower = NULL, upper = NULL, fn = BetaSEstimationAR2,
                    x = z1ij, jac = NULL)
    nls.out.B <- nls.lm(par, lower = NULL, upper = NULL,
                      fn = BetaSEstimationBIPAR2, x = z1ij, jac = NULL)
    beta.s <- nls.out[[1]]
    beta.sB <- nls.out.B[[1]]
    for (i in 2:k) {
      for (j in 2:k) {
        w1ij.1[i, j] = beta.s[1] * z1ij[i-1, j] + beta.s[2] * z1ij[i, j-1]
          + beta.s[3] * z1ij[i-1, j-1]
        w1ij.1B[i, j] = beta.sB[1] * z1ij[i-1, j] + beta.sB[2] * z1ij[i, j-1]
          + beta.sB[3] * z1ij[i-1, j-1]
      }
    }
  }
}

```

```

MSE <- sum((zlij - wlij.1) ^ 2) / length(zlij)
MSE.B <- sum((zlij - wlij.1B) ^ 2) / length(zlij)
if (MSE <= MSE.B) {
  beta.sM <- beta.sB
} else {
  beta.sM <- be.s
}
s1 <- uniroot(PhiFunction, c(1/10^15, 10*(10^15)), tol = 0.01,
             u = ResiduesAR2(zlij, beta.s))[[1]]
s2 <- uniroot(PhiFunction, c(1/10^15, 10*(10^15)), tol = 0.01,
             u = ResiduesBIPAR2(zlij, beta.sB))[[1]]
sigma <- min(s1, s2)
nls.out2 <- nls.lm(beta.sM, lower = NULL, upper = NULL,
                  fn = BetaMEstimationAR2, x = zlij, s = sigma, jac = NULL)
nls.out2.B <- nls.lm(beta.sM, lower = NULL, upper = NULL,
                    fn = BetaMEstimationBIPAR2, x = zlij, s = sigma, jac = NULL)
beta.BMM <- numeric()
suma.1 <- sum((BetaMEstimationAR2(zlij, nls.out2[[1]], sigma)) ^ 2)
suma.2 <- sum((BetaMEstimationBIPAR2(zlij, nls.out2.B[[1]], sigma)) ^ 2)
if (suma.1 <= suma.2) {
  beta.BMM = nls.out2[[1]]
} else {
  beta.BMM = nls.out2.B[[1]]
}
for (i in 2:k) {
  for (j in 2:k) {
    wlij[i, j] = beta.BMM[1] * zlij[i-1, j] + beta.BMM[2] * zlij[i, j-1]
    + beta.BMM[3] * zlij[i-1, j-1]
  }
}
wlij = wlij + mediaij
w1[((i-1)*(k-1)+2):(i*(k-1)+1), ((j-1)*(k-1)+2):(j*(k-1)+1)] = wlij[2:k, 2:k]
}
}
# Muestra la imagen reconstruida w1
plot(imagematrix(normalize(w1)))
# Genera la imagen diferencia:
image.segmented <- y-w1
# Muestra la imagen diferencia y-w1

```

```
plot(imagematrix(normalize(y-w1)))
```


Bibliografía

- Alata O, Olivier C (2003) Choice of a 2-d causal autoregressive texture model using information criteria. *Pattern Recognition Letters* 24(9-10):1191–1201
- Allende H, Galbiati J (2004) A non-parametric filter for digital image restoration, using cluster analysis. *Pattern Recognition Letters* 25(8):841–847
- 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(11):1219–1231
- Baran S, Pap G, Van Zuijlen M (2004) Asymptotic inference for a nearly unstable sequence of stationary spatial ar models. *Statistics & probability letters* 69(1):53–61
- Basu S, Reinsel G (1993) Properties of the spatial unilateral first-order arma model. *Advances in applied Probability* 25(3):631–648
- Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)* 36(2):192–225
- Bhandari A, Kumar A, Singh G (2015) Improved knee transfer function and gamma correction based method for contrast and brightness enhancement of satellite image. *AEU-International Journal of Electronics and Communications* 69(2):579–589
- Billingsley P (2013) *Convergence of probability measures*. John Wiley & Sons
- 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(2):141–165
- Chao J, Woyczynski W (2006) *Martingale theory in harmonic analysis and Banach spaces: proceedings of the NSF-CBMS Conference held at the Cleveland State University, Cleveland, Ohio, July 13-17, 1981, vol 939*. Springer

- Chuvieco E (2008) Teledetección ambiental: La observación de la tierra desde el espacio. Tech. rep.
- Cohen G (2016) A clt for multi-dimensional martingale differences in a lexicographic order. *Stochastic Processes and their Applications* 126(5):1503–1510
- Dieudonné J (2013) Foundations of modern analysis. Read Books Ltd
- 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(5):609–628
- Gottardo R, Raftery A, Yee Yeung K, Bumgarner R (2006) Bayesian robust inference for differential gene expression in microarrays with multiple samples. *Biometrics* 62(1):10–18
- Guyon X (1995) Random fields on a network: modeling, statistics, and applications. Springer Science & Business Media
- Huang H, Lee T (2006) Data adaptive median filters for signal and image denoising using a generalized sure criterion. *IEEE Signal Processing Letters* 13(9):561–564
- Huber P (1964) Robust estimation of a location parameter. *The Annals of Mathematical Statistics* 35(1):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(8):1313–1325
- Landi M, Ojeda S, Di Bella C, Salvatierra P, Argañaraz J, Bellis L (2017) Selección de parcelas control para estudios de la dinámica post-incendio: desempeño de rutinas no paramétricas y autorregresivas. *Revista de Teledetección* 49:70–90
- 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(6)
- Lehmann E (2004) Elements of large-sample theory. Springer Science & Business Media
- Maronna R, Martin R, Yohai V (2006) Robust statistics. John Wiley and Sons
- Marquardt D (1963) An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics* 11(2):431–441
- Martori J, Hoberg K (2008) Nuevas técnicas de estadística espacial para la detección de clusters residenciales de población inmigrante. *Scripta Nova-Revista Electronica De Geografia Y Ciencias Sociales* XII(263)
- Muler N, Yohai V (2002) Robust estimates for arch processes. *J Time Ser Anal* 23:341–375

- Muler N, Peña D, Yohai V (2009) Robust estimation for arma models. *The Annals of Statistics* 37(2):816–840
- Ojeda S (1999) Robust ra estimators for bidimensional autoregressive models. Tesis doctoral, Universidad Nacional de Córdoba. Facultad de Matemática, Astronomía y Física
- Ojeda S, Vallejos R, Lucini M (2002) Performance of robust ra estimator for bidimensional autoregressive models. *Journal of Statistical Computation and Simulation* 72(1):47–62
- Ojeda S, Vallejos R, Bustos O (2010) A new image segmentation algorithm with applications to image inpainting. *Computational Statistics and Data Analysis* 54(9):2082–2093
- Ojeda S, Vallejos R, Lamberti P (2012) Measure of similarity between images based on the codispersion coefficient. *Journal of Electronic Imaging* 21(2):023,019
- Ojeda S, Britos G, Vallejos R (2018) An image quality index based on coefficients of spatial association with an application to image fusion. *Spatial Statistics* 23:1–16
- Ponomarenko N, Lukin V, Zelensky A, Egiazarian K, Carli M, Battisti F (2009) Tid2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics* 10(4):30–45
- Ponomarenko N, Jin L, Ieremeiev O, Lukin V, Egiazarian K, Astola J, Vozel B, Chehdi K, Carli M, Battisti F, et al (2015) Image database tid2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication* 30:57–77
- Prastawa M, Bullitt E, Ho S, Gerig G (2004) A brain tumor segmentation framework based on outlier detection. *Medical image analysis* 8(3):275–283
- 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(1):569,759.
- Quintana C, Ojeda S, Tirao G, Valente M (2011) Mammography image detection processing for automatic micro-calcification recognition. *Chilean Journal of Statistics* 2(2)
- Sadabadi M, Shafiee M, Karrari M (2009) Two-dimensional arma model order determination. *ISA transactions* 48(2):247–253
- 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(4):311–320
- Sain S, Cressie N (2007) A spatial model for multivariate lattice data. *Journal of Econometrics* 140(1):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(3):132–139
- Vallejos R (1998) Estimación robusta en procesos autorregresivos bidimensionales. Tesis de maestría, Universidad Técnica Federico Santa María, Departamento de Matemáticas
- Vallejos R, García-Donato G (2006) Bayesian analysis of contaminated quarter plane moving average models. *Journal of Statistical Computation and Simulation* 76(2):131–147
- 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(3):674–682
- Wang Z, Bovik A (2002) Image and multidimensional signal processing-a universal image quality index. *IEEE Signal Processing Letters* 9(3):81–84
- Whittle P (1954) On stationary processes in the plane. *Biometrika* pp 434–449
- Yao Q, Brockwell P (2006) Gaussian maximum likelihood estimation for arma models ii: spatial processes. *Bernoulli* 12(3):403–429
- Yohai V (1985) High breakdown-point and high efficiency robust estimates for regression. Tech. Rep. 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