



Programa de Doctorado en Estadística, Optimización  
y Matemática Aplicada

**Análisis computacional para la comparación de  
estimadores y métodos de resolución de Modelos de  
Ecuaciones Simultáneas**

---

**María Belén Pérez Sánchez**

Directora de la tesis

**Dra. Dña. Carmen Perea Marco**

Codirector de la tesis

**Dr. D. Jose Juan López Espín**

---

UNIVERSIDAD MIGUEL HERNÁNDEZ DE ELCHE

2022



UNIVERSIDAD MIGUEL HERNÁNDEZ DE ELCHE

Programa de Doctorado en Estadística, Optimización y

Matemática Aplicada

ANÁLISIS COMPUTACIONAL PARA LA  
COMPARACIÓN DE ESTIMADORES Y  
MÉTODOS DE RESOLUCIÓN DE  
MODELOS DE ECUACIONES  
SIMULTÁNEAS

María Belén Pérez Sánchez

Tesis presentada para optar al grado de  
Doctora por la Universidad Miguel Hernández de Elche,  
realizada bajo la dirección de Carmen Perea Marco  
y la codirección de Jose Juan López Espín



*Daría todo lo que sé, por la mitad de lo que ignoro.*

*René Descartes*





El Dr. D. *Domingo Morales González*, Coordinador del Programa de Doctorado en **Estadística, Optimización y Matemática Aplicada**

**INFORMA:**

Que Dña. *María Belén Pérez Sánchez* ha realizado bajo la supervisión de nuestro Programa de Doctorado el trabajo titulado **Análisis computacional para la comparación de estimadores y métodos de resolución de modelos de ecuaciones simultáneas** conforme a los términos y condiciones definidos en su Plan de Investigación y de acuerdo al Código de Buenas Prácticas de la Universidad Miguel Hernández de Elche, cumpliendo los objetivos previstos de forma satisfactoria para su defensa pública como tesis doctoral.

Lo que firmo para los efectos oportunos, en Elche a 8 de marzo de 2022

*Dr. D. Domingo Morales González*  
Coordinador del Programa de Doctorado en Estadística, Optimización  
y Matemática Aplicada





La Dra. Dña. *Carmen Perea Marco*, directora y el Dr. D. *Jose Juan López Espín*, codirector de la tesis doctoral titulada **Análisis computacional para la comparación de estimadores y métodos de resolución de modelos de ecuaciones simultáneas**

**INFORMAN:**

Que Dña. *María Belén Pérez Sánchez* ha realizado bajo nuestra supervisión el trabajo titulado **Análisis computacional para la comparación de estimadores y métodos de resolución de modelos de ecuaciones simultáneas** conforme a los términos y condiciones definidos en su Plan de Investigación y de acuerdo al Código de Buenas Prácticas de la Universidad Miguel Hernández de Elche, cumpliendo los objetivos previstos de forma satisfactoria para su defensa pública como tesis doctoral.

Lo que firmamos para los efectos oportunos, en Elche a 8 de marzo de 2022

Directora de la tesis

Dra. Dña. *Carmen Perea Marco*

Codirector de la tesis

Dr. D. *Jose Juan López Espín*





La presente Tesis Doctoral, titulada **Análisis computacional para la comparación de estimadores y métodos de resolución de modelos de ecuaciones simultáneas**, se presenta bajo la modalidad de **tesis convencional con el siguiente indicio de calidad**:

- Belén Pérez-Sánchez, Martín González, Carmen Perea and Jose J. López-Espín. A New Computational Method for Estimating Simultaneous Equations Model using Entropy as a Parameter Criteria. *Mathematics* 2021, 9, 700.  
DOI:10.3390/math9070700.



---

# Agradecimientos

---

En primer lugar deseo expresar mi agradecimiento a los directores de esta tesis doctoral, Dr. Jose Juan López Espín y Dra. Carmen Perea Marco, por la dedicación que han brindado a este trabajo, por el respeto a mis sugerencias e ideas y por la dirección y el rigor de las mismas. Gracias por la confianza que me han ofrecido desde el inicio.

A mis queridos padres, Castora y Pepe y a mis hermanos, Jose, Sonia y Felipe por su apoyo incondicional.

No quisiera que nadie fuese olvidado, por lo que estas líneas de agradecimiento son para todas las personas que me han acompañado durante esta etapa animando a presentar esta tesis. Algunos siguen, otros no, pero todos ellos me han ofrecido a su manera, sus sabios e ideales consejos que tanto he necesitado.

Y para el final, el más importante, un infinito *gracias* para Jose, Lucía y Daniela por su paciencia, tiempo robado y confianza que siempre han manifestado en el buen fin de esta tesis.



---

# Resumen

---

Predecir el efecto que una o varias variables *causan* en otra variable es un problema a resolver muy frecuente en cualquier disciplina. Una de las herramientas estadísticas para abordar este problema es la modelización de las relaciones a través de la regresión, técnica inferencial aplicada en muchas disciplinas científicas y de la vida cotidiana con muy buenos resultados bajo ciertos supuestos que no siempre se pueden asumir. Una situación que invalida la aplicación de un modelo de regresión múltiple es que no todas las relaciones pueden describirse como un modelo uniecuacional, lo que hace necesario contemplar esas relaciones de manera más realista y por lo tanto más compleja. Los *Modelos de Ecuaciones Simultáneas* consiguen modelar relaciones permitiendo la influencia bidireccional de las variables implicadas en el estudio.

Esta tesis se centra en la resolución de modelos de ecuaciones simultáneas. En concreto, se propone un *nuevo método de estimación* y se lleva a cabo un estudio computacional comparativo frente a diferentes métodos de estimación. Así mismo, y ante la necesidad de disponer de un *parameter criteria* para la selección del mejor método de estimación, se investiga y analizan diversos criterios de información y se propone una nueva medida basada en la *entropía*. Las características propias del problema son consideradas a la hora de la selección del método de estimación, proponiendo unos *límites* a la variabilidad del problema y a su dimensión. Además, se investiga la aplicación de las *redes neuronales artificiales* para la resolución del modelo de ecuaciones simultáneas, augurando buenos resultados a la vista de la experimentación llevada a cabo.

Durante la investigación se han generado múltiples modelos de ecuaciones simultáneas de manera aleatoria, han sido estimados por diferentes métodos de estimación y se han calculado diversas medidas de calidad de las estimaciones obtenidas tras múltiples iteraciones. Para la estimación, se han aplicado métodos frecuentistas, como Mínimos Cuadrados en Dos Etapas, y técnicas de inferencia Bayesiana, como Markov Chain Monte Carlo. Además, se propone un nuevo método,  $Bmom_{OPT}$ , basado en la optimización de unos parámetros del Método Bayesiano de Momentos. Para la comparación de los resultados de los estimadores, se han calculado medidas como la *Norma Euclídea* de los parámetros del modelo y sus estimaciones, el *Criterio de información de Akaike*, una nueva medida basada en la entropía,  $H_2(e)$ , y el *tiempo de cómputo*, entre otras. De los resultados se puede extraer que, el nuevo método de estimación propuesto proporciona un menor error en la estimación de los parámetros del modelo, y que la medida propuesta como parameter criteria basada en la entropía puede ser tomada en cuenta como alternativa a los criterios ya existentes. Del estudio también se concluye que sería deseable reducir el tiempo de ejecución utilizado por el método de estimación propuesto.

Finalmente, y ante la necesidad de modelar relaciones no necesariamente lineales, la implementación de la simultaneidad en redes neuronales artificiales deja abierta una nueva línea de investigación de aplicación en aquellos casos en los que, el objetivo no sea la estimación de los coeficientes del modelo sino la obtención de los valores predichos con el modelo.

---

---

# Summary

---

Predicting the effect that one or more variables *causes* on another variable is a very common problem to solve in any discipline. One of the statistical tools to solve this problem is the modeling of relationships through regression, an inferential technique applied in many scientific disciplines and everyday life, that it has very good results under certain assumptions that cannot always be assumed. A situation that invalidates the application of a multiple regression model is not all relationships can be described as a single-equation model. For that, it is necessary to contemplate those relationships more realistic and therefore more complex. The *Simultaneous Equations Models* manage to model relationships allowing the bidirectional influence of the variables involved in the study.

This thesis focuses on solving simultaneous equation models. Specifically, a *new estimation method* is proposed and a comparative computational study is carried out against different estimation methods. Likewise, it is necessary a *parameter criteria* to select the best method of estimation and, for that, several information criteria are investigated and analyzed and a new measure based on *entropy* is proposed. The characteristics of the problem are considered when a method of estimation is selected, proposing a *limits* to the variability and dimension of the problem. In addition, the application of *neural networks* to solve a simultaneous equations model is studied, predicting good results in view of the experimentation carried out.

In this work, multiple simultaneous equations models have been generated randomly, and they have been estimated by different methods of estimation and several quality measures of estimation after multiple iterations have been calculated. For estimation, frequentist methods such as Least Squares Two Step, and Bayesian inference techniques such as Markov Chain Monte Carlo, have been applied. In addition,  $Bmom_{OPT}$ , is proposed as a new method based on the optimization of some parameters of the Bayesian Method of Moments. For the comparison, measures such as the *Euclidean Norm* of the model parameters and their estimates, the *Akaike Information Criterion*, a new measure based on entropy,  $H_2(e)$ , and the *computation time*, have been calculated. From results it can be concluded that, the new proposed method of estimation provides a lower error in the estimated parameters of model. Also, the proposed measure as parameter criteria based on entropy, can be taken into account as an alternative to the criteria already used. It is also concluded from the study that it would be desirable to reduce the execution time used by the proposed estimation method.

Finally, when the relationships of variables are not necessarily linear, the implementation of the simultaneity relationship in *artificial neural networks* leaves open a new line of research. If the objective is to predict the endogenous variable, this way can be a good alternative.

---

---

# Índice general

---

<b>Resumen</b>	<b>III</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Breve introducción histórica . . . . .	2
1.2. Objetivos del presente trabajo . . . . .	3
1.3. Aportaciones y metodología empleada . . . . .	3
1.4. Estructura de la tesis . . . . .	4
<b>2. El MES y Métodos de Estimación</b>	<b>7</b>
2.1. Introducción al Modelo de Ecuaciones Simultáneas . . . . .	7
2.2. Definición del modelo . . . . .	10
2.3. Métodos de estimación . . . . .	11
2.3.1. Mínimos Cuadrados Ordinarios (MCO) . . . . .	11
2.3.2. Mínimos Cuadrados Indirectos (MCI) . . . . .	12
2.3.3. Mínimos Cuadrados en Dos Etapas (MC2E) . . . . .	13
2.3.4. Mínimos Cuadrados en Tres Etapas (MC3E) . . . . .	13
2.3.5. Método Bayesiano de Momentos (BMOM) . . . . .	14

2.3.6.	Enfoque Bayesiano en dos etapas ( <i>Bayes<sub>2E</sub></i> ) . . . . .	15
2.3.7.	Inferencia MCMC . . . . .	15
2.4.	Conclusiones . . . . .	16
<b>3.</b>	<b>Criterio de Información y Entropía</b>	<b>19</b>
3.1.	Conceptos básicos de Teoría de la información. Entropía. . . . .	20
3.2.	Tipos de Entropía . . . . .	22
3.2.1.	Entropía conjunta y entropía condicional . . . . .	23
3.2.2.	Entropía relativa e información mutua . . . . .	25
3.2.3.	Relación entre la entropía y la información mutua . . . . .	26
3.2.4.	Regla de la cadena para entropía, entropía relativa e información mutua . . . . .	26
3.2.5.	La desigualdad de Jensen y sus propiedades . . . . .	27
3.2.6.	La desigualdad de Log Sum y sus aplicaciones . . . . .	29
3.2.7.	Desigualdad de procesamiento de datos . . . . .	29
3.3.	Criterios de información en la obtención de modelos estadísticos . . . . .	30
3.3.1.	El criterio de información de Akaike . . . . .	31
3.3.2.	El criterio de información Bayesiano . . . . .	32
3.3.3.	El Criterio de información de Hannan y Quinn . . . . .	32
3.4.	Criterio de Información de Akaike y Bayesiano aplicado al MES . . . . .	33
3.5.	Conclusiones . . . . .	34
<b>4.</b>	<b>Red Neuronal Artificial y MES No Lineal</b>	<b>35</b>
4.1.	Introducción . . . . .	35
4.1.1.	Estimación del modelo de ecuaciones simultáneas no lineal . . . . .	36

---

---

4.2. Introducción a las redes neuronales artificiales . . . . .	38
4.2.1. Clasificación de las redes neuronales artificiales . . . . .	41
4.2.2. Algoritmos de una RNA . . . . .	42
4.2.3. Principales modelos de RNA . . . . .	44
4.3. Aproximación de las redes neuronales a la modelización econométrica . .	44
4.3.1. Las redes neuronales artificiales como respuesta al modelo de ecuaciones simultáneas no lineales . . . . .	46
4.4. Conclusiones . . . . .	47
<b>5. Nuevo Método de Estimación y Criterio de Información</b>	<b>49</b>
5.1. Nuevo método de estimación: BMOM optimizado . . . . .	50
5.2. Entropía como parámetro de criterio de información . . . . .	51
5.3. Estudio experimental . . . . .	54
5.4. Conclusiones . . . . .	61
<b>6. Método de Estimación según Variabilidad</b>	<b>63</b>
6.1. Introducción . . . . .	63
6.2. Estudio experimental . . . . .	64
6.3. Conclusiones . . . . .	72
<b>7. Resolución de MES mediante RNA</b>	<b>75</b>
7.1. Resolución de un MES mediante una RNA: caso 1 . . . . .	75
7.2. Resolución de un MES mediante una RNA: caso 2 . . . . .	80
7.3. Conclusiones . . . . .	81
<b>8. Ejemplos</b>	<b>83</b>
8.1. Modelo de la Oferta de mano de obra de trabajadoras casadas . . . . .	84

---

8.2. MES para un problema de relación entre hormonas y otras variables . . .	88
8.3. Conclusiones . . . . .	93
<b>9. Conclusiones y Trabajos futuros</b>	<b>95</b>
9.1. Conclusiones . . . . .	95
9.2. Trabajos futuros . . . . .	97
<b>Índice de tablas</b>	<b>97</b>
<b>Índice de figuras</b>	<b>101</b>
<b>Siglas y acrónimos</b>	<b>105</b>
<b>Anexo I</b>	<b>107</b>

---

# CAPÍTULO 1

---

## Introducción

---

Esta tesis se centra en el Modelo de Ecuaciones Simultáneas (MES) como estructura que modela la relación entre variables que permite simultaneidad en ellas y que es una técnica de inferencia utilizada históricamente en el ámbito de la economía. En esta tesis se investiga y desarrolla un nuevo método de estimación así como el análisis y comparación con otros métodos. Además, se estudian los criterios de información y se propone un nuevo parameter criteria basado en la entropía para la elección del mejor método de estimación de un MES, analizado desde el punto de vista de la variabilidad de los datos, para ajustar la decisión según la variabilidad presentada. A pesar de que en muchas situaciones el interés radica en la estimación de los parámetros del modelo, esto es, en estimar el efecto que una variable *ceteris paribus* causa en otra, esto no siempre es imprescindible. Para situaciones en las que sea suficiente la estimación de la variable explicada y no necesaria la de los coeficientes del modelo, es posible modelar las relaciones no necesariamente de manera lineal y es cuando la simultaneidad puede implementarse como una red neuronal artificial.

## 1.1. Breve introducción histórica

El modelo de ecuaciones simultáneas es una técnica utilizada principalmente en la rama de la Econometría. Esta disciplina tuvo un enorme desarrollo en 1930 impulsada con la creación de la Econometric Society <sup>1</sup>, proporcionando instrumentos para abarcar este tipo de problemas. No fue hasta bien avanzados los años treinta cuando se realizaron estudios en Estados Unidos, basados en la estimación de una única ecuación y centrándose en el análisis de la demanda y los mercados en equilibrio. También se llevaron a cabo estudios en Europa en los cuales se centraban en estimar varias ecuaciones que modelizaban el equilibrio de los mercados. En este contexto, Jan Tinbergen concluye un proyecto para la League of Nations, en el que desarrolla por primera vez un modelo de ecuaciones simultáneas para una economía completa permitiendo efectuar estudios sobre teorías alternativas de los ciclos económicos y estimar las principales elasticidades [1]. Uno de los efectos de este trabajo fue su influencia en el programa de investigación de la Cowles Commission, dedicando todos los recursos de la Cowles al estudio y desarrollo de los modelos de ecuaciones simultáneas [2]. El periodo 1945-1975 marca el desarrollo de la econometría moderna, siendo la última década en la que los desarrollos de la econometría empírica son numerosos, sus aplicaciones son múltiples y se abre un gran mercado que abarca tanto el sector público como la empresa privada. Los modelos econométricos son entendidos como un nuevo instrumento útil y válido para la planificación.

Con la crisis económica de los setenta producida por la subida de los precios del petróleo, provocó un cambio en los intereses de los econométricos, al encontrarse con limitaciones que estos grandes modelos macroeconómicos presentaban a la hora de proporcionar soluciones útiles para la creación de políticas económicas, precisamente cuando más se necesitaban, provocando el alejamiento en la investigación de estos problemas. Estos modelos macroeconómicos fueron descartados tanto por su gran tamaño, que era difícil modelar como por el problema de cómputo, aunque los avances informáticos cada vez propiciaban más el uso de estos modelos.

Con el desuso de estos modelos en economía provocado por la crisis, se aceptó la

---

<sup>1</sup><https://www.econometricsociety.org/>

---

resolución de los MES como regresiones individuales, compensando la simplicidad del problema con el error cometido. Esta situación ha cambiado en los últimos tiempos cuando investigadores de otras disciplinas como medicina [3], biología [4], etc., han empezado a utilizar los MES para modelar las relaciones entre sus variables.

## 1.2. Objetivos del presente trabajo

Los objetivos del presente trabajo son:

1. La propuesta de un nuevo método de estimación de Modelos de Ecuaciones Simultáneas (MES) y un estudio comparativo.
2. Determinar un nuevo criterio de información basado en la entropía, alternativo a los ya existentes.
3. Valorar el marco de los datos dentro de los cuales aplicar el método de estimación que proporciona mejores resultados.
4. Resolver un MES a través de una red neuronal artificial (RNA).

Para el primer objetivo se minimiza la función del criterio de información de Akaike de tal manera que los errores sean mínimos. Para el segundo objetivo se estudian las propiedades de la entropía para encontrar una nueva versión que se adapta a las características del problema. El tercer objetivo se basa en varios estudios experimentales que ofrecen información suficiente para la observación del comportamiento de diferentes métodos al controlar la variabilidad de los datos. Finalmente, para el cuarto objetivo, se investigan las diferentes estructuras de RNA para su aplicación en la resolución de un MES.

## 1.3. Aportaciones y metodología empleada

Las aportaciones más importantes y de carácter novedoso obtenidas en este trabajo pueden ordenarse de la siguiente manera:

---

- Una primera aportación es la propuesta de un nuevo método de estimación y la comparación experimental con otros métodos de estimación, tanto clásicos como Bayesianos.
- La segunda aportación es la propuesta de un nuevo criterio de información para la selección de un método de estimación en base a esta medida. Este criterio, basado en la entropía es una alternativa a los criterios existentes para el problema de los MES, que resultaban de adaptaciones de los modelos de regresión uniecuacionales.

Para ambos objetivos, se han simulado un gran número de MES que permiten comparar diferentes medidas de calidad de la estimación así como del tiempo utilizado en la misma.

- La tercera aportación ha sido el estudio experimental para determinar el mejor método de estimación en función de los rangos de variabilidad de las variables endógenas del problema. Esto se ha llevado a cabo simulando MES en los que se ha variado tanto la dimensión en cuanto a variables endógenas, exógenas y tamaño de la muestra así como la varianza de los datos.
- La cuarta aportación es la introducción de las redes neuronales artificiales a la resolución de un MES, cuya aplicación apenas consta en la bibliografía revisada.

## **1.4. Estructura de la tesis**

Esta tesis está estructurada en 9 capítulos. El capítulo 1, como ya se ha visto, hace una introducción para facilitar al lector el abordaje al problema.

El capítulo 2 describe el Modelo de Ecuaciones Simultáneas y se citan algunas de las aplicaciones que hay en la literatura. Se hace una revisión de los diferentes métodos de estimación, clasificados por el tipo de inferencia utilizada, frecuentista o Bayesiana, y por la cantidad de información utilizada en la estimación, métodos de información limitada o información completa. Concretamente, se hace una revisión de la metodología usada por los siguientes métodos: Mínimos Cuadrados Ordinarios, Mínimos Cuadrados Indirectos, Mínimos Cuadrados en Dos Etapas, Mínimos Cuadrados en Tres Etapas, Método

---

de Momentos Bayesiano, inferencia Bayesiana usando una familia de distribuciones conjugadas y Markov Chain Monte Carlo (MCMC).

En el capítulo 3 se introduce el concepto de entropía, tipos y propiedades, haciendo una breve descripción de algunas de ellas. Se relaciona la entropía como medida de complejidad y de calidad de estimación de los errores, de forma similar al Criterio de Información de Akaike y el Criterio de Información Bayesiano, de los que se hace una breve presentación y su aplicación a los MES.

En el capítulo 4 se presenta una revisión de la literatura de los métodos de estimación de MES no lineales así como algunas aplicaciones. En relación a esta modelización no lineal, se introduce el concepto de Red Neuronal Artificial (RNA) y se presentan brevemente algunos de los tipos de RNA más populares.

La propuesta de un nuevo método de estimación de un MES es presentado en el capítulo 5. Este nuevo método, basado en el Método de Momentos Bayesiano, estima los coeficientes del modelo optimizando los parámetros al minimizar los errores en la estimación, (diferencia entre la variable dependiente y su estimación). Existen varios *parameter criteria* que permiten la selección del mejor método de estimación y en este trabajo se ha estudiado y experimentado con nuevas medidas, proponiendo un nuevo criterio de información basado en la entropía para la selección del método con menor error de estimación de un MES.

El capítulo 6 contiene el estudio experimental que permite abordar el problema de cómo la variabilidad de las variables endógenas puede ser determinante para la selección del método de estimación con menor error. Para este estudio se han generado múltiples MES con diferentes dimensiones en variables, tamaño muestral y varianzas de las variables endógenas.

La aplicación de las RNA a los MES se expone en el capítulo 7. Gracias a la flexibilidad de esta técnica de Machine Learning (ML) para entrenar y predecir valores de salida a partir de un conjunto de observaciones, se han entrenado dos RNA y los resultados se han comparado con los obtenidos con el método propuesto en el capítulo 5.

La aplicación del trabajo realizado a problemas reales se recoge en el capítulo 8. La

---

primera de ellas se hace con un problema clásico en Economía que relaciona el salario percibido y las horas trabajadas, entre otras variables, medidas en un conjunto de mujeres trabajadoras casadas que es comparado con los resultados de las estimaciones obtenidas con métodos utilizados en esta tesis [5]. La segunda aplicación se lleva a cabo con datos clínicos. Concretamente se han recopilado variables relativas a niveles hormonales en adolescentes así como variables de tipo clínico y social. Este conjunto de variables ha sido propuesto por expertos en esa materia y en el presente trabajo se ofrece un modelo y una solución a la estimación de los efectos que las hormonas producen en otras cuando se tiene en cuenta que la relación existente entre dichas hormonas es bidireccional o simultánea.

Y por último, en el capítulo 9 se recogen las conclusiones y se definen líneas futuras de trabajo relacionadas y complementarias llevadas a cabo en esta tesis. Además se presentan las publicaciones y aportaciones a congresos más importantes durante la realización de este trabajo.

---

# CAPÍTULO 2

---

## El Modelo de Ecuaciones Simultáneas y Métodos de Estimación

---

En la primera parte de este capítulo se hace una descripción del Modelo de Ecuaciones Simultáneas (MES) y se presentan algunas de sus aplicaciones. En la segunda parte, se describen métodos de estimación, tanto frecuentistas como Bayesianos.

### 2.1. Introducción al Modelo de Ecuaciones Simultáneas

Un MES [6] es un modelo estadístico formado por un conjunto de ecuaciones de regresión que reflejan la simultaneidad entre el conjunto de variables dependientes e independientes del modelo, es decir, se utiliza cuando existe una relación de influencia bidireccional entre ambos tipos de variables. Los métodos de estimación pueden agruparse en dos grandes grupos: por un lado los pertenecientes al enfoque frecuentista o clásico y por otro lado los métodos o técnicas Bayesianas.

Con respecto a la inferencia frecuentista, la estimación de un MES se puede realizar

mediante métodos de información limitada o de información completa. Los métodos de información limitada estiman cada una de las ecuaciones de la forma estructural sin hacer uso de la información contenida en la especificación detallada del resto del modelo, considerando únicamente las variables endógenas y exógenas que se incluyen en cada ecuación. Mínimos Cuadrados Ordinarios (MCO), Mínimos Cuadrados Indirectos (MCI) y Mínimos Cuadrados de Dos Etapas (MC2E) son ejemplos de métodos de información limitada [6]. Los métodos de información completa consideran la estimación conjunta de todo el modelo en la forma estructural. Estos métodos requieren la especificación de todas las ecuaciones y todas ellas deben ser identificadas. En general, son asintóticamente más eficientes que los métodos de información limitada, ya que incorporan toda la información del sistema, con el inconveniente de que si alguna ecuación se especifica incorrectamente, las estimaciones generadas pueden ser inconsistentes con las otras ecuaciones. Ejemplos de este tipo de métodos son Máxima Verosimilitud de Información Completa (MVIC) o Mínimos Cuadrados en Tres Etapas (MC3E) [6].

La metodología Bayesiana proporciona un enfoque probabilístico a la inferencia de los parámetros del modelo. Está basado en la suposición de que las cantidades de interés son gobernadas por distribuciones de probabilidad y que se pueden tomar decisiones óptimas razonando sobre estas probabilidades junto con los datos obtenidos. No utiliza supuestos muestrales, aunque introduce un alto grado de complejidad debido a la especificación previa de la distribución, así como a la obtención de la distribución posterior. Algunas técnicas de inferencia Bayesiana son el Método Bayesiano de Momentos (BMOM) [7] o los métodos utilizados por Chao y Phillips [8], Geweke [9] y Kleibergen y Van Dijk [10].

El reciente desarrollo de las Cadenas de Markov Monte Carlo (MCMC) ha sido clave para hacer posible el cálculo de grandes modelos que requieren integraciones sobre cientos o incluso miles de parámetros desconocidos. El algoritmo Metropolis-Hastings y el muestreo de Gibbs [11] son algunos ejemplos de MCMC.

En cuanto a la elección de una u otra metodología, la literatura se limita a una comparación entre inferencia Bayesiana y frecuentista concluyendo que los métodos Bayesianos funcionan mejor en el caso de una muestra pequeña [7]. Sin embargo, falta realizar un estudio más amplio que permita determinar, bajo diferentes condiciones de tamaño del

---

problema, variabilidad de los datos, etc., qué metodología ofrece mejor resultados a la hora de estimar un MES.

Las aplicaciones de los MES se pueden encontrar principalmente en el marco económico, disciplina que los puso en práctica en los años 30 [1], aunque posteriormente surgieron aplicaciones en áreas muy diversas. Algunos de los ejemplos los podemos encontrar en los efectos de la descentralización fiscal sobre la desigualdad de ingresos regionales en Indonesia se han estudiado utilizando datos a nivel provincial durante el período 2001-2014 [12], o encontrando un MES que relacione el empleo con la salud mental [13], otros han estudiado el impacto del comercio exterior en la eficiencia energética en la industria textil de China [14] o el consumo de energía de biomasa, el crecimiento económico y las emisiones de carbono en África Occidental [4]. Otros estudios se han centrado en comparar los resultados obtenidos por MES con el modelado de regresión lineal en áreas como los fenotipos complejos [3], un modelo para pronosticar la demanda de vuelos y adaptar así la frecuencia de los mismos y la capacidad de los aviones [15] o un MES para modelar prescripciones en atención primaria [16].

En el enfoque Bayesiano, la aplicación del modelado a través de un MES también es amplio: se ha utilizado en ciencias agrícolas [17], o para estudiar el impacto de la información del producto en sitios web de terceros en el mecanismo de retroalimentación entre las recomendaciones de los clientes y las ventas minoristas en Download.com<sup>1</sup> y Amazon.com<sup>2</sup> (01 03 2020), [18]. También se ha utilizado para analizar la interdependencia de la audiencia de un programa de televisión entre los cónyuges [19], para explorar los efectos de la pareja en el comportamiento de los juegos de azar en los casinos [20], para modelar la interacción entre la percepción de riesgo para la salud de las personas y los hábitos de la masticación de betel en Taiwan [21], y para estudiar los efectos del bloqueo repetitivo de la tiroides con yodo en el desarrollo del cerebro y la tiroides en ratas [22].

---

<sup>2</sup><https://www.amazon.com/>

---

## 2.2. Definición del modelo

Consideremos  $m$  variables dependientes o endógenas las cuales dependen de  $k$  variables independientes o exógenas. Supongamos que cada variable endógena puede expresarse como una combinación lineal de las otras variables endógenas, exógenas y de ruido blanco que representa la interferencia estocástica. De esta manera, un MES puede expresarse de la siguiente forma [6]:

$$\begin{aligned}
 y_1 &= B_{1,2}y_2 + B_{1,3}y_3 + \dots + B_{1,m}y_m + \Gamma_{1,1}x_1 + \dots + \Gamma_{1,k}x_k + u_1 \\
 y_2 &= B_{2,1}y_1 + B_{2,3}y_3 + \dots + B_{2,m}y_m + \Gamma_{2,1}x_1 + \dots + \Gamma_{2,k}x_k + u_2 \\
 &\vdots \\
 y_m &= B_{m,1}y_1 + B_{m,2}y_2 + \dots + B_{m,m-1}y_{m-1} + \Gamma_{m,1}x_1 + \dots + \Gamma_{m,k}x_k + u_m.
 \end{aligned} \tag{2.1}$$

Las ecuaciones pueden representarse de la siguiente forma matricial:

$$YB^T + X\Gamma^T + U = 0 \tag{2.2}$$

donde  $B \in \mathbb{R}^{m \times m}$  y  $\Gamma \in \mathbb{R}^{m \times k}$  son matrices de coeficientes,  $Y \in \mathbb{R}^{n \times m}$  es la matriz de variables endógenas,  $X \in \mathbb{R}^{n \times k}$  es la matriz de variables exógenas, y  $U \in \mathbb{R}^{n \times m}$  es la matriz de variables de ruido blanco, siendo  $m$  el número de variables endógenas o dependientes que coincide con el número de ecuaciones,  $k$  el número de variables exógenas o predeterminadas y  $n$  el tamaño de la muestra. Tal y como se observa en la expresión (2.1), las variables endógenas también forman parte del conjunto de las variables explicativas, lo que caracteriza la relación bidireccional o simultánea que existe entre las variables.

El modelo estructural (ecuación (2.2)) se puede expresar también en forma reducida como sigue:

$$Y = X\Pi + V \tag{2.3}$$

con

$$\Pi = -\Gamma^T (B^T)^{-1}, V = U (B^T)^{-1} \quad (2.4)$$

Algunos coeficientes de  $B$  y  $\Gamma$  son cero y conocidos a priori. El número de variables endógenas y exógenas en la  $i$ -ésima ecuación de 2.1 se denota por  $m_i$  y  $k_i$ . Para que un MES pueda resolverse, sus ecuaciones deben estar identificadas, para lo cual, una condición necesaria pero no suficiente, es que el número de variables (endógenas y exógenas) en la ecuación debe ser menor o igual a  $k + 1$ , esto es  $m_i - 1 \leq k - k_i$  (condición de orden). Cuando  $m_i - 1 = k - k_i$ , la ecuación es exactamente identificada y cuando  $m_i - 1 < k - k_i$  es sobreidentificada. Una condición necesaria y suficiente para que una ecuación sea identificable es que el rango de la matriz  $G$  sea igual a  $m - 1$ , siendo  $G$  la matriz formada por los parámetros que las variables excluidas en dicha ecuación tienen asociados en las restantes ecuaciones.

Resolver el modelo es equivalente a obtener la estimación de  $B$  y  $\Gamma$  en (2.2) a partir de una muestra representativa del modelo (un conjunto de valores del conjunto de datos  $X$  y  $Y$ ), en orden de explicar la ecuación que representa la relación entre ambos conjuntos de variables.

## 2.3. Métodos de estimación

Entre los diferentes tipos de estimadores de los parámetros de un MES, dentro del enfoque frecuentista, los métodos MC3E, MC2E, MCO, estimadores de clase  $k$  [23], etc. son algunos ejemplos, y en el enfoque Bayesiano, algoritmos MCMC y diversas distribuciones conjugadas son algunas de estas técnicas. A continuación se realiza una breve revisión a algunos de los métodos más utilizados.

### 2.3.1. Mínimos Cuadrados Ordinarios (MCO)

En este método, la estimación del modelo es realizada ecuación a ecuación, con la ventaja de la simplicidad del cálculo, pero claramente resulta inconsistente, porque las variables endógenas incluidas en cada ecuación están correlacionadas con las perturbaciones.

Para las  $n$  observaciones, los coeficientes distintos de cero en la ecuación  $t$ -ésima son

$$\begin{aligned} Y_t &= Y_t\beta + X_t\Gamma + e_t \\ &= Z_t\delta_t + e_t \end{aligned} \quad (2.5)$$

siendo  $Z_t = [Y_t|X_t]$  y  $\delta_t = [\beta|\Gamma]$

Las  $m$  ecuaciones en forma reducida son

$$Y = X\Pi + V \quad (2.6)$$

La expresión del estimador MCO es  $(X^T X)^{-1} X^T Y$ , cuyo vector resultante es el estimador de  $\Pi$ , que denotamos por  $\hat{\Pi}$ . Con los coeficientes estimados, es posible obtener estimaciones de la variable  $y$  a partir de  $\hat{Y} = X\hat{\Pi}$ .

Aunque se puede afirmar con certeza que este estimador es inconsistente, no es posible determinar la gravedad de este problema. La ventaja de MCO es la simplicidad del cálculo, pero con la alta capacidad computacional existente actualmente, esta ventaja es extremadamente modesta. Otra ventaja es que este método no requiere el estudio de la identificación del modelo, con lo que todos los MES pueden ser estimados.

### 2.3.2. Mínimos Cuadrados Indirectos (MCI)

Este estimador [24] solo es aplicable cuando las ecuaciones son exactamente identificadas, por lo que, según la condición de orden, debe cumplirse  $m_i - 1 + k_i = k$ . MCI estima los coeficientes de la matriz  $\Pi$  de la ecuación (2.3) mediante MCO utilizando todas las variables endógenas como variables dependientes y todas las exógenas como explicativas. Una vez estimada  $\Pi$ , se obtienen los valores de la forma estructural de la ecuación a ser resuelta. Para estimar los parámetros de la forma estructural, MCI resuelve las ecuaciones que se derivan de  $-B_i\Pi = \Gamma_i$  (2.4). El hecho de que la ecuación sea exactamente identificada asegura una única solución a la estimación.

### 2.3.3. Mínimos Cuadrados en Dos Etapas (MC2E)

MC2E es el método de estimación más común [6], desarrollado independientemente por Theil (1953) [25] y Basman (1957) [26]. Sabemos que no se debe usar MCO como estimador en un MES debido a la correlación de las variables endógenas con el término aleatorio y una solución a este problema de correlación la ofrece MC2E. Este método resuelve el sistema en el que previamente ha sustituido las variables endógenas que actúan como explicativas en todas las ecuaciones del sistema por otras variables (muy parecidas a las endógenas originales), y que no están correlacionadas con el término de error.

Este método se llama *dos etapas*, porque en su expresión matemática, se aplica MCO dos veces. En la primera etapa, esas nuevas variables que sustituyen a las variables endógenas que actúan como explicativas, a las que se denominan *proxys*, son calculadas por MCO usando las variables exógenas del modelo como variables independientes. En la segunda etapa, las variables endógenas son sustituidas por las *proxys*, y MCO es aplicado de nuevo obteniendo la estimación de los parámetros. Ambos, MC2E y MCO son estimadores de la clase simple K [23], expresado en (2.7) cuando  $K_1 = K_2 = 0$  (MCO), y cuando  $K_1 = K_2 = 1$  (MC2E). Este estimador puede ser utilizado en cualquier ecuación del MES que esté identificada, sin importar si es exactamente identificada o sobreidentificada. Según la condición de orden, para resolver una ecuación por MC2E debe cumplirse  $m_i - 1 + k_i \leq k$ .

### 2.3.4. Mínimos Cuadrados en Tres Etapas (MC3E)

MC3E estima conjuntamente las ecuaciones que forman el modelo, por lo que requiere que todas ellas estén identificadas. En la estimación utiliza la matriz de varianzas-covarianzas de los términos de error estructurales. Al incorporar toda la información del sistema, los estimadores obtenidos son asintóticamente más eficientes [24], pero hay situaciones en las que esta información no mejora la eficacia de los estimadores. Si alguna ecuación del modelo está mal especificada provoca estimaciones no consistentes en los estimadores del resto de ecuaciones. Cuando no hay correlación contemporánea entre los errores de las distintas ecuaciones, este método no gana en eficiencia. Además, si

---

todas las ecuaciones están exactamente identificadas, las estimaciones de MC2E y MC3E coinciden.

### 2.3.5. Método Bayesiano de Momentos (BMOM)

Este método (del inglés Bayesian Method of Moments), propuesto por Zellner en 1998 [7], aplica el principio de máxima entropía y genera una estimación óptima evaluada por estimadores de clase doble K mostrados en la tabla 2.1. Cuando no hay suficiente información disponible para obtener la función de probabilidad, este método permite un análisis de datos sin especificar una función de probabilidad.

Considerando, por ejemplo, la primera ecuación estructural,  $y_1 = Y_1\beta_1 + X_1\gamma_1 + u_1$ , siendo  $Y_1 \in \mathbb{R}^{n \times m}$  y  $X_1 \in \mathbb{R}^{n \times k}$  la matriz de variables endógenas y exógenas y  $u_1 \in \mathbb{R}^{n \times 1}$ , el vector de ruido blanco, los parámetros del modelo denotados por  $\delta_1 = (\beta_1 \gamma_1)^T \in \mathbb{R}^{1 \times (m+k)}$  estimados por BMOM vienen dados por:

$$\hat{\delta}_1(K_1, K_2) = \begin{bmatrix} Y_1^T Y_1 - K_1 \hat{V}_1^T \hat{V}_1 & Y_1^T Z_1 \\ Z_1^T Y_1 & Z_1^T Z_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_1 - K_2 \hat{V}_1)^T y_1 \\ Z_1^T y_1 \end{bmatrix} \quad (2.7)$$

con  $Z_1 = (X_1 \Pi_1)^T$ , donde  $\Pi_1$  es la matriz de coeficientes de la forma reducida  $Y_1$ , siendo  $Y_1 = X_1 \Pi_1 + V_1$ . Estos coeficientes pueden ser calculados por mínimos cuadrados obteniendo  $\hat{\Pi}_1 = (X^T X)^{-1} X^T Y_1$  y  $\hat{V}_1 = Y_1 - X \hat{\Pi}_1$ .

La tabla 2.1 muestra los parámetros  $K_1$  y  $K_2$  para las funciones de pérdida utilizando el enfoque BMOM, siendo  $n$  el tamaño de la muestra,  $k$  el número de variables exógenas y  $m_i$  el número de variables endógenas de la ecuación  $i$ -ésima.

Tabla 2.1: Valores de los parámetros  $K_1$  y  $K_2$  propuestos para minimizar la función de pérdida.

Función de pérdida	Enfoque BMOM	Enfoque Bayesiano Tradicional
1. Bondad de ajuste $L_g = (y_1 - Z_1 \hat{\delta}_1)^T (y_1 - Z_1 \hat{\delta}_1)$	$K_1 = 1 - k/(n - k)$ $K_2 = 1$	$K_1 = 1 - k/(n - k - m_i - 2)$ $K_2 = 1$
2. Precisión de estimación $L_p = (\delta_1 - \hat{\delta}_1)^T Z_1^T Z_1 (\delta_1 - \hat{\delta}_1)$	$K_1 = K_2 = 1 - k/(n - k)$	$K_1 = K_2 = 1 - k/(n - k - m_i - 2)$
3. Función de pérdida balanceada $L_b = wL_g + (1 - w)L_p$	$K_1 = 1 - k/(n - k)$ $K_2 = 1 - (1 - w)k/(n - k)$	$K_1 = 1 - k/(n - k - m_i - 2)$ $K_2 = 1 - (1 - w)k/(n - k - m_i - 2)$

### 2.3.6. Enfoque Bayesiano en dos etapas ( $Bayes_{2E}$ )

En inferencia Bayesiana, una solución pragmática para elegir la distribución a priori es seleccionar un miembro de la familia de una distribución de tal manera que la distribución a posteriori pertenezca a la misma familia que la previa. Si el previo es conjugado, la distribución posterior después de la primera observación pertenece, por definición, al mismo tipo y se usa como la nueva distribución previa en la siguiente observación. Al incorporar esta segunda observación, la nueva distribución posterior, también pertenece a la clase de conjugación. Así, este proceso secuencial solo actualiza el valor de los parámetros de la distribución [11]. Concretamente, en este estudio este enfoque será aplicado en dos pasos, emulando así la técnica de mínimos cuadrados en dos etapas.

### 2.3.7. Inferencia MCMC

En inferencia Bayesiana es esencial seleccionar la distribución previa. Sin embargo, hay situaciones en las que esta selección es algo difícil debido a la ausencia de información previa del modelo. La metodología de MCMC proporciona un amplio alcance para el modelado estadístico y se usa ampliamente para calcular las distribuciones posteriores en modelos econométricos complejos. En particular, estos métodos necesitan integrar la

distribución posterior de los parámetros del modelo y por esa razón MCMC toma muestras de esta distribución posterior. En general, para simular valores de una distribución a posteriori, estos métodos los obtienen de manera que dependen únicamente de su predecesor, asegurando la convergencia independientemente de los valores iniciales. Por ello, es necesario simular la cadena para un número elevado de iteraciones y desechar los primeros valores por considerar que no están en el estado estacionario. Hay muchas formas de construir estas muestras, como el muestreo de Gibbs o casos especiales del marco general de Metrópolis y Hastings [11].

## 2.4. Conclusiones

Tras una revisión de la literatura, se ha presentado el MES y algunas de sus aplicaciones en áreas que van desde la economía hasta la medicina. Se han descrito algunos de los métodos de estimación más utilizados, tanto en el enfoque clásico o frecuentista como en inferencia Bayesiana.

Dentro de la inferencia frecuentista parece intuitivo que son preferibles métodos sistemáticos, es decir, que utilizan toda la información del sistema, como por ejemplo MC3E, frente a métodos que estiman ecuación a ecuación, sin información de otras ecuaciones, como los métodos MC2E y MCO. Pero, a pesar de que los sistemáticos son asintóticamente mejores, tienen dos grandes problemas. El primero es que cualquier especificación errónea de la estructura del modelo se extenderá por todo el sistema, lo que afectará al resto de ecuaciones aunque éstas estén correctamente especificadas, mientras que los métodos con información limitada limitarán el problema a la ecuación concreta en la cual aparece ese error de especificación. El segundo problema es que la variación para muestras finitas de la matriz estimada de varianzas-covarianzas se transmite por todo el sistema, de modo que la varianza de MC3E puede ser tan grande o más que la de MC2E.

Respecto a los métodos de inferencia Bayesiana, la comparación no es clara, siendo evidente la gran diferencia que presentan unos métodos de otros, además de contar o no con información previa. Al existir en MCMC un gran número de algoritmos diferentes así como funciones de aceptación, la comparación entre ellos puede resultar muy compleja

---

de llevar a cabo.

Todo ello, sin información del problema, conlleva a la conclusión de que no hay un método con mejores bondades que otros, al menos inicialmente.



## CAPÍTULO 3

---

### Criterio de Información y Entropía

---

En este capítulo se hace una revisión de la Entropía como medida de incertidumbre, su origen, propiedades y aplicaciones. Además, se desarrollan los conceptos y resultados más importantes para el estudio y análisis de los criterios de selección en modelos de regresión y su adaptación a los MES. El problema de selección de modelos, básicamente se reduce a seleccionar de entre un conjunto de modelos candidatos aquél que tenga menor error en el ajuste o más próximo esté del modelo verdadero, medida a partir de una distancia o criterio de información.

La primera parte del capítulo contiene el estudio de los conceptos de la teoría de la información, concretamente los conceptos de entropía, entropía relativa (Divergencia de Kullback-Leibler) e información mutua y algunas de sus propiedades, así como una revisión a la entropía dada por Shannon. La segunda parte recoge los criterios de información usados en econometría para la selección del mejor modelo de entre un conjunto de candidatos y que han sido adaptados a los MES.

## 3.1. Conceptos básicos de Teoría de la información.

### Entropía.

La teoría de la información responde dos preguntas fundamentales en la teoría de la comunicación: cuál es la compresión de datos final (la entropía  $H$ ) y cuál es la tasa de transmisión de comunicación final (la capacidad del canal  $C$ ). Por esta razón, algunos consideran que la teoría de la información es un subconjunto de la teoría de la comunicación [27].

El concepto de entropía se introdujo inicialmente en la termodinámica, donde se utilizó para proporcionar la base de la segunda ley de la termodinámica, en la que se define el grado de desorden de un sistema. Posteriormente, la estadística mecánica proporcionó una conexión entre las propiedades macroscópicas de la entropía y los estados del sistema. Desde un punto de vista matemático, son funciones no negativas de distribuciones de probabilidad con múltiples aplicaciones, como el uso de la teoría de la información para medir la estabilidad de un sistema [28]. Las aplicaciones de la entropía se pueden encontrar en los campos tan diversos como el de finanzas, [29], ingeniería ambiental y del agua [30], sistemas urbanos [31] y aplicaciones para encuestas de satisfacción del cliente [32].

A continuación se describe brevemente la relación de la teoría de la información con otros campos [28]:

**Ingeniería eléctrica (Teoría de la comunicación)** A principios de los años 40 se pensaba que aumentar la velocidad de transmisión de información a través de un canal de comunicación aumentaba la probabilidad de error. Shannon sorprendió a la comunidad de la teoría de la comunicación al demostrar que esto no era cierto siempre que la velocidad de comunicación estuviera por debajo de la capacidad del canal, pudiendo calcular esta capacidad a partir de las características de ruido del canal. Shannon argumentó además que los procesos aleatorios como la música y el habla tienen una complejidad irreducible por debajo de la cual la señal no se puede comprimir. A esto lo llamó entropía, en deferencia al uso paralelo de esta palabra en

---

termodinámica, y argumentó que si la entropía de la fuente es menor que la capacidad del canal, entonces se puede lograr una comunicación asintóticamente libre de errores.

El trabajo moderno sobre los aspectos de comunicación de la teoría de la información se ha concentrado en la teoría de la información de red: la teoría de las velocidades simultáneas de comunicación de muchos emisores a muchos receptores en una red de comunicación.

**Informática (Complejidad de Kolmogorov)** Kolmogorov, Chaitin y Solomonoff plantearon la idea de que la complejidad de una cadena puede definirse por la longitud del programa binario más corto para calcular la cadena, por lo que la complejidad es la longitud mínima de descripción. Esta definición de complejidad es universal, independiente de la computadora y es de fundamental importancia al sentar las bases para la teoría de la complejidad. La complejidad  $K$  de Kolmogorov es aproximadamente igual a la entropía  $H$  de Shannon si la secuencia se extrae al azar de una distribución que tiene entropía  $H$ .

**Física (Termodinámica)** La mecánica estadística es el lugar de nacimiento de la entropía y la segunda ley de la termodinámica. La entropía siempre aumenta y entre otras cosas, la segunda ley permite descartar cualquier reclamo de máquinas de movimiento perpetuo.

**Matemáticas (Teoría de la probabilidad y estadística)** Las cantidades fundamentales de la teoría de la información, entropía, entropía relativa e información mutua, se definen como funciones de las distribuciones de probabilidad. Además, caracterizan el comportamiento de secuencias largas de variables aleatorias y consiguen estimar probabilidades de eventos raros y encontrar la mejor estimación del error en las pruebas de hipótesis.

**Filosofía de la ciencia (Navaja de Occam)** William de Occam dijo "Las causas no se multiplicarán más allá de la necesidad", o parafraseando, "La explicación más simple es la mejor". Solomonoff, y luego Chaitin, argumentan que uno obtiene un procedimiento de predicción universalmente bueno si toma una combinación

---

ponderada de todos los programas que explican los datos y observa lo que imprimen a continuación.

**Economía (Inversión).** La inversión repetida en un mercado de valores estacionario resulta en un crecimiento exponencial de la riqueza. La tasa de crecimiento de la riqueza es el doble de la tasa de entropía del mercado de valores. Existe un gran paralelismo entre la teoría de la inversión óptima en el mercado de valores y la teoría de la información.

**Computación vs. Comunicación** A medida que se construyen computadoras más grandes con componentes más pequeños, se encuentra un límite de cómputo y un límite de comunicación. La computación es comunicación limitada y la comunicación es computación limitada. Estos conceptos se entrelazan y por lo tanto, todos los desarrollos en la teoría de la comunicación a través de la teoría de la información deberían tener un impacto directo en la teoría de la computación.

## 3.2. Tipos de Entropía

Sea  $X$  una variable aleatoria discreta de  $\mathcal{X}$  y función de masa de probabilidad  $p(x) = \Pr\{X = x\}, x \in \mathcal{X}$ . Se denota la función de masa de probabilidad por  $p(x)$  en lugar de  $p_X(x)$  por conveniencia. Por lo tanto,  $p(x)$  y  $p(y)$  se refieren a dos variables aleatorias diferentes y de hecho, son funciones de masa de probabilidad diferentes,  $p_X(x)$  y  $p_Y(y)$  respectivamente.

### Definición 1

La entropía  $H(X)$  de una variable aleatoria discreta  $X$  se define por

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x) \quad (3.1)$$

También se escribe  $H(p)$  para definir la cantidad anterior. El registro está en base 2 y la entropía se expresa en bits. Por ejemplo, la entropía de un lanzamiento de una moneda es de 1 bit. Se usa la convención de que  $0 \log 0 = 0$ , que se justifica fácilmente

---

por la continuidad ya que  $x \log x \rightarrow 0$  cuando  $x \rightarrow 0$ . Por lo tanto, agregar términos de probabilidad 0 no cambia la entropía.

Si la base del logaritmo es  $b$ , se denota la entropía como  $H_b(X)$  y si la base del logaritmo es  $e$ , entonces la entropía se mide en nats. Se debe tener en cuenta que la entropía es una función de la distribución de  $X$ , por lo que no depende de los valores reales tomados por la variable aleatoria  $X$ , sino solo de las probabilidades. Se denota esperanza por  $E$  y por lo tanto, si  $X \sim p(x)$ , entonces el valor esperado de la variable aleatoria  $g(x)$  es,

$$E_p[g(X)] = \sum_{x \in \mathcal{X}} g(x)p(x) \quad (3.2)$$

o más simplemente como  $E[g(X)]$ . Tiene un interés peculiar la esperanza autorreferencial de  $g(x)$  bajo  $p(x)$  cuando  $g(x) = \log \frac{1}{p(x)}$ .

La entropía de  $X$  también puede interpretarse como el valor esperado de  $\frac{1}{p(x)}$ , donde  $X$  se dibuja de acuerdo con la función de probabilidad de masa  $p(x)$ . Así

$$H(X) = E_p \left[ \log \frac{1}{p(x)} \right] \quad (3.3)$$

A continuación, algunas consecuencias inmediatas de la definición:

**Lema 1**

$$H(X) \geq 0$$

*Demostración:*

$$0 \leq p(x) \leq 1 \text{ implica } \log(1/p(x)) \geq 0$$

**Lema 2**

$$H_b(X) = (\log_b a) H_a(X)$$

*Demostración:*

$$\log_b p = \log_b a \log_a p$$

### 3.2.1. Entropía conjunta y entropía condicional

La definición de entropía de una sola variable aleatoria del punto anterior se extiende ahora a la entropía de dos variables aleatorias.

---

**Definición 2**

La entropía conjunta  $H(X,Y)$  de dos variables aleatorias discretas  $(X,Y)$  con una distribución conjunta  $p_X(x)$  se define como

$$H(X,Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(x,y) \quad (3.4)$$

lo cual se puede expresar como

$$H(X,Y) = -E [\log p(x,y)] \quad (3.5)$$

A continuación se define la entropía condicional de una variable aleatoria dada otra:

**Definición 3**

Si  $(X,Y) \sim p(x,y)$ , entonces la entropía condicional  $H(Y|X)$  es definida como

$$\begin{aligned} H(Y|X) &= \sum_{x \in \mathcal{X}} p(x) H(Y|X=x) \\ &= - \sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x) \\ &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(y|x) \\ &= -E_{p(x,y)} [\log p(y|x)] \end{aligned} \quad (3.6)$$

es decir, es el valor esperado de las entropías de las distribuciones condicionales, promediadas por la variable condicionante.

**Teorema 1**

*Regla de la cadena*

$$H(X,Y) = H(X) + H(Y|X) \quad (3.7)$$

*Demostración*

$$\begin{aligned} H(X,Y) &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(x,y) \\ &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(x)p(y|x) \\ &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(x) - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(y|x) \\ &= - \sum_{x \in \mathcal{X}} p(x) \log p(x) - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(y|x) \\ &= H(X) + H(Y|X) \end{aligned} \quad (3.8)$$

**Corolario 1**

$$H(X, Y|Z) = H(X|Z) + H(Y|X, Z) \quad (3.9)$$

**3.2.2. Entropía relativa e información mutua**

La entropía de una variable aleatoria es una medida de la incertidumbre de la variable aleatoria. Es una medida de la cantidad de información requerida en promedio para describir la variable aleatoria. La entropía relativa es una medida de la distancia entre dos distribuciones.

**Definición 4**

La entropía relativa o la distancia de Kullback Leibler entre dos funciones de masa de probabilidad  $p(x)$  y  $q(x)$  se define como

$$\begin{aligned} D(p||q) &= \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \\ &= E_p \left[ \log \frac{p(x)}{q(x)} \right] \end{aligned} \quad (3.10)$$

La entropía relativa es no negativa y solo es 0 si  $p=q$ .

La entropía relativa no es una distancia real entre distribuciones ya que no es simétrica y no satisface la desigualdad del triángulo. Sin embargo, a menudo se piensa en la entropía relativa como una distancia entre distribuciones.

Ahora se presenta la información mutua, que es una medida de la cantidad de información que contiene una variable aleatoria sobre otra variable aleatoria. Es la reducción en la incertidumbre de una variable aleatoria debido al conocimiento de la otra.

**Definición 5**

Sean dos variables aleatorias  $X$  e  $Y$  con una función de masa de probabilidad conjunta  $p(x, y)$  y funciones de masa de probabilidad marginal  $p(x)$  y  $p(y)$ . La información mutua  $I(X; Y)$  es la entropía relativa entre la distribución conjunta y la distribución del producto

---

$p(x)p(y)$ , es decir

$$\begin{aligned}
 I(X;Y) &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\
 &= D(p(x,y) \| p(x)p(y)) \\
 &= E_{p(x,y)} \left[ \log \frac{p(x,y)}{p(x)p(y)} \right]
 \end{aligned} \tag{3.11}$$

### 3.2.3. Relación entre la entropía y la información mutua

#### Teorema 2

*Información mutua y entropía*

$$\begin{aligned}
 I(X;Y) &= H(X) - H(X|Y) \\
 I(X;Y) &= H(Y) - H(Y|X) \\
 I(X;Y) &= H(X) + H(Y) - H(X,Y) \\
 I(X;Y) &= I(Y;X) \\
 I(X;X) &= H(X)
 \end{aligned} \tag{3.12}$$

### 3.2.4. Regla de la cadena para entropía, entropía relativa e información mutua

#### Teorema 3

*Regla de la cadena para la entropía*

Sean  $X_1, X_2, \dots, X_n$  que pueden ser representados de acuerdo a  $p(x_1, x_2, \dots, x_n)$ . Entonces,

$$H(X_1, X_2, \dots, X_n) = \sum_{i=1}^n H(X_i | X_{i-1}, \dots, X_1) \tag{3.13}$$

#### Definición 6

*La información mutua condicional de las variables aleatorias X e Y dada Z se define por*

$$\begin{aligned}
 I(X;Y|Z) &= H(X|Z) - H(X|Y,Z) \\
 &= E_{p(x,y,z)} \left[ \log \frac{p(x,y|z)}{p(x|z)p(y|z)} \right]
 \end{aligned} \tag{3.14}$$

**Teorema 4**

*Regla de la cadena para Información*

$$I(X_1, X_2, \dots, X_n; Y) = \sum_{i=1}^n I(X_i; Y | X_{i-1}, X_{i-2}, \dots, X_1) \quad (3.15)$$

**Definición 7**

La entropía relativa condicional  $D(p(y|x)||q(y|x))$  es el promedio de las entropías relativas entre las funciones de probabilidad condicional  $p(x|y)$  y  $q(y|x)$  promediadas sobre la función de probabilidad  $p(x)$ , es decir,

$$\begin{aligned} D(p(y|x)||q(y|x)) &= \sum_x p(x) \sum_y p(y|x) \log \frac{p(y|x)}{q(y|x)} \\ &= E_{p(x,y)} \left[ \log \frac{p(y|x)}{q(y|x)} \right] \end{aligned} \quad (3.16)$$

**Teorema 5**

*Regla de la cadena para la entropía relativa*

$$D(p(x,y)||q(x,y)) = D(p(x)||q(x)) + D(p(y|x)||q(y|x)) \quad (3.17)$$

**3.2.5. La desigualdad de Jensen y sus propiedades**

Algunas de las más notables propiedades de la entropía son consecuencia directa de la convexidad de la función logaritmo y la propia entropía es también una función cóncava, por lo que, a continuación se muestran las propiedades de estas funciones.

**Definición 8**

Se dice que una función  $f(x)$  es convexa en un intervalo  $(a, b)$  si para cada  $x_1, x_2 \in (a, b)$  y  $0 \leq \lambda \leq 1$

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \quad (3.18)$$

Se dice que una función  $f$  es estrictamente convexa si la igualdad se cumple solo si  $\lambda = 0$  ó  $\lambda = 1$

**Definición 9**

Una función  $f$  es cóncava si  $-f$  es convexa. Una función es convexa si siempre se encuentra debajo de cualquier cuerda. Una función es cóncava si siempre se encuentra por encima de cualquier cuerda.

**Teorema 6**

Si la función  $f$  tiene una segunda derivada que no es negativa (positiva) en todas partes, entonces la función es convexa (estrictamente convexa).

**Teorema 7**

Desigualdad de Jensen.

Si  $f$  es una función convexa y  $X$  es una variable aleatoria, entonces

$$E[f(X)] \geq f(E[X]) \quad (3.19)$$

Además, si  $f$  es estrictamente convexo, entonces la igualdad en 3.19 implica que  $X = E[X]$  con probabilidad 1, es decir,  $X$  es una constante.

**Teorema 8**

Desigualdad de información.

Sea  $p(x)$ ,  $q(x)$ ,  $x \in \mathcal{X}$ , dos funciones de probabilidad. Entonces

$$D(p||q) \geq 0 \quad (3.20)$$

con la igualdad si y solo si

$$p(x) = q(x) \text{ para todo } x$$

**Corolario 2**

No negatividad de la información mutua.

Para dos variables aleatorias,  $X$ ,  $Y$

$$I(X;Y) \geq 0 \quad (3.21)$$

con la igualdad si y solo si  $X$  e  $Y$  son independientes

**Corolario 3**

$$D(p(y|x)||q(y|x)) \geq 0 \quad (3.22)$$

con la igualdad si y solo si  $p(y|x) = q(y|x)$  para todo  $x$  e  $y$  con  $p(x) > 0$

**Teorema 9**

$H(X) \leq \log|\mathcal{X}|$ , donde  $|\mathcal{X}|$  denota el número de elementos en el rango de  $X$ , con igualdad si y solo si  $X$  tiene una distribución uniforme sobre  $|\mathcal{X}|$

---

**Teorema 10**

*El acondicionamiento reduce la entropía.*

$$H(X|Y) \leq H(X) \quad (3.23)$$

*con igualdad si y solo si X e Y son independientes.*

**3.2.6. La desigualdad de Log Sum y sus aplicaciones**

Ahora se demuestra una consecuencia simple de la concavidad del logaritmo, que se utilizará para probar algunos resultados de concavidad para la entropía.

**Teorema 11**

*Desigualdad de suma logarítmica.*

*Para números no negativos,  $a_1, a_2, \dots, a_n$  y  $b_1, b_2, \dots, b_n$*

$$\sum_{i=1}^n a_i \log \frac{a_i}{b_i} \geq \left( \sum_{i=1}^n a_i \right) \log \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i} \quad (3.24)$$

*con igualdad si y solo si  $\frac{a_i}{b_i} = \text{constante}$ .*

**3.2.7. Desigualdad de procesamiento de datos**

La desigualdad en el procesamiento de datos se puede usar para mostrar que ninguna manipulación inteligente de los datos puede mejorar las inferencias que se pueden hacer a partir de los datos.

**Definición 10**

*Se dice que las variables aleatorias X, Y, Z forman una cadena de Markov en ese orden (denotado por  $X \rightarrow Y \rightarrow Z$ ) si la distribución condicional de Z depende solo de Y y es condicionalmente independiente de X. Específicamente, X, Y y Z forman una cadena de Markov  $X \rightarrow Y \rightarrow Z$  si la función de probabilidad conjunta se puede escribir como*

$$p(x, y, z) = p(x)p(y|x)p(z|y) \quad (3.25)$$

A continuación se muestran algunas consecuencias:

---

- $X \rightarrow Y \rightarrow Z$  si y solo si  $X$  y  $Z$  son condicionalmente independientes dado  $Y$ . La propiedad Markoviana implica independencia condicional porque

$$p(x, z|y) = \frac{p(x, y, z)}{p(y)} = \frac{p(x, y)p(z|y)}{p(y)} = p(x|y)p(z|y) \quad (3.26)$$

Esta es la caracterización de las cadenas de Markov que se pueden extender para definir los campos de Markov, que son procesos aleatorios  $n$ -dimensionales en los que el interior y el exterior son independientes dados los valores en la frontera.

- $X \rightarrow Y \rightarrow Z$  implica que  $Z \rightarrow Y \rightarrow X$ . Esta propiedad se suele escribir como  $X \leftrightarrow Y \leftrightarrow Z$
- Si  $Z = f(Y)$ , entonces  $X \rightarrow Y \rightarrow Z$

#### **Teorema 12**

*Desigualdad de procesamiento de datos. Si  $X \rightarrow Y \rightarrow Z$ , entonces  $I(X; Y) \geq I(X; Z)$*

#### **Corolario 4**

*En particular, si  $Z = g(Y)$  tenemos que  $I(X; Y) \geq I(X; g(Y))$*

#### **Corolario 5**

*Si  $X \rightarrow Y \rightarrow Z$  entonces  $I(X; Y|Z) \leq I(X; Y)$*

### **3.3. Criterios de información en la obtención de modelos estadísticos**

Cuando se obtienen estimaciones a partir de la aplicación de diferentes métodos es necesario disponer de un criterio de parámetro o criterio de información para la selección de aquél que menor error produzca.

Estos criterios suelen utilizarse como simples herramientas para elegir entre una serie de modelos estimados que difieren en bondad de ajuste y complejidad, ofreciendo una estimación relativa de la información perdida cuando se utiliza un modelo determinado para representar el proceso que genera los datos. Intuitivamente, un modelo más complejo describirá mejor las relaciones que se están analizando y al mismo tiempo, un modelo

complejo pierde la capacidad de explicar esas relaciones. El objetivo es, por tanto, encontrar un modelo que posea el equilibrio entre la complejidad y la bondad del ajuste, pero, en caso de disponer de un conjunto de modelos estimados, estos criterios no avisan si ninguno de ellos es un buen candidato.

Existe una gran variedad de criterios aplicados a los modelos de regresión con una sola ecuación, siendo el Criterio de Información de Akaike (AIC) [33, 34], su versión corregida (AICc) [35], el Criterio de Información Bayesiano o de Schwarz (BIC/SIC)[36], el criterio de Hannan y Quinn (HQ) [37] y el Criterio de selección del modelo basado en la Divergencia simétrica de Kullback - Leibler [38], los más utilizados.

A continuación se hace una breve descripción de algunos de estos criterios de información.

### 3.3.1. El criterio de información de Akaike

El precursor del uso de los criterios de información para la selección de modelos fue Hirotosugu Akaike en los años 70 [33], cuyo método, para el caso general se determina a partir de la siguiente expresión:

$$AIC = -2\ln(L) + 2k \quad (3.27)$$

donde  $L$  es la función de verosimilitud para el modelo estimado y  $k$  el número de variables o parámetros a estimar.

Como puede observarse, AIC está formado por dos componentes: la primera componente,  $-2\ln L$  es el valor de la función de verosimilitud, que es la probabilidad de obtener los datos dado el modelo candidato. Se puede observar que, sin tener en cuenta la segunda componente de la expresión, el modelo con el AIC mínimo es el que tiene el valor más alto para la función de verosimilitud. Sin embargo, a este primer componente se agrega un ajuste basado en el número de parámetros estimados. Cuanto mayor es el número de parámetros del modelo, mayor es la cantidad que se suma al primer componente, aumentando así el valor del AIC y, de este modo, se penaliza al modelo por ser este más complejo.

---

En 1989, C. M. Hurvich y C. Tsay [35] proponen una expresión corregida del AIC para el caso de modelos en los que el tamaño muestral no es elevado y/o el número de parámetros es relativamente grande, resultando un criterio que proporciona modelos más adecuados en estas circunstancias. La corrección consiste en la modificación del segundo término, el de penalización, quedando la expresión:

$$AIC_c = AIC + \frac{2k^2 + 2k}{n - k - 1} \quad (3.28)$$

donde  $k$  es el número de parámetros a estimar y  $n$  el tamaño de la muestra.

Ambos,  $AIC$  y  $AIC_c$  son similares cuando el cociente  $n/k$  es suficientemente grande.

### 3.3.2. El criterio de información Bayesiano

El criterio de información bayesiano [36], propuesto por Schwarz en 1978 y por lo tanto también conocido como el criterio de información de Schwarz y el criterio bayesiano de Schwarz, es otro criterio de selección de modelo basado en la teoría de la información pero establecido dentro de un contexto bayesiano. La diferencia entre BIC y AIC es la mayor penalización impuesta por el número de parámetros por el primero que por el segundo. Burnham y Anderson proporcionan argumentos teóricos a favor del AIC, en particular  $AIC_c$  sobre BIC. Además, en el caso del análisis de regresión multivariante, Yang explica por qué AIC es mejor que BIC en la selección de modelos [19]. El BIC se calcula según la siguiente expresión:

$$BIC = -2\ln(L) + 2k \ln(n) \quad (3.29)$$

donde  $L$  es la función de verosimilitud para el modelo estimado,  $k$  el número de parámetros a estimar y  $n$  el tamaño de la muestra. El mejor modelo es el que proporciona menor BIC.

### 3.3.3. El Criterio de información de Hannan y Quinn

En 1979, E. J. Hannan y B. G. Quinn [37] proponen una variación del BIC en el que el tamaño muestral influye más suavemente que en éste, pero es constante como ocurría en

---

el caso del AIC. Su expresión viene dada por:

$$HQIC = -2\ln(L) + 2k \ln(\ln(n)) \quad (3.30)$$

donde  $L$  es la función de verosimilitud para el modelo estimado,  $k$  el número de parámetros a estimar y  $n$  el tamaño de la muestra.

### 3.4. Criterio de Información de Akaike y Bayesiano aplicado al MES

Algunos de los criterios revisados en el punto anterior, tales como AIC y BIC, son muy utilizados en modelos de regresión multivariante y se han adaptado a los MES [39] como criterio de selección de modelo.

Las expresiones de AIC y BIC para su aplicación a los MES se presentan a continuación:

$$AIC = n \ln |\hat{\Sigma}_e| + 2 \sum_{i=1}^m (m_i + k_i - 1) + m(m+1) \quad (3.31)$$

$$BIC = n \ln |\hat{\Sigma}_e| + \ln n \left( \sum_{i=1}^m (m_i + k_i - 1) + \frac{m}{2}(m+1) \right) \quad (3.32)$$

donde  $|\hat{\Sigma}_e|$  es el determinante de la matriz de varianzas-covarianzas de los errores  $e_i$ ,  $i = 1, \dots, m$ , donde  $e_i$  es la diferencia entre la variable endógena  $Y_i$  y su estimación  $\hat{Y}_i$  en la ecuación  $i$ -ésima,  $m$  es el número de ecuaciones del modelo,  $n$  es el tamaño de la muestra y  $m_i$  y  $k_i$  el número de variables endógenas y exógenas respectivamente en la ecuación  $i$ -ésima.

Ambos criterios, que pueden descomponerse como la suma de un indicador de la *bondad de ajuste* y otro que indica la *complejidad* del modelo, solo difieren en cómo calcula ese factor de complejidad, que en ambos casos depende de la dimensión del problema, tanto del tamaño de la muestra como del número de variables endógenas. Y dado que en este estudio se comparan métodos y no modelos, el factor de complejidad en ambas medidas es constante en esa comparación, por lo que ambos criterios son equivalentes.

Uno de los objetivos de esta tesis es encontrar un criterio de información para la selección del método de estimación de un MES con menor error de estimación y no la selección de un mejor modelo de entre varios candidatos. En definitiva, el problema a resolver consiste en encontrar un criterio para la selección del mejor método que estime la relación entre las variables endógenas y exógenas, permitiendo que esta relación sea simultánea. No se trata de comparar estos criterios entre sí, si no de comparar dos o más modelos idénticos en dimensión tanto de variables endógenas, exógenas y el tamaño muestral, entre los que solo existe diferencia en la bondad del ajuste y por tanto en los errores  $e_i$ , no en su complejidad.

### **3.5. Conclusiones**

El estudio de los criterios de información llevado a cabo en este capítulo conduce al desarrollo de nuevas medidas basadas en los errores de estimación que permiten valorar la bondad del modelo estimado. Cómo utilizar estos errores en la obtención de un criterio para la selección de un método de estimación es uno de los objetivos de esta tesis cuyo desarrollo y resultados se describen en la sección 5.2, en la que el concepto de entropía es usada como base para su desarrollo.

---

# CAPÍTULO 4

---

## Modelo de Ecuaciones Simultáneas No Lineal y Red Neuronal Artificial

---

En este capítulo se hace una introducción a los métodos de estimación de modelos de ecuaciones simultáneas no lineales y a las redes neuronales artificiales.

### 4.1. Introducción

En los modelos uniecuacionales, la extensión del modelo lineal a ecuaciones no lineales es relativamente sencillo y, generalmente, solo requiere pequeños cambios en el análisis consistentes en la simple *linealización* del problema. En los MES, esta simplificación no se podrá aplicar excepto en unas pocas situaciones concretas: si la no linealidad aparece causada por las variables exógenas y de los parámetros, entonces, mediante una redefinición apropiada se puede obtener un modelo lineal. Sin embargo, si son las variables endógenas las que presentan no linealidad, el sistema de ecuaciones no garantiza la existencia de la forma reducida o que la función no lineal sea en variables exógenas y perturbaciones [24]. La aparición de las redes neuronales artificiales (RNA)

permiten dar un paso en la reformulación de un MES no lineal en la que no es necesario modelizar previamente las relaciones entre las variables. Las redes neuronales constituyen una técnica de análisis de datos que desde hace más de tres décadas se ha extendido su utilización en muchos ámbitos y han adquirido una gran importancia al dar respuestas a problemas que son difíciles de resolver usando la programación basada en reglas. Estas redes toman su nombre por el parecido de su arquitectura a una red neuronal humana, a pesar de las diferencias con el funcionamiento del cerebro humano [40].

#### **4.1.1. Estimación del modelo de ecuaciones simultáneas no lineal**

De forma similar a la estimación de un modelo lineal, en esta sección se hace un repaso a las técnicas usadas para estimar relaciones que no pueden ser modeladas linealmente, y más concretamente, se hace para el caso de no linealidad en parámetros.

Amemiya en 1974 [41] muestra que las propiedades de optimalidad de MC2E se extienden al MC2ENL (mínimos cuadrados en dos etapas no lineal) siempre que el modelo sea lineal en variables y no lineal en parámetros. El mismo autor, en 1977 [42] compara el estimador máximo verosímil (EMV) con el estimador mínimos cuadrados en tres etapas no lineal (MC3ENL) en el que prueba que la consistencia y la normalidad asintótica del EMV en los MES no lineales depende del supuesto de normalidad del término de error, a diferencia del caso lineal. También concluye que el EMV es asintóticamente más eficiente que MC3ENL si las especificaciones son correctas, pero es menos robusto, siendo el MC3ENL consistente incluso cuando el supuesto de normalidad no se cumple.

Posteriormente, Besley en 1979 [43] compara la competitividad computacional del método de máxima verosimilitud con información completa (MVIC) con su competidor más popular, MC3E, en la estimación de una variedad de modelos lineales y no lineales (en parámetros y variables). Describe el estimador no lineal de máxima verosimilitud de información completa (MVICNL) y define una aproximación computacionalmente eficiente, el TRUNFIML. La aplicación de MC3ENL se logra mediante el método de Jorgenson-Laffont y las comparaciones se realizan sobre la base de la cantidad de

---

iteraciones a la convergencia, la cantidad de evaluaciones de funciones y el tiempo total requerido de la CPU de la computadora.

Por otro lado, Blundell [44], en un estudio más reciente en 2013 ha revisado dos clases alternativas de enfoques de sistemas de estimación en modelos de ecuaciones simultáneas no lineales no aditivas. El primer enfoque utiliza una suposición de independencia condicional y el segundo adopta restricciones de exclusión sobre variables instrumentales. En particular, centrándose en un caso especial del primer enfoque, la función de control, se ha observado que en modelos no lineales la estimación utilizando el estimador de función de control puede ser gravemente inconsistente a menos que se impongan restricciones fuertes al modelo simultáneo.

En lo que respecta a estudios experimentales, Goldfeld en 1968 [45] analiza las propiedades de muestras pequeñas de las estimaciones de parámetros de dos sistemas no lineales específicos. Como tal, puede considerarse como una extensión natural de estudios previos, el cuál fue motivado por las siguientes preguntas: (1) ¿En qué medida los experimentos previos con modelos lineales tienen relevancia para los modelos no lineales, es decir, en qué medida las conclusiones sustantivas derivadas de casos lineales son válidas para los modelos no lineales? (2) ¿Existe una variedad de métodos de estimación más o menos igualmente razonables para modelos no lineales? (3) ¿Se puede idear un algoritmo que tenga propiedades de convergencia aceptables, es decir, un algoritmo que converja rápidamente al máximo verdadero, para el cálculo de estimaciones de máxima probabilidad de información completa?.

Por otro lado, Winegarden en 1984 [46] explica la nueva economía doméstica en función de la fecundidad, trabajo y estado civil. Presenta un modelo de tres ecuaciones simultáneas y tres variables endógenas. Además, hace referencia temporal, en intervalos de diez años y territorial, utilizando variables por país. Debido a la no linealidad de las variables clave y la simultaneidad del sistema, se estimó utilizando el método MC3ENL.

El método de mínimos cuadrados no lineales (MCONL), se puede entender como una extensión del procedimiento del método de estimación MCO en modelos lineales, de hecho las hipótesis de partida del método mínimo-cuadrático no exige en ningún momento la linealidad del modelo, si bien la resolución analítica del mismo se complica bastante

---

cuando el modelo no es lineal.

$$Y_t = f(Y_t, X_t, \beta, \gamma) + \varepsilon_t$$

Una diferencia importante de los modelos no lineales es que el número de parámetros no se relaciona directamente con el número de variables  $X$  e  $Y$  en el modelo.

## 4.2. Introducción a las redes neuronales artificiales

Antes de introducir al lector en redes neuronales artificiales, conviene situar este campo en el entorno de la Inteligencia Artificial (IA). En la figura 4.1 se muestran varios conceptos relacionados:

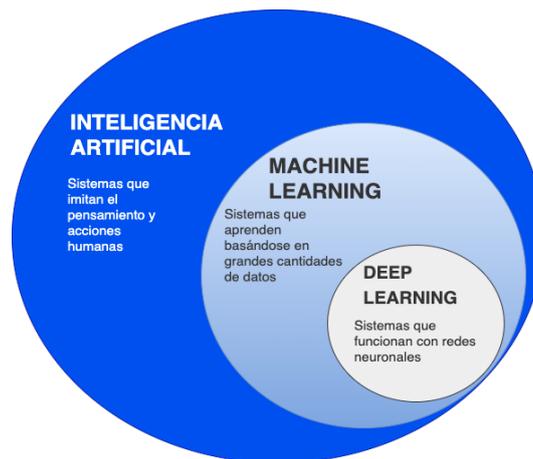


Figura 4.1: Inteligencia Artificial, Machine Learning y Deep Learning

Básicamente, se llama Inteligencia Artificial a la capacidad de las máquinas de adquirir capacidades propias de los humanos, como el razonamiento, el aprendizaje o la predicción. Dentro de la IA se encuentra el Machine Learning (ML), que puede entenderse cómo los computadores adquieren la capacidad de identificar patrones o elaborar predicciones a partir de algoritmos. El Deep Learning (DL), traducido como aprendizaje profundo, usa algoritmos similares al cerebro humano, siendo una Red Neuronal Artificial (RNA) la que le permite llevar a cabo el proceso de ML, es decir, DL es ML llevado a cabo por RNA.

Una RNA (o ANN, del inglés Artificial Neural Network) es una estructura matemática que busca relaciones en un conjunto de datos y es el resultado de búsqueda por parte de los humanos de la creación de un modelo computacional que simule funciones del cerebro humano [47].

En 1943, el neurobiólogo Warren McCulloch y el estadístico Walter Pitts publicaron el artículo "A logical calculus of Ideas Imminent in Nervous Activity" [48] y fueron los pioneros en el modelado artificial de una neurona imitando a las que constituyen el cerebro humano. Pero no fue hasta 1957 cuando el psicólogo Frank Rosenblatt creó una estructura llamada perceptrón [49]. Esta estructura, con una o varias entradas, una función de activación y una salida, posee aspectos similares al del cerebro y receptores de sistemas biológicos [50]. La figura 4.2 presenta de manera gráfica las similitudes entre una neurona biológica y otra artificial y muestra como las entradas del perceptrón actúan como las dendritas o señales de entrada hacia la neurona. La función de activación realiza el papel del soma que decide en qué momento disparar esa señal eléctrica y la salida modela al axón, que al activarse envía información hacia las neuronas con las que se conecta.

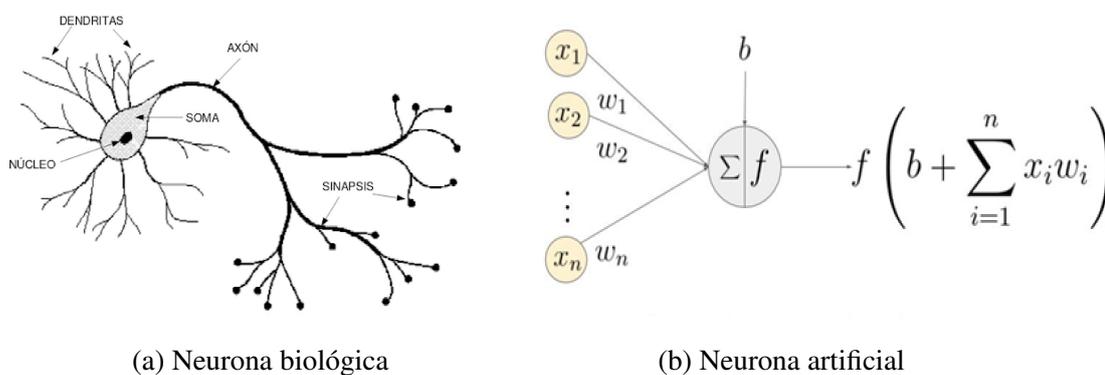


Figura 4.2: Similitudes entre una neurona biológica y una artificial

De manera simplificada, una RNA está formada por nodos interconectados organizados en una o varias capas, tal y como se muestra en la figura 4.3.

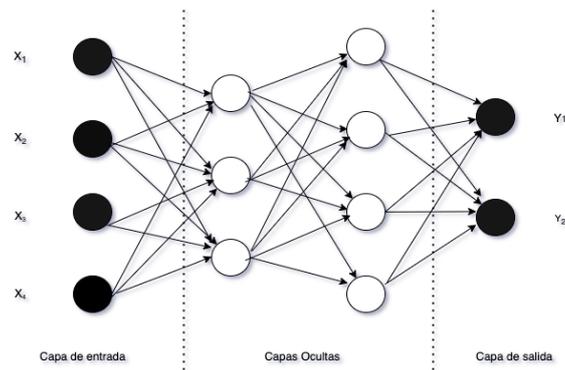


Figura 4.3: Ejemplo de estructura típica de una red neuronal artificial

Además de la similitud en la estructura, las RNA presentan otras características que se asemejan a las del cerebro como son el aprendizaje de la experiencia, generalización de casos previos a casos nuevos y el poder de abstracción de algunas características que poseen un conjunto de datos.

Entre otras aplicaciones, a través de una RNA se puede aproximar una amplia gama de modelos predictivos haciendo suposiciones mínimas sobre la forma y asunción de modelos. La forma de las relaciones queda determinada durante el proceso de aprendizaje, de tal manera que si la relación lineal entre el objetivo y los predictores es la correcta, los resultados de la red neuronal deben aproximarse mucho a los del modelo lineal clásico. Por el contrario, si una relación no lineal es más apropiada, la red neuronal aproximará esa estructura correcta del modelo. Esta capacidad de aproximación provoca que la red neuronal no sea fácilmente interpretable. Por lo tanto, si lo que se intenta es explicar el proceso subyacente que genera las relaciones entre el objetivo y los predictores, se debería utilizar mejor un modelo estadístico más tradicional. Sin embargo, si la interpretabilidad del modelo no es importante, pueden obtenerse buenas predicciones utilizando una red neuronal.

Algunas de las principales ventajas de las RNA es su habilidad para completar tareas con infinitas combinaciones, la capacidad para dar sentido a datos incompletos, ambiguos o contradictorios, y la capacidad para usar procesos controlados cuando no existe un modelo exacto que seguir. La combinación de los dos campos al abordar una red neuronal artificial, el biológico y el matemático proporcionan métodos cuantitativos y modelos de

aprendizaje aplicables a múltiples disciplinas. A lo largo de su evolución, han demostrado ser muy competitivas en la resolución de problemas del mundo real en comparación con otros tradicionales de análisis de datos, generalmente basados en el modelado estadístico [47]. Y de modo más concreto, la aplicación de las RNA a problemas de grandes dimensiones en los que las relaciones entre las variables no es fácil de determinar puede ser decisiva, tanto por las predicciones obtenidas como por el tiempo necesario para la resolución del problema.

### 4.2.1. Clasificación de las redes neuronales artificiales

Hay un elevado número de estructuras diferentes que pueden tener las redes neuronales, pudiendo incluir la mayoría tipologías como perceptrón multicapa (P), redes neuronales convolucionales (RNC), redes neuronales recurrentes (RNN) y redes neuronales de base radial (RNBR). La clasificación de las redes neuronales se puede llevar a cabo atendiendo al número de capas y al tipo y grado de conexiones que forman su arquitectura.

Por el número de capas, se pueden dividir en dos tipos:

**Redes neuronales monocapa** Estas redes son las más sencillas. La capa de entrada también puede considerarse una capa, pero al no hacer cálculos sobre ella no se tiene en cuenta a la hora de clasificarse. La capa de entrada se conecta a la capa de neuronas de salida que realizan ciertos cálculos.

**Redes neuronales multicapa** Entre las conexiones de entrada y de salida, existen diversas capas de neuronas que hacen de intermediarias, denominadas capas ocultas y que pueden o no conectarse entre ellas.

Atendiendo al tipo de conexiones, también se pueden clasificar en dos tipos:

**Redes neuronales no recurrentes** En este tipo de redes, la información viaja en un solo sentido, no existe una retroalimentación y carecen de memoria.

---

**Redes neuronales recurrentes** Las neuronas tienen la posibilidad de realizar conexiones de retroalimentación, ya sea entre neuronas de una misma capa o entre diferentes capas. La retroalimentación permite que las redes neuronales recurrentes tengan memoria. Este tipo tiene más ventajas en general y suelen ser más potentes que las no recurrentes.

Ambos tipos tienen en común algunos algoritmos como la fase de entrenamiento.

Según el grado de conexiones:

**Redes neuronales totalmente conectadas** Las capas anteriores y las capas posteriores, están totalmente conectadas

**Redes parcialmente conectadas** No todas las neuronas están conectadas entre ellas.

Y por último, atendiendo a su topología:

**Red neuronal Convolutiva (CNN)** En esta red, cada neurona no se une con todas y cada una de las capas siguientes, solo con un subgrupo de ellas consiguiendo especializarse, reduciendo el número de neuronas y la complejidad computacional necesaria para su ejecución.

**Red neuronal recurrente (RNN)** Estas redes no tienen una estructura de capas, sino que permiten conexiones arbitrarias entre las neuronas, crear ciclos y así aportar temporalidad y permitiendo que la red tenga memoria.

**Redes de base radial (RBF)** Calculan la salida de la función según la distancia a un punto denominado centro. La salida es una combinación lineal de las funciones de activación radiales utilizadas por las neuronas individuales. Tienen la ventaja de que no presentan mínimos locales y así evitar que la retropropagación se quede bloqueada.

### 4.2.2. Algoritmos de una RNA

Existen infinitos algoritmos para llevar a cabo una red neuronal y el objetivo de esta sección es citar y describir brevemente algunos de ellos así como los cálculos que más

---

se utilizan. Estos algoritmos son los encargados de añadir no linealidad a cada una de las neuronas, facilitando la resolución de problemas no triviales con pocas neuronas. Estas funciones reciben la salida de cada neurona y son las encargadas de pasar el resultado a la siguiente capa.

**Descenso de gradiente** Se trata de uno de los algoritmos de optimización más conocido con el que las redes neuronales consiguen aprender basando su objetivo en alcanzar el punto mínimo de un error. Además de utilizarse para redes neuronales, también es compatible con cálculos de regresión lineal, regresión logística o máquinas de soporte vectorial, haciendo que sean más complejos. Existen varios tipos de descenso de gradiente, como Batch Gradient, Mini-Batch o descenso de gradiente estocástico.

**Función sigmoide** Se puede observar en procesos de aprendizaje automático e incluso en procesos naturales y se utiliza para trazar la evolución de un proceso en un determinado espacio de tiempo. Por ejemplo, en la inteligencia artificial se utiliza para la evolución y conseguir llegar a una tasa de errores más baja y mejorar las predicciones.

**Backpropagation o Propagación hacia atrás** Este cálculo se utiliza para trabajar en diversas capas de una red neuronal. El algoritmo se propaga desde las capas de entrada hasta las de la salida: al llegar a la capa de salida, se calcula un margen de error con el resultado de la salida y el resultado deseado. La señal de error se propaga a las capas anteriores, comenzando por las de salida, hasta las capas intermedias, de forma que se ajusta el peso de las neuronas artificiales. Al reducir el margen de error, se consigue una predicción más exacta. Con la función sigmoide se puede observar la evolución en la que el margen de error baja.

**Descenso de Newton** Es una aplicación del cálculo diferencial que se aplica para encontrar los ceros y raíces de una función derivable de  $n$ -ésimo grado. Este método, en las redes neuronales se utiliza para la optimización.

**Aprendizaje de cuantificación vectorial** A partir de la creación de un diccionario de

---

prototipos, se realizan entrenamientos de las capas supervisadas de una red neuronal, facilitando a esta técnica el aprendizaje para interpretar diversos patrones.

### 4.2.3. Principales modelos de RNA

**Red de Hopfield** Se usa como sistema de memoria asociativa con unidades binarias y están diseñadas para converger a un mínimo local, aunque la convergencia a uno de los patrones almacenados no está garantizada.

**Self-Organizing-Map (SOM)** Es una red entrenada usando aprendizaje no supervisado para producir una representación discreta del espacio de las muestras de entrada, llamado mapa.

**Máquinas de Boltzman y Cauchy** Son redes que usan estados binarios, conexiones bidireccionales, transiciones probabilísticas y puede tener unidades ocultas.

**Teoría de la resonancia adaptativa (ART)** Esta red facilita el aprendizaje de nueva información sin destruir la ya almacenada, basada en el dilema de la estabilidad y plasticidad del aprendizaje.

**Red de Elman** Esta red incorpora la retroalimentación, lo que permite incluir temporalidad consiguiendo a la red que tenga memoria.

**Red de Jordan** Es una red recurrente y puede considerarse como una red de propagación directa con neuronas adicionales de contexto en la capa de entrada.

## 4.3. Aproximación de las redes neuronales a la modelización econométrica

La capacidad de las RNAs para aprender, generalizar y retener conocimiento de los datos consiguen que, desde un punto de vista econométrico y estadístico, puedan ser considerados como modelos de regresión o modelos discriminantes no lineales [51]. Los modelos neuronales son, básicamente, una forma de inferencia estadística no paramétrica

---

inspirados en los sistemas biológicos, que junto al uso de algoritmos en el proceso de aprendizaje, muestran evidencia de lo próximo que se encuentran de la definición de los modelos de regresión no lineales. Modelos como por ejemplo, las redes feedforward o perceptrón, poseen la capacidad de ser utilizadas como aproximadores universales de funciones. Desde la perspectiva econométrica, a los modelos neuronales se les puede asociar el siguiente planteamiento: existe una función desconocida a priori,  $f(x)$  con cierto componente estocástico, donde el proceso de aprendizaje del modelo consiste en calcular un estimador de la función desconocida,  $f(x; w) = \hat{f}(x)$ , siendo  $w$  el vector de parámetros desconocidos y  $x$  el conjunto de datos observados. El modelo neuronal definido es por lo tanto un estimador no paramétrico de la esperanza matemática de  $y$  condicionada a  $x$ , es decir,  $E(y/x)$ , planteamiento semejante a la interpretación de la inferencia clásica a través de los métodos estadísticos.

A continuación se presentan algunos modelos econométricos junto a la topología de la red neuronal equivalente:

**Modelo de regresión lineal** Estos modelos pueden ser representados por una red neuronal de dos capas con función de transferencia lineal. Esta red también se denomina Adaline (ADaptive LINear Element), y en su versión múltiple Madaline, que permite representar un modelo de ecuaciones aparentemente no relacionadas.

**Modelo autorregresivo** Si se utilizan outputs retardados como entradas en una red Adaline, se obtiene una ecuación formada por elementos temporales de carácter lineal AR(p), es decir, un modelo autorregresivo cuyo orden es igual al número de ponderaciones del modelo neuronal, con un bias nulo.

**Modelo Probit y Logit** Estos modelos son adecuados para problemas de clasificación a los que puede programarse una red Adaline pero asociando a la salida una función binaria.

Resolver un modelo econométrico a través de uno neuronal presenta algunos inconvenientes, como por ejemplo:

---

- i. No existe ninguna fórmula para determinar la estructura óptima de un modelo neuronal, tanto en la determinación del número de capas, el número de neuronas en la capa oculta, etc.
- ii. No existe algoritmo óptimo que proporcione el mínimo global en la superficie de error cuando hay mínimos locales.
- iii. Las propiedades estadísticas de las redes neuronales generalmente no están disponibles y es muy difícil interpretar los parámetros del modelo neuronal una vez terminado el proceso de aprendizaje.

Por otro lado, existen ventajas al aplicar una RNA a un modelo econométrico, proporcionando un gran peso a favor de su uso. Por ejemplo:

- i. Es válida cuando no se tiene un modelo exacto del proceso suficientemente preciso para programarlo.
- ii. Útil para inferir el conocimiento subyacente en las observaciones.
- iii. Aconsejable cuando los datos son tan complejos que es imposible resolverlo en tiempos razonables con los métodos tradicionales.

#### **4.3.1. Las redes neuronales artificiales como respuesta al modelo de ecuaciones simultáneas no lineales**

Los modelos de ecuaciones simultáneas pueden ser resueltos con métodos que obtienen buenos resultados, pero bajo ciertos supuestos que no siempre son admisibles. La modelización lineal no siempre está justificada y el uso de las redes neuronales puede ofrecer buenos resultados dada su capacidad para completar tareas con múltiples combinaciones, dar sentido a datos incompletos o ambiguos y la capacidad para usar procesos controlados cuando no existe un modelo exacto que seguir.

Las aplicaciones de las RNAs a los MESs son diversas y a continuación se citan algunas de ellas. Kumar en 1991 [52] presenta un enfoque para formular y estimar un MES de

---

la economía estadounidense como un problema de redes neuronales en el que concluye que este nuevo enfoque es prometedor aunque con reservas debido al pequeño tamaño del problema. Un estudio muestra los resultados de la comparación al estimar el modelo de Klein I a través de métodos estadísticos frente a una red neuronal con dos capas ocultas, en el que destaca la factibilidad y los buenos resultados de aplicar RNAs a este tipo de problema [53]. Un trabajo similar, de Caporletti, también resuelve el problema económico de Klein planteado como un MES de dos ecuaciones comparando los resultados de estimar por métodos clásicos como MC2E y MC3E con una RNA [54], obteniendo resultados muy parecidos entre los diferentes métodos. Por otro lado, Yaxue en 2021 [55] estudia la difusión de artículos científicos del mundo académico a través de las redes sociales usando un MES y una red neuronal artificial.

#### 4.4. Conclusiones

Una vez revisados los métodos de estimación para los modelos de ecuaciones simultáneas no lineales, se observa que en muchas ocasiones los métodos de estimación del caso lineal sufren variaciones para dar solución al no lineal. La estimación de los modelos no lineales se vuelve compleja según aumentan las no linealidades a lo largo de todas las ecuaciones del modelo, lo que puede resultar inviable en modelos de grandes dimensiones. Una alternativa para resolver estos problemas es la aplicación de las redes neuronales artificiales, técnica que no necesita un modelado rígido previo, al estar basadas en un concepto de entrenamiento-aprendizaje. La elección de una u otra metodología para resolver el problema dependerá sobre qué parámetros se tiene interés en estimar: si el objetivo es estimar el efecto que una variable causa sobre otra u otras variables, se debe utilizar una metodología de estimación estadística, pues proporciona una estimación de los coeficientes del modelo y por tanto, del efecto que causan unas variables en otras. Por el contrario, si el interés radica únicamente en estimar los valores predichos de las variables endógenas, ambas técnicas serán válidas para el propósito.

---



## CAPÍTULO 5

---

# **Nuevo Método de Estimación y Criterio de Información de un Modelo de Ecuaciones Simultáneas basado en la Entropía**

---

En este capítulo se propone un nuevo método de estimación de un MES y un nuevo criterio de información para la comparación y selección del modelo de entre un conjunto de modelos estimados. El nuevo método que se presenta obtiene la estimación de los parámetros del modelo en base a la optimización de los coeficientes del método BMOM, minimizando el Criterio de Información de Akaike (AIC). Para el estudio se han generado MES de manera aleatoria a partir de un algoritmo generador, comparándose las estimaciones obtenidas por los diferentes métodos así como la bondad de las mismas, bondad que ha sido medida a través del AIC y la entropía. Los resultados muestran que el método de estimación propuesto ofrece mejores aproximaciones y que la nueva medida de entropía puede ser una alternativa al AIC como criterio de selección del mejor método

de estimación de un MES.

## 5.1. Nuevo método de estimación: BMOM optimizado

En el capítulo 2 (sección 2.3) se introdujo el método BMOM en el que el autor propone valores concretos a los parámetros  $K_1$  y  $K_2$  en función del tipo de estimación: *bondad de ajuste*, *precisión en la estimación* o *balanceado*, tanto para el Enfoque Bayesiano como el Tradicional. En este trabajo se propone una variación de BMOM obteniendo  $K_1$  y  $K_2$  mediante la optimización de un parameter criteria. Concretamente, se propone la optimización de los parámetros  $K_1$  y  $K_2$  que minimizan el AIC [33], que es una medida de calidad de los modelos estadísticos basado en el ajuste de la muestra y en la bondad de la estimación. Por lo tanto, dada una colección de modelos estimados con diferentes métodos para una muestra dada, AIC estima la calidad de cada método, proporcionando así una herramienta para la selección del método de estimación con menor error.

La expresión para esta medida es:

$$AIC = n \ln |\hat{\Sigma}_e| + 2 \sum_{i=1}^m (m_i + k_i - 1) + m(m+1) \quad (5.1)$$

donde  $n$  es el tamaño muestral,  $m$  es el número de ecuaciones,  $m_i$  y  $k_i$  son el número de variables endógenas y exógenas en la ecuación  $i$ -ésima y  $\hat{\Sigma}_e$  la matriz de varianza-covarianza de los errores definidos como  $e_j = Y_j - \hat{Y}_j$ ,  $j=1, \dots, m$ .

Para el estudio experimental se ha generado un gran número de MES de manera aleatoria y bajo ciertas condiciones (ver sección 5.3). A partir de estos modelos, a los que llamaremos modelos *reales* se han calculado unos nuevos errores,  $e_j^r$ , obtenidos como diferencia entre los valores *reales*,  $Y_j^r$  y los valores estimados por cada método,  $\hat{Y}_j$ . Estos errores, que constituyen la matriz de varianzas-covarianzas, se sustituyen en la ecuación (5.1), obteniendo una nueva medida que se denota por  $AIC_{real}$ .

$$AIC_{real} = n \ln |\hat{\Sigma}_e^r| + 2 \sum_{i=1}^m (m_i + k_i - 1) + m(m+1) \quad (5.2)$$

Para alcanzar el valor mínimo de  $AIC_{real}$ , se utiliza un algoritmo basado en el método

Cuasi-Newton, de modo que en cada iteración este algoritmo busca una aproximación de la inversa de la matriz Hessiana, garantizando así la aproximación en cada paso del proceso. Concretamente, se ha utilizado la función *optim* implementada en  para obtener los valores óptimos con el método L-BFGS-B.

## 5.2. Entropía como parámetro de criterio de información

El desarrollo de la entropía y sus propiedades han sido presentadas en el capítulo 3 y en esta sección se describen las fórmulas de las medidas que han sido obtenidas en el experimento como criterio de selección del método de estimación.

Clausius [56] en 1879 introdujo la entropía en termodinámica para medir la cantidad de energía en un sistema que no puede producir trabajo, este concepto aparece en muchos contextos (mecánica estadística, teoría de la información, etc.) como desorden, incertidumbre, aleatoriedad, complejidad, etc. Claude Shannon [57] en 1948 construyó su teoría de la información y la comunicación, siendo generalizada por algunos autores como Tsallis [58].

La expresión para la entropía de Shannon [59] es:

$$H_S(x) = - \sum_{i=1}^n p(x_i) \log p(x_i) \quad (5.3)$$

Se han obtenido tres nuevas medidas como criterio de parámetro de calidad de un MES basadas en entropía, que se enumeran a continuación:

- i. Como primera medida de entropía calculada en este trabajo se ha utilizado una versión la expresión desarrollada por Amigó et al. [60] como una variación de la entropía generalizada, que permite contemplar valores pequeños de  $p$  en la distribución. La expresión para esta entropía, denotada por  $H_A(p)$ , se define como:

$$H_A(p) = \prod_{i=1}^n (2 - (p_i)^{p_i}) \quad (5.4)$$

En nuestro problema, esta expresión debe ser aplicada a un conjunto de errores  $e_1, e_2, \dots, e_n$  para las  $m$  variables endógenas,  $Y_1, Y_2, \dots, Y_m$ , obteniendo  $m$  medidas de entropía, una para cada variable endógena:

$$H_j(e) = \prod_{i=1}^n (2 - (p_{ij})^{p_{ij}}) \quad j = 1, 2, \dots, m \quad (5.5)$$

donde  $n$  es el tamaño de la muestra,  $m$  el número de variables endógenas, y los valores  $p_{ij}$  han sido calculados de la siguiente manera:

$$p_{ij} = \frac{e_{ij}}{\sum_{i=1}^n e_{ij}} \quad j = 1, 2, \dots, m, \quad \text{donde} \quad e_{ij} = Y_{ij} - \widehat{Y}_{ij} \quad (5.6)$$

es decir,  $p_{ij}$  se puede interpretar como la masa de error en cada variable endógena calculada a partir de la matriz de error como la diferencia entre las variables endógenas y su estimación a través de cada método. Finalmente, para el modelo estimado, el promedio del logaritmo de la expresión (5.5) se calcula como un criterio de parámetro para seleccionar un método de estimación:

$$H(e) = \frac{\sum_{j=1}^m \ln H_j(e)}{m} \quad (5.7)$$

El valor mínimo de  $H(e)$  se alcanza cuando  $e_{ij}$  son homogéneos y, por tanto, los valores de  $\widehat{Y}_{ij}$  están más equilibrados.

- ii. La segunda medida basada en la entropía se deriva directamente de la expresión (5.4). Considerando los valores  $p_{ij}$  calculados en (5.6) esta medida se obtiene como:

$$H_2(e) = \frac{\prod_{j=1}^m (2 - (p_{ij})^{p_{ij}})}{m} \quad (5.8)$$

- iii. Y por último, para la introducción a la tercera medida basada en entropía, se comienza con la definición para un conjunto de datos  $x_1, x_2, \dots, x_n$ , de una media generalizada o media de Hölder con parámetro  $w$ , que se define como:

$$M_w(x) = \left( \frac{1}{n} \sum_{i=1}^n x_i^w \right)^{\frac{1}{w}} \quad (5.9)$$

y utilizando la definición de norma se tiene la siguiente igualdad:

$$\|x\|_w = M_w(|x|) \sqrt[w]{n} \quad (5.10)$$

Sustituyendo  $x_1, x_2, \dots, x_n$ , por  $p_{1j}, p_{2j}, \dots, p_{nj}$ , con  $j = 1, \dots, m$

$$M_{1/2}(e) = \left( \frac{\sum_{j=1}^m p_{ij}^{1/2}}{m} \right)^2 \quad (5.11)$$

que es la media de Hölder con parámetro  $w = \frac{1}{2}$ , quedando finalmente la expresión para la medida como:

$$\|e\|_{1/2} = M(|e|) \sqrt[1/2]{m} \quad (5.12)$$

En la tabla 5.1 se muestra un resumen de las tres medidas basadas en entropía que serán utilizadas como criterio de parámetro para la selección del método de estimación en el estudio experimental llevado a cabo. Dado que las medidas se obtienen del concepto de entropía, la elección de una u otra dependerá de la simplicidad de cálculo según los datos del problema, puesto que todas ellas conducen a seleccionar el mismo método.

Tabla 5.1: Medidas de entropía como criterio de parámetro para la selección del método de estimación

Medida	Expresión
$H(e)$	$\frac{\sum_{j=1}^m \ln H_j(e)}{m}$
$H_2(e)$	$\frac{\prod_{j=1}^m (2 - (p_{ij})^{p_{ij}})}{m}$
$M_{1/2}(e)$	$\left( \frac{\sum_{j=1}^m p_{ij}^{1/2}}{m} \right)^2$

### 5.3. Estudio experimental

En esta sección se muestran los resultados experimentales llevados a cabo para el estudio comparativo de los métodos de estimación y las medidas de entropía propuestas anteriormente. Para el estudio se han generado varios MES, cuyos coeficientes se han comparado con las estimaciones obtenidas con los diferentes métodos. Los MES, tanto los coeficientes del modelo como los valores de las variables, se han generado aleatoriamente sin ningún patrón a seguir en las ecuaciones, exigiendo tan solo que todas las ecuaciones sean identificadas. Concretamente, la generación de los MES se ha llevado de la siguiente manera: los valores de las matrices  $B$  y  $\Gamma$  se generan aleatoriamente siguiendo una distribución uniforme en  $[0, 10]$ , la matriz  $X$  una distribución normal multivariante y, finalmente, la matriz  $Y$  como  $\Pi X$  más una distribución Normal con media 0 y sigma 0.1, siendo  $\Pi = -\Gamma B^{-1}$ .

Los métodos utilizados en el estudio comparativo son MC2E, BMOM, ambos revisados en las secciones 2.3.3 y 2.3.5 respectivamente, el nuevo método propuesto  $Bmom_{OPT}$  desarrollado en la sección 5.1 y los métodos  $Bayes_{2E}$  y MCMC que son una particularización de lo introducido también en las secciones 2.3.6 y 2.3.7.

Para llevar a cabo la estimación, se han utilizado dos funciones del software  [61]: en MCMC, las cadenas de  $D_{\delta, \hat{\delta}}$  se han simulado a través de la función *MCMCregress* del paquete MCMCpack sin información previa y con el algoritmo de Gibbs, obteniendo el promedio para inferir los parámetros del modelo.

También se ha recurrido a  en la aplicación del método  $Bmom_{OPT}$ , concretamente se ha utilizado la función *optim* para obtener los valores óptimos  $K_1$  y  $K_2$ . Se ha utilizado el método quasi-newton L-BFGS-B acotado en  $[-3, +3]$ . En las pruebas llevadas a cabo, este método, que hace un uso limitado de memoria, ha sido el que ha resultado el mejor en la convergencia al óptimo, tanto en tiempo como en los saltos en los que concurría cuando alguna de las ecuaciones del modelo carecía de variables endógenas como variables explicativas.

Por otro lado, el método  $Bayes_{2E}$  que aquí se propone, puede considerarse una

---

técnica híbrida de inferencia: usa una distribución Normal Inversa Gamma antes de obtener expresiones analíticas exactas para la distribución posterior de los coeficientes estructurales  $B$  y  $\Gamma$  del MES en *dos etapas*. De manera análoga a la metodología MC2E, este método aplica en los dos pasos *mínimos cuadrados Bayesianos* en lugar de MCO, obteniendo en primera etapa las *proxys* y en la segunda, después de sustituir éstas por las variables endógenas, obtiene los parámetros del modelo estimados. A continuación se explica brevemente este proceso de inferencia.

Sea la primera ecuación del MES en primera etapa para estimar las *proxys*:

$$y_1 = Y_1\beta_1 + X_1\gamma_1 + u_1 \quad (5.13)$$

$$Y = X\Pi + E, \quad E \sim N(0, \sigma^2 I) \quad (5.14)$$

La distribución a priori Normal Inversa Gamma (NIG) para  $(\pi, \sigma)$  viene dada por:

$$\pi | \sigma^2 \sim N(\alpha_0, \sigma^2 V_0) \quad y \quad \sigma^2 \sim IG(a_0, b_0) \quad (5.15)$$

La distribución posterior se puede calcular analíticamente, siendo una distribución a priori NIG, la posterior también es una distribución NIG, entonces

$$\pi, \sigma^2 \sim NIG(\alpha_1, V_1, a_1, b_1) \quad (5.16)$$

donde

$$\begin{aligned} \alpha_1 &= (V_0^{-1} + X^T X)^{-1} (V_0^{-1} \alpha_0 + X^T y) \\ V_1 &= (V_0^{-1} + X^T X)^{-1} \\ a_1 &= a_0 + \frac{1}{2}n \\ b_1 &= b_0 + \frac{1}{2} (\alpha_0^T V_0^{-1} \beta_1 + y^T y - \beta_1^T V_1^{-1} \alpha_1) \end{aligned} \quad (5.17)$$

Sustituyendo las *proxys* en la ecuación (5.13):

$$y_1 = \hat{Y}_1 \beta_1 + X_1 \gamma_1 + u_1 \Rightarrow Y = (\hat{Y}X) \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + U \Rightarrow Y = Z\Phi + U \quad (5.18)$$

Y por último, en la segunda etapa, se vuelve a aplicar el mismo procedimiento. Los parámetros iniciales en la distribución previa Normal Inversa Gamma han sido un promedio de 5 y una precisión de 0,2.

Finalmente, para llevar a cabo la comparación entre los diferentes métodos, se han obtenido las siguientes medidas de calidad de los modelos estimados:

- i. La Norma Euclídea entre  $\delta$  y  $\hat{\delta}$ , denotada por  $D_{\delta, \hat{\delta}}$ , donde  $\delta = [B \Gamma]$  es la matriz de coeficientes y  $\hat{\delta}$  es la matriz de coeficientes estimados.
- ii.  $AIC_{real}$ , basado en los errores calculados como diferencia entre las variables endógenas generadas y sus estimaciones.
- iii.  $AIC$ , basado en los errores calculados como diferencia entre las variables endógenas y sus estimaciones.
- iv.  $AIC_{MC2E}$ , basado en los errores calculados como diferencia entre las estimaciones de las variables endógenas obtenidas con MC2E y las obtenidas por cada método .
- v. La entropía  $H(e)$ .
- vi. La entropía  $H_2(e)$ .
- vii. La entropía  $M_{1/2}(e)$ .
- viii. El tiempo de ejecución.

Para todas las medidas anteriores se han obtenido el promedio y la desviación estándar de 50 simulaciones de varios MES que difieren en número de variables endógenas, exógenas y tamaño muestral. Los resultados mostrados en las tablas 5.2 - 5.9 se han obtenido aplicando un valor de sigma igual a 0,1 en la generación de los MES.

La norma euclídea,  $D_{\delta, \hat{\delta}}$  se ha obtenido como diferencia entre los coeficientes del modelo generado  $\delta$ , y los coeficientes estimados  $\hat{\delta}$  por cada método. Los resultados se muestran

---

en la tabla 5.2, y se observa que en todos los modelos simulados la media de la norma más baja es proporcionada por el método  $Bmom_{OPT}$ , aumentando el valor de la media según aumenta el número de parámetros a estimar y disminuyendo según el tamaño de la muestra aumenta. Le siguen las normas proporcionadas por los métodos MC2E y  $Bayes_{2E}$ .

Los métodos BMOM tienen resultados muy similares, con muy poca diferencia entre ellos en todos los modelos. El método que proporciona estimaciones con la norma más alta es MCMC. Y como era de esperar, aumentar el tamaño de la muestra manteniendo el número de variables endógenas y exógenas fijas, reduce la norma de los coeficientes en todos los métodos, excepto MCMC, en el que no se aprecia ninguna tendencia.

Además de los resultados mostrados, se han simulado MES con otros valores de sigma, obteniendo así diferentes variabilidades en los datos. En general, se ha observado que, cuando esta variabilidad es elevada, no hay un buen método para estimar, algo que por otro lado es fácil de intuir. Por ejemplo, en el caso sigma 0,25, el valor promedio de la norma nuevamente es menor en  $Bmom_{OPT}$ , excepto en un modelo, que es menor en  $Bayes_{2E}$ , siendo la diferencia entre ambos muy pequeña. El método MCMC es el que tiene la norma más alta respecto al resto de métodos en todos los MES simulados. Respecto a la desviación estándar, los valores medios obtenidos son similares en todos los modelos y métodos.

La tabla 5.3 muestra la media de  $AIC_{real}$  en las 50 simulaciones. Esta medida es una versión del AIC cuya fórmula usa los errores obtenidos como diferencia entre las variables endógenas generadas y las estimadas. El valor mínimo de esta medida se obtiene con  $Bmom_{OPT}$  en todos los modelos simulados seguido de MC2E y  $Bayes_{2E}$ , que tienen valores muy similares entre ellos. Los BMOM vuelven a obtener medias muy próximas y MCMC es el método que obtiene esta medida con los valores más elevados. Al comparar entre modelos con diferente número de variables,  $AIC_{real}$  aumenta a medida que aumenta el número de variables y, al comparar modelos con el mismo número de variables, esta medida aumenta a medida que aumenta el tamaño de la muestra. Al aumentar el valor de sigma y, por tanto, la variabilidad de las variables endógenas, el resultado sigue un patrón similar al obtenido con sigma 0,1. Los valores medios obtenidos de la desviación

---

estándar de  $AIC_{real}$  aumenta según aumenta la dimensión del MES, no presentando grandes diferencias entre los diferentes métodos para un mismo modelo.

En la tabla 5.4 se muestran los valores del AIC, observando que el valor medio de las 50 simulaciones es más pequeño en los métodos BMOM, con muy poca diferencia entre ellos en todos los modelos simulados. MC2E es el segundo método que proporciona valores medios más pequeños, seguido de  $Bayes_{2E}$ ,  $Bmom_{OPT}$  y finalmente MCMC. Al igual se observaba en las medidas anteriores, al variar el valor sigma en las simulaciones, los resultados no varían significativamente de los obtenidos con sigma 0,1. Las desviaciones obtenidas varían proporcionalmente a las dimensiones del modelo, con un resultado similar entre métodos de estimación.

La medida denotada por  $AIC_{MC2E}$  se ha obtenido en base a los errores calculados como diferencia entre los estimados con MC2E y cada uno del resto de métodos. Los resultados de en la tabla 5.5 muestran conclusiones similares a los valores  $AIC_{real}$  comentados anteriormente, obteniendo el valor mínimo con el método  $Bmom_{OPT}$  en todos los modelos simulados y aumentando según aumenta la dimensión del problema. La desviación típica también presenta un resultado similar, aumentando al incrementar el tamaño del MES.

La primera medida de entropía calculada,  $H(e)$ , se muestra en la tabla 5.6. Los resultados muestran que no hay un método que proporcione el valor medio mínimo en todos los modelos. Por el contrario, son los métodos  $Bmom_{OPT}$ , MC2E y  $Bayes_{2E}$  los que de manera arbitraria ofrecen el valor medio mínimo de  $H(e)$ . En las simulaciones llevadas a cabo con sigma 0,5 se han obtenido resultados similares a estos. La desviación típica aumenta al aumentar el tamaño del problema.

Respecto a la segunda medida de entropía calculada,  $H_2(e)$ , los resultados se recogen en la tabla 5.7, observando que el valor medio mínimo se alcanza con  $Bmom_{OPT}$ , obteniendo empate en casi todos los modelos con MC2E. Les siguen los métodos BMOM,  $Bayes_{2E}$  y por último MCMC. Esta entropía lleva a los mismos resultados que la norma  $D_{\delta, \hat{\delta}}$ .

La tercera entropía calculada,  $M_{1/2}(e)$ , cuyos resultados quedan reflejados en la tabla 5.8, muestra valores muy próximos entre los métodos  $Bmom_{OPT}$ ,  $Bayes_{2E}$  y MC2E.

Y finalmente, el tiempo de ejecución recogido en la tabla 5.9 muestra que  $Bayes_{2E}$  y

MC2E necesitan un menor tiempo para todas los modelos simulados. Les siguen los métodos BMOM y finalmente, MCMC y  $Bmom_{OPT}$ , que son los que necesitan un tiempo más elevado para la estimación. El tiempo aumenta como consecuencia del aumento del tamaño de la muestra y el número de variables y no cambian entre modelos cuando aumenta la variabilidad. En cuanto a la desviación típica de las 50 simulaciones, estas son proporcionales al valor medio del tiempo usado en el cómputo.

Tabla 5.2: Media y desviación estándar de 50 simulaciones de  $D_{\delta, \hat{\delta}}$ . Sigma 0,1.

m	k	n	MC2E	BMOM			$Bmom_{OPT}$	$Bayes_{2E}^a$	MCMC <sup>b</sup>
				Bondad ajuste	Precisión estimación	Tradicional			
10	20	100	27,670 <sub>7,778</sub>	40,914 <sub>7,538</sub>	40,966 <sub>7,546</sub>	42,133 <sub>7,514</sub>	<b>20,647</b> <sub>8,826</sub>	33,673 <sub>12,084</sub>	71,410 <sub>10,491</sub>
10	40	100	40,927 <sub>9,104</sub>	58,635 <sub>8,039</sub>	58,932 <sub>8,100</sub>	61,058 <sub>7,818</sub>	<b>26,769</b> <sub>8,209</sub>	56,340 <sub>13,721</sub>	91,076 <sub>4,918</sub>
20	60	100	115,852 <sub>8,537</sub>	141,029 <sub>6,294</sub>	140,906 <sub>6,250</sub>	147,628 <sub>5,166</sub>	<b>94,999</b> <sub>12,498</sub>	146,640 <sub>5,904</sub>	163,771 <sub>6,092</sub>
10	20	400	16,563 <sub>11,257</sub>	27,508 <sub>6,955</sub>	27,534 <sub>6,974</sub>	27,782 <sub>6,981</sub>	<b>10,619</b> <sub>5,868</sub>	22,233 <sub>17,026</sub>	70,449 <sub>7,899</sub>
10	40	400	15,199 <sub>4,366</sub>	30,538 <sub>6,467</sub>	30,494 <sub>6,446</sub>	30,851 <sub>6,490</sub>	<b>7,923</b> <sub>2,576</sub>	26,009 <sub>22,856</sub>	90,357 <sub>5,568</sub>
10	40	1000	7,394 <sub>2,830</sub>	17,218 <sub>5,450</sub>	17,229 <sub>5,458</sub>	17,330 <sub>5,477</sub>	<b>5,130</b> <sub>1,944</sub>	9,233 <sub>7,540</sub>	95,210 <sub>10,424</sub>

<sup>a</sup> Todos los parámetros iniciales tomaron el valor 5 y precisión 0,2. <sup>b</sup> Sin información previa. Tamaño de la cadena igual a 10.000.

Tabla 5.3: Media y desviación estándar de 50 simulaciones de  $AIC_{real}$ . Sigma 0,1.

m	k	n	MC2E	BMOM			$Bmom_{OPT}$	$Bayes_{2E}^a$	MCMC <sup>b</sup>
				Bondad ajuste	Precisión estimación	Tradicional			
10	20	100	1361,764 <sub>781,683</sub>	1895,361 <sub>801,349</sub>	1896,600 <sub>800,971</sub>	1951,348 <sub>801,768</sub>	<b>1156,465</b> <sub>792,277</sub>	1993,402 <sub>831,727</sub>	4598,074 <sub>395,783</sub>
10	40	100	1915,270 <sub>651,931</sub>	2345,388 <sub>636,213</sub>	2351,286 <sub>635,582</sub>	2414,261 <sub>630,277</sub>	<b>1718,000</b> <sub>626,334</sub>	2531,052 <sub>715,689</sub>	4547,361 <sub>286,781</sub>
20	60	100	5941,844 <sub>779,559</sub>	6458,904 <sub>762,773</sub>	6456,303 <sub>763,063</sub>	6620,917 <sub>760,210</sub>	<b>5692,839</b> <sub>793,354</sub>	6732,681 <sub>777,638</sub>	10172,018 <sub>206,247</sub>
10	20	400	4438,187 <sub>3271,180</sub>	7319,586 <sub>3449,337</sub>	7323,685 <sub>3450,560</sub>	7390,043 <sub>3453,242</sub>	<b>3409,573</b> <sub>3232,656</sub>	7517,898 <sub>5184,224</sub>	22875,120 <sub>1641,293</sub>
10	40	400	4645,919 <sub>2358,815</sub>	6989,900 <sub>2744,681</sub>	6983,853 <sub>2744,491</sub>	7045,436 <sub>2752,133</sub>	<b>3877,930</b> <sub>2299,173</sub>	7382,701 <sub>4943,642</sub>	22057,334 <sub>917,641</sub>
10	40	1000	6824,562 <sub>8001,673</sub>	11738,582 <sub>8294,696</sub>	11742,499 <sub>8294,949</sub>	11799,305 <sub>8299,763</sub>	<b>5913,439</b> <sub>8123,421</sub>	9046,596 <sub>8850,159</sub>	63247,410 <sub>2586,679</sub>

<sup>a</sup> Todos los parámetros iniciales tomaron el valor 5 y precisión 0,2. <sup>b</sup> Sin información previa. Tamaño de la cadena igual a 10.000.

Tabla 5.4: Media y desviación estándar de 50 simulaciones de  $AIC$ . Sigma 0,1.

m	k	n	MC2E	BMOM			$Bmom_{OPT}$	$Bayes_{2E}^a$	MCMC <sup>b</sup>
				Bondad ajuste	Precisión estimación	Tradicional			
10	20	100	2168,030 <sub>854,334</sub>	1784,122 <sub>808,851</sub>	1783,602 <sub>808,150</sub>	<b>1765,455</b> <sub>804,426</sub>	2419,415 <sub>893,327</sub>	2391,887 <sub>785,620</sub>	4413,688 <sub>421,952</sub>
10	40	100	2009,636 <sub>639,153</sub>	1850,753 <sub>626,141</sub>	<b>1850,716</b> <sub>626,059</sub>	1853,913 <sub>624,351</sub>	2372,355 <sub>692,930</sub>	2254,874 <sub>728,737</sub>	4348,852 <sub>256,390</sub>
20	60	100	3866,102 <sub>1051,585</sub>	3647,221 <sub>1056,690</sub>	<b>3645,439</b> <sub>1056,018</sub>	3723,252 <sub>1059,006</sub>	4543,780 <sub>1054,504</sub>	4119,416 <sub>1128,524</sub>	9856,712 <sub>241,454</sub>
10	20	400	15027,648 <sub>3347,331</sub>	13448,372 <sub>3161,253</sub>	13446,459 <sub>3161,039</sub>	<b>13420,826</b> <sub>3156,957</sub>	15524,231 <sub>3459,419</sub>	15626,142 <sub>3279,198</sub>	22160,759 <sub>1516,667</sub>
10	40	400	12849,587 <sub>2699,761</sub>	11990,079 <sub>2509,827</sub>	11991,659 <sub>2509,734</sub>	<b>11976,753</b> <sub>2506,040</sub>	13479,303 <sub>2744,344</sub>	13606,665 <sub>2961,619</sub>	21421,728 <sub>858,468</sub>
10	40	1000	37879,770 <sub>9621,598</sub>	36438,251 <sub>9459,390</sub>	36437,331 <sub>9459,167</sub>	<b>36425,550</b> <sub>9457,322</sub>	38479,010 <sub>9605,284</sub>	38035,942 <sub>9483,024</sub>	61720,600 <sub>2138,345</sub>

<sup>a</sup> Todos los parámetros iniciales tomaron el valor 5 y precisión 0,2. <sup>b</sup> Sin información previa. Tamaño de la cadena igual a 10.000.

Tabla 5.5: Media y desviación estándar de 50 simulaciones de  $AIC_{MC2E}$ . Sigma 0,1.

m	k	n	BMOM			$Bmom_{OPT}$	$Bayes_{2E}^a$	MCMC <sup>b</sup>
			Bondad ajuste	Precisión estimación	Tradicional			
10	20	100	-318,962 <sub>1980,392</sub>	-308,709 <sub>1987,418</sub>	-114,869 <sub>1981,466</sub>	<b>-2673,276</b> <sub>2377,238</sub>	860,668 <sub>1064,956</sub>	4494,185 <sub>638,266</sub>
10	40	100	-188,458 <sub>1826,964</sub>	-164,944 <sub>1845,007</sub>	88,455 <sub>1847,359</sub>	<b>-1123,007</b> <sub>2053,466</sub>	929,625 <sub>1021,578</sub>	4344,843 <sub>503,577</sub>
20	60	100	75,935 <sub>1183,877</sub>	58,759 <sub>1183,941</sub>	1079,081 <sub>1185,327</sub>	<b>-1079,661</b> <sub>2008,975</sub>	2139,387 <sub>1343,573</sub>	9912,858 <sub>248,055</sub>
10	20	400	-448,690 <sub>8411,143</sub>	-439,480 <sub>8415,159</sub>	-245,942 <sub>8411,416</sub>	<b>-12183,687</b> <sub>11212,626</sub>	1448,893 <sub>7302,126</sub>	22615,837 <sub>2380,486</sub>
10	40	400	-707,417 <sub>5559,027</sub>	-728,139 <sub>5561,745</sub>	-509,476 <sub>5553,990</sub>	<b>-8332,300</b> <sub>6753,422</sub>	334,995 <sub>7490,973</sub>	21890,892 <sub>1336,942</sub>
10	40	1000	-10316,546 <sub>14901,662</sub>	-10295,502 <sub>14900,521</sub>	-10069,827 <sub>14898,311</sub>	<b>-37959,671</b> <sub>18784,233</sub>	-14900,925 <sub>11966,366</sub>	62977,169 <sub>3595,350</sub>

<sup>a</sup> Todos los parámetros iniciales tomaron el valor 5 y precisión 0,2. <sup>b</sup> Sin información previa. Tamaño de la cadena igual a 10.000.

Tabla 5.6: Media y desviación estándar de 50 simulaciones de  $H(e)$ . Sigma 0,1.

m	k	n	MC2E	BMOM			$Bmom_{OPT}$	$Bayes_{2E}^a$	MCMC <sup>b</sup>
				Bondad ajuste	Precisión estimación	Tradicional			
10	20	100	58,840 <sub>0,758</sub>	59,250 <sub>0,706</sub>	59,247 <sub>0,716</sub>	59,291 <sub>0,736</sub>	<b>58,841</b> <sub>0,749</sub>	59,445 <sub>1,050</sub>	60,169 <sub>0,956</sub>
10	40	100	58,927 <sub>0,733</sub>	59,177 <sub>0,565</sub>	59,181 <sub>0,561</sub>	59,241 <sub>0,602</sub>	<b>58,899</b> <sub>0,811</sub>	59,423 <sub>0,676</sub>	59,583 <sub>0,801</sub>
20	60	100	59,565 <sub>0,465</sub>	59,611 <sub>0,518</sub>	59,624 <sub>0,525</sub>	59,612 <sub>0,540</sub>	<b>59,553</b> <sub>0,508</sub>	59,536 <sub>0,507</sub>	59,661 <sub>0,549</sub>
10	20	400	264,779 <sub>1,400</sub>	264,966 <sub>1,418</sub>	264,972 <sub>1,423</sub>	264,988 <sub>1,421</sub>	<b>264,746</b> <sub>1,410</sub>	266,956 <sub>5,807</sub>	273,052 <sub>2,240</sub>
10	40	400	<b>267,673</b> <sub>1,364</sub>	267,752 <sub>1,318</sub>	267,752 <sub>1,318</sub>	267,765 <sub>1,317</sub>	267,727 <sub>1,405</sub>	268,744 <sub>3,959</sub>	270,982 <sub>1,795</sub>
10	40	1000	685,429 <sub>2,393</sub>	685,812 <sub>2,308</sub>	685,812 <sub>2,307</sub>	685,821 <sub>2,307</sub>	<b>685,401</b> <sub>2,423</sub>	687,027 <sub>9,618</sub>	712,114 <sub>2,725</sub>

<sup>a</sup> Todos los parámetros iniciales tomaron el valor 5 y precisión 0,2. <sup>b</sup> Sin información previa. Tamaño de la cadena igual a 10.000.

Tabla 5.7: Media y desviación estándar de 50 simulaciones de  $H_2(e)$ . Sigma 0,1.

m	k	n	MC2E	BMOM			$Bmom_{OPT}$	$Bayes_{2E}^a$	MCMC <sup>b</sup>
				Bondad ajuste	Precisión estimación	Tradicional			
10	20	100	4,074 <sub>0,013</sub>	4,081 <sub>0,012</sub>	4,081 <sub>0,012</sub>	4,082 <sub>0,013</sub>	<b>4,074</b> <sub>0,013</sub>	4,084 <sub>0,018</sub>	4,096 <sub>0,016</sub>
10	40	100	4,076 <sub>0,012</sub>	4,080 <sub>0,010</sub>	4,080 <sub>0,010</sub>	4,081 <sub>0,010</sub>	<b>4,075</b> <sub>0,014</sub>	4,084 <sub>0,010</sub>	4,087 <sub>0,014</sub>
20	60	100	4,086 <sub>0,008</sub>	4,087 <sub>0,009</sub>	4,087 <sub>0,009</sub>	4,087 <sub>0,009</sub>	<b>4,086</b> <sub>0,009</sub>	4,086 <sub>0,009</sub>	4,088 <sub>0,009</sub>
10	20	400	5,579 <sub>0,005</sub>	5,579 <sub>0,005</sub>	5,579 <sub>0,005</sub>	5,580 <sub>0,005</sub>	<b>5,579</b> <sub>0,005</sub>	5,587 <sub>0,021</sub>	5,609 <sub>0,008</sub>
10	40	400	5,590 <sub>0,005</sub>	5,590 <sub>0,005</sub>	5,590 <sub>0,005</sub>	5,590 <sub>0,005</sub>	<b>5,590</b> <sub>0,005</sub>	5,593 <sub>0,014</sub>	5,602 <sub>0,007</sub>
10	40	1000	6,530 <sub>0,003</sub>	6,531 <sub>0,003</sub>	6,531 <sub>0,003</sub>	6,531 <sub>0,003</sub>	<b>6,530</b> <sub>0,003</sub>	6,532 <sub>0,013</sub>	6,568 <sub>0,003</sub>

<sup>a</sup> Todos los parámetros iniciales tomaron el valor 5 y precisión 0,2. <sup>b</sup> Sin información previa. Tamaño de la cadena igual a 10.000.

Tabla 5.8: Media y desviación estándar de 50 simulaciones de  $M_{1/2}(e)$ . Sigma 0,1.

m	k	n	MC2E	BMOM			$Bmom_{OPT}$	$Bayes_{2E}^a$	MCMC <sup>b</sup>
				Bondad ajuste	Precisión estimación	Tradicional			
10	20	100	84,461 <sub>0,013</sub>	84,764 <sub>0,012</sub>	84,760 <sub>0,012</sub>	84,793 <sub>0,013</sub>	<b>84,434</b> <sub>0,013</sub>	84,841 <sub>0,018</sub>	85,309 <sub>0,006</sub>
10	40	100	84,521 <sub>0,580</sub>	84,660 <sub>0,464</sub>	84,662 <sub>0,465</sub>	84,711 <sub>0,504</sub>	<b>84,518</b> <sub>0,639</sub>	84,854 <sub>0,546</sub>	84,902 <sub>0,617</sub>
20	60	100	84,935 <sub>0,378</sub>	84,954 <sub>0,401</sub>	84,972 <sub>0,422</sub>	84,967 <sub>0,412</sub>	84,958 <sub>0,429</sub>	<b>84,932</b> <sub>0,423</sub>	85,015 <sub>0,414</sub>
10	20	400	336,549 <sub>1,088</sub>	336,602 <sub>1,120</sub>	336,608 <sub>1,117</sub>	336,624 <sub>1,116</sub>	<b>336,528</b> <sub>1,084</sub>	337,845 <sub>3,630</sub>	341,229 <sub>1,493</sub>
10	40	400	<b>338,121</b> <sub>0,920</sub>	338,154 <sub>0,928</sub>	338,153 <sub>0,930</sub>	338,170 <sub>0,929</sub>	338,187 <sub>0,940</sub>	338,768 <sub>2,527</sub>	339,958 <sub>1,269</sub>
10	40	1000	834,826 <sub>1,754</sub>	835,014 <sub>1,632</sub>	835,014 <sub>1,631</sub>	835,022 <sub>1,633</sub>	<b>834,813</b> <sub>1,773</sub>	835,703 <sub>5,995</sub>	849,850 <sub>1,776</sub>

<sup>a</sup> Todos los parámetros iniciales tomaron el valor 5 y precisión 0,2. <sup>b</sup> Sin información previa. Tamaño de la cadena igual a 10.000.

Tabla 5.9: Media y desviación estándar de 50 simulaciones del Tiempo de ejecución. Sigma 0,1.

m	k	n	MC2E	BMOM			$Bmom_{OPT}$	$Bayes_{2E}^a$	MCMC <sup>b</sup>
				Bondad ajuste	Precisión estimación	Tradicional			
10	20	100	0,073 <sub>0,145</sub>	0,732 <sub>0,246</sub>	0,732 <sub>0,246</sub>	0,732 <sub>0,246</sub>	258,227 <sub>98,562</sub>	<b>0,056</b> <sub>0,017</sub>	294,453 <sub>17,949</sub>
10	40	100	0,143 <sub>0,047</sub>	1,022 <sub>0,395</sub>	1,022 <sub>0,395</sub>	1,022 <sub>0,395</sub>	274,088 <sub>345,929</sub>	<b>0,140</b> <sub>0,050</sub>	499,554 <sub>68,116</sub>
20	60	100	<b>0,314</b> <sub>0,047</sub>	2,495 <sub>0,395</sub>	2,495 <sub>0,395</sub>	2,495 <sub>0,395</sub>	748,504 <sub>345,929</sub>	0,407 <sub>0,050</sub>	1435,791 <sub>68,116</sub>
10	20	400	0,125 <sub>0,031</sub>	4,265 <sub>0,864</sub>	4,265 <sub>0,864</sub>	4,265 <sub>0,864</sub>	2586,186 <sub>1005,068</sub>	<b>0,109</b> <sub>0,030</sub>	328,571 <sub>18,174</sub>
10	40	400	0,235 <sub>0,037</sub>	4,533 <sub>0,765</sub>	4,533 <sub>0,765</sub>	4,533 <sub>0,765</sub>	2281,255 <sub>804,186</sub>	<b>0,214</b> <sub>0,032</sub>	507,791 <sub>38,694</sub>
10	40	1000	0,426 <sub>0,066</sub>	21,385 <sub>1,595</sub>	21,385 <sub>1,595</sub>	21,385 <sub>1,595</sub>	14534,080 <sub>21689,539</sub>	<b>0,376</b> <sub>0,107</sub>	524,904 <sub>26,564</sub>

<sup>a</sup> Todos los parámetros iniciales tomaron el valor 5 y precisión 0,2. <sup>b</sup> Sin información previa. Tamaño de la cadena igual a 10.000.

## 5.4. Conclusiones

En este capítulo se ha desarrollado un nuevo método de estimación de MES, el método  $Bmom_{OPT}$ , consistente en la optimización de unos parámetros de tal manera que minimiza el AIC basado en los errores entre los coeficientes reales del modelo y los estimados. Este método, experimentando bajo ciertas condiciones de variabilidad, ofrece estimaciones con menor norma euclídea entre los parámetros del modelo y sus estimaciones comparado con el resto de métodos estudiados.

Uno de los criterios más usados para la elección de un modelo estimado, el AIC, ha sido obtenido en las simulaciones como criterio de selección de método de estimación. Los resultados han mostrado que AIC no lleva a la elección del método con menor error de

estimación de los parámetros. Esto ha motivado al desarrollo de un nuevo criterio para la selección del método de estimación con menor error. Concretamente, se han desarrollado tres entropías para su aplicación al modelo de ecuaciones simultáneas, observando tras el experimento que, a pesar de que todas ellas ofrecen resultados muy parecidos, la entropía  $H_2(e)$  selecciona el método de estimación con menor diferencia entre los parámetros y sus estimaciones.

Resaltar que el uso de la entropía como criterio de selección lleva a la elección del método que minimiza la norma del error en un mayor porcentaje que el AIC. Finalmente, notar que minimizar el error no conlleva a la elección del método con menor tiempo de ejecución, por lo que hay que valorar en función de la naturaleza del problema si prevalece el tiempo de cómputo o el error cometido.

---

# CAPÍTULO 6

---

## **Método de Estimación de un Modelo de Ecuaciones Simultáneas según Variabilidad de los Datos**

---

En este capítulo se lleva a cabo el estudio del efecto de la variabilidad de las variables endógenas sobre el método de estimación de un MES con menor error. Algunos MES estudiados en esta tesis tienen grandes dimensiones, tanto en número de variables endógenas como exógenas y muestra, por lo que las combinaciones del número de variables y el tamaño de la muestra produce un alto número de modelos distintos, hecho que ha motivado a controlar únicamente la variabilidad de las variables endógenas.

### **6.1. Introducción**

El primer problema al que hacer frente en una investigación es modelizar un conjunto de datos según un MES para lo cual, un análisis preliminar que describe el conjunto de variables es fundamental.

Como ya se describió en el capítulo 2, en los MES hay dos tipos de variables implícitas en la definición del modelo: por un lado las variables endógenas, variables que explican a la vez que son explicadas, y, por otro lado, las variables exógenas, variables puramente explicativas. Las variables endógenas son doblemente interesantes por el papel que juegan en un MES, propiedades que ha motivado a controlar su variabilidad a la hora de estudiar qué método ofrece mejores estimaciones según el escenario analizado. La variabilidad de las variables endógenas es determinante en el ajuste de un modelo, puesto que no se obtienen los mismos resultados al trabajar con datos muy variables que con datos con poca variabilidad. Por lo tanto, como consecuencia de esta propiedad, el modelo estimado se ajustará en menor o mayor medida a los datos en función de esa variabilidad. Al mismo tiempo, es deseable que el conjunto de variables exógenas que intervienen en el MES tenga un alto grado de variabilidad, pues es esta característica la que define su capacidad explicativa y la varianza de los estimadores.

Por todo ello, el objetivo de este capítulo es determinar el método de estimación de entre los estudiados en el capítulo 5 que ofrece menor error de estimación en función de la variabilidad de las variables endógenas y de la dimensión del modelo.

## 6.2. Estudio experimental

Para llevar a cabo este estudio se han simulado MES con diferentes variabilidades en lo que respecta a las variables endógenas, con el objetivo de encontrar el mejor método de estimación en base a esa variabilidad y a las dimensiones del modelo.

Para este experimento se ha considerado un gran rango de variabilidades en la generación de los MES descrita en la sección 5.1 y, una vez estimados, se han seleccionado los resultados más representativos según la desviación típica de las variables endógenas. Estos resultados se muestran en las tablas 6.1 - 6.4, las cuales recogen los valores medios de 30 iteraciones de un MES de las siguientes medidas: norma euclídea  $D_{\delta, \hat{\delta}}$ , AIC, entropía  $H_2(e)$  y el tiempo de ejecución. El resto de entropías, (ver 5.7 y 5.11), no han sido incluidas por llevar a conclusiones idénticas a  $H_2(e)$ . El número de menor tamaño que acompaña a la media es la desviación típica. Se recogen los resultados para MES

---

en los que se han combinado 10 y 20 variables endógenas, 20, 40, 60 y 100 variables exógenas y tamaño muestral igual a 100, 400 y 1000. En cada MES el valor  $DT(Y)$  es la media de la desviación típica de las  $m$  variables endógenas de ese modelo.

Para facilitar la comprensión al lector, a continuación se describe la estructura de la tabla 6.1 y los resultados de los seis modelos presentados en ella: el primer bloque es el resultado de simular 30 MES con 10 variables endógenas, 40 exógenas, tamaño muestral igual a 100 y la media de la desviación típica de las 10 endógenas en las 30 simulaciones es igual a 1,971. Dentro de este primer bloque de la tabla, el valor mínimo de la media de la norma euclídea de las 30 simulaciones (23,306) se ha obtenido con el método  $Bmom_{OPT}$ . A este valor mínimo le siguen los valores obtenidos con  $MC2E$ ,  $Bayes_{2E}$ , los BMOM y por último, MCMC.

Respecto al AIC, el valor obtenido por el método  $Bmom_{OPT}$  se sitúa por detrás de  $MC2E$ , los BMOM y  $Bayes_{2E}$ , quedando solo por detrás de él, el obtenido con MCMC.

En cuanto a la entropía  $H_2(e)$ , el valor mínimo vuelve a alcanzarse con el método  $Bmom_{OPT}$ , seguido por  $MC2E$ ,  $Bayes_{2E}$ , los BMOM y MCMC.

Finalmente, el tiempo de ejecución es máximo con  $Bmom_{OPT}$ , presentando además una gran diferencia respecto al resto de métodos.

Analizando modelos con mayor número de variables y tamaño muestral, pero con  $DT(Y)$  próxima a la comentada anteriormente, los resultados presentan una estructura similar, siendo  $Bmom_{OPT}$  el método que, en valor medio, tiene menor norma euclídea y entropía y mayor tiempo de ejecución.

Al aumentar un 36,88% la variabilidad de las variables endógenas, tabla 6.2, las simulaciones y estimaciones muestran un patrón similar al comentado anteriormente: el valor mínimo en norma euclídea y entropía se alcanza con  $Bmom_{OPT}$ .

La tabla 6.3 recoge los resultados de modelos que han aumentado un 43,45% la variabilidad de las  $m$  variables endógenas respecto a los modelos de la tabla 6.2: se puede observar que los resultados son similares a los obtenidos en las tablas 6.1 y 6.2: norma euclídea y entropía mínimas con  $Bmom_{OPT}$ . Sin embargo, estos resultados difieren cuando  $DT(Y)$  aumenta un 180%, ver tabla 6.4, situación en la que no se observa que ningún de

---

los métodos sea mejor que el resto, ni en norma euclídea ni en entropía.

Tabla 6.1: Media y desviación típica de la norma, AIC, entropía y tiempo de ejecución de 30 simulaciones (I). Desviación media total de las variables endógenas 2,133.

Medida	MC2E	BMOM			$Bmom_{OPT}$	$Bayes2E$	MCMC
		Bondad ajuste	Precisión estimación	Tradicional			
		m = 10	k = 20	n = 100	DT(Y) = 1,971		
$D_{\delta,\hat{\delta}}$	29,570 <sub>7,292</sub>	43,233 <sub>6,553</sub>	43,509 <sub>6,567</sub>	44,653 <sub>6,511</sub>	<b>23,306</b> <sub>8,462</sub>	31,962 <sub>10,139</sub>	68,373 <sub>4,435</sub>
AIC	2511,737 <sub>567,795</sub>	2085,189 <sub>539,692</sub>	2080,392 <sub>538,473</sub>	<b>2056,552</b> <sub>535,776</sub>	2853,010 <sub>627,993</sub>	2684,088 <sub>571,362</sub>	4438,846 <sub>235,293</sub>
$H_2(e)$	4,056 <sub>0,014</sub>	4,068 <sub>0,013</sub>	4,068 <sub>0,013</sub>	4,070 <sub>0,013</sub>	<b>4,054</b> <sub>0,014</sub>	4,065 <sub>0,015</sub>	4,097 <sub>0,012</sub>
Tiempo	<b>0,043</b> <sub>0,013</sub>	0,603 <sub>0,168</sub>	0,603 <sub>0,168</sub>	0,603 <sub>0,168</sub>	316,834 <sub>148,244</sub>	0,047 <sub>0,010</sub>	288,370 <sub>24,462</sub>
		m = 10	k = 40	n = 100	DT(Y) = 2,188		
$D_{\delta,\hat{\delta}}$	38,819 <sub>6,609</sub>	58,029 <sub>7,433</sub>	58,216 <sub>7,441</sub>	60,449 <sub>7,329</sub>	<b>25,942</b> <sub>9,070</sub>	57,643 <sub>13,762</sub>	88,757 <sub>3,850</sub>
AIC	2536,259 <sub>374,421</sub>	2332,463 <sub>358,592</sub>	<b>2332,076</b> <sub>358,549</sub>	2332,531 <sub>356,661</sub>	2876,489 <sub>414,139</sub>	2765,889 <sub>495,545</sub>	4471,742 <sub>179,768</sub>
$H_2(e)$	4,063 <sub>0,010</sub>	4,077 <sub>0,011</sub>	4,077 <sub>0,011</sub>	4,079 <sub>0,011</sub>	<b>4,059</b> <sub>0,012</sub>	4,079 <sub>0,014</sub>	4,091 <sub>0,015</sub>
Tiempo	<b>0,130</b> <sub>0,049</sub>	1,024 <sub>0,279</sub>	1,024 <sub>0,279</sub>	1,024 <sub>0,279</sub>	275,434 <sub>130,434</sub>	0,139 <sub>0,054</sub>	491,668 <sub>36,030</sub>
		m = 20	k = 60	n = 100	DT(Y) = 2,360		
$D_{\delta,\hat{\delta}}$	110,935 <sub>7,739</sub>	137,417 <sub>5,445</sub>	137,891 <sub>5,448</sub>	145,608 <sub>4,630</sub>	<b>89,024</b> <sub>10,135</sub>	145,483 <sub>7,610</sub>	160,597 <sub>3,866</sub>
AIC	4647,691 <sub>509,027</sub>	<b>4417,764</b> <sub>510,143</sub>	4418,854 <sub>509,616</sub>	4516,954 <sub>506,291</sub>	5323,162 <sub>556,887</sub>	4991,909 <sub>862,818</sub>	9908,837 <sub>240,800</sub>
$H_2(e)$	4,083 <sub>0,011</sub>	4,084 <sub>0,008</sub>	4,085 <sub>0,008</sub>	4,085 <sub>0,009</sub>	<b>4,082</b> <sub>0,010</sub>	4,084 <sub>0,010</sub>	4,086 <sub>0,009</sub>
Tiempo	<b>0,311</b> <sub>0,116</sub>	2,436 <sub>0,318</sub>	2,436 <sub>0,318</sub>	2,436 <sub>0,318</sub>	743,402 <sub>333,421</sub>	0,405 <sub>0,099</sub>	1426,786 <sub>60,510</sub>
		m = 10	k = 20	n = 400	DT(Y) = 1,549		
$D_{\delta,\hat{\delta}}$	14,327 <sub>6,779</sub>	26,212 <sub>6,177</sub>	26,256 <sub>6,194</sub>	26,504 <sub>6,187</sub>	<b>11,006</b> <sub>14,884</sub>	17,047 <sub>12,569</sub>	66,636 <sub>4,192</sub>
AIC	17171,472 <sub>2861,319</sub>	15636,993 <sub>2436,438</sub>	15632,856 <sub>2436,027</sub>	<b>15605,275</b> <sub>2428,614</sub>	17740,722 <sub>3037,985</sub>	17514,730 <sub>2831,427</sub>	22414,282 <sub>1035,850</sub>
$H_2(e)$	5,560 <sub>0,007</sub>	5,562 <sub>0,007</sub>	5,562 <sub>0,007</sub>	5,562 <sub>0,007</sub>	<b>5,559</b> <sub>0,007</sub>	5,567 <sub>0,018</sub>	5,613 <sub>0,012</sub>
Tiempo	0,115 <sub>0,038</sub>	3,942 <sub>0,671</sub>	3,942 <sub>0,671</sub>	3,942 <sub>0,671</sub>	2360,125 <sub>861,793</sub>	<b>0,094</b> <sub>0,025</sub>	323,575 <sub>14,397</sub>
		m = 10	k = 40	n = 400	DT(Y) = 2,131		
$D_{\delta,\hat{\delta}}$	19,486 <sub>5,255</sub>	37,365 <sub>6,768</sub>	37,391 <sub>6,771</sub>	37,784 <sub>6,782</sub>	<b>11,146</b> <sub>3,504</sub>	24,802 <sub>17,350</sub>	89,056 <sub>4,633</sub>
AIC	16538,739 <sub>1844,884</sub>	15279,305 <sub>1598,832</sub>	15277,803 <sub>1598,426</sub>	<b>15255,394</b> <sub>1593,465</sub>	17443,120 <sub>1991,016</sub>	16814,836 <sub>2095,821</sub>	21809,350 <sub>705,884</sub>
$H_2(e)$	5,567 <sub>0,005</sub>	5,569 <sub>0,007</sub>	5,569 <sub>0,007</sub>	5,570 <sub>0,007</sub>	<b>5,567</b> <sub>0,007</sub>	5,571 <sub>0,010</sub>	5,601 <sub>0,006</sub>
Tiempo	0,232 <sub>0,035</sub>	4,437 <sub>0,769</sub>	4,437 <sub>0,769</sub>	4,437 <sub>0,769</sub>	2448,428 <sub>919,992</sub>	<b>0,219</b> <sub>0,035</sub>	503,499 <sub>41,883</sub>
		m = 20	k = 100	n = 1000	DT(Y) = 2,596		
$D_{\delta,\hat{\delta}}$	42,808 <sub>6,468</sub>	88,271 <sub>8,806</sub>	88,262 <sub>8,800</sub>	88,999 <sub>8,814</sub>	<b>19,849</b> <sub>3,787</sub>	47,486 <sub>7,043</sub>	197,665 <sub>3,093</sub>
AIC	99642,703 <sub>4799,050</sub>	92828,617 <sub>4459,255</sub>	92829,991 <sub>4458,503</sub>	<b>92738,984</b> <sub>4454,168</sub>	104887,960 <sub>5183,736</sub>	99824,959 <sub>4824,378</sub>	126819,415 <sub>1197,312</sub>
$H_2(e)$	6,558 <sub>0,004</sub>	6,559 <sub>0,003</sub>	6,559 <sub>0,003</sub>	6,559 <sub>0,003</sub>	<b>6,558</b> <sub>0,003</sub>	6,559 <sub>0,003</sub>	6,564 <sub>0,003</sub>
Tiempo	1,555 <sub>0,088</sub>	50,089 <sub>2,990</sub>	50,089 <sub>2,990</sub>	50,089 <sub>2,990</sub>	22795,800 <sub>8931,299</sub>	<b>1,397</b> <sub>0,220</sub>	2650,516 <sub>85,397</sub>

Tabla 6.2: Media y desviación típica de la norma, AIC, entropía y tiempo de ejecución de 30 simulaciones (II). Desviación media total de las variables endógenas 2,921.

Medida	MC2E	BMOM			<i>Bmom<sub>OPT</sub></i>	<i>Bayes<sub>2E</sub></i>	MCMC
		Bondad ajuste	Precisión estimación	Tradicional			
		m = 10	k = 20	n = 100	DT(Y) = 2,404		
$D_{\delta, \hat{\delta}}$	30,851 <sub>6,821</sub>	44,885 <sub>6,331</sub>	45,080 <sub>6,358</sub>	46,131 <sub>6,369</sub>	<b>25,407</b> <sub>6,638</sub>	33,150 <sub>8,399</sub>	67,589 <sub>4,840</sub>
AIC	2459,170 <sub>738,623</sub>	2009,605 <sub>688,520</sub>	2006,469 <sub>688,226</sub>	<b>1986,016</b> <sub>685,967</sub>	2741,467 <sub>725,342</sub>	2599,981 <sub>678,672</sub>	4296,245 <sub>351,974</sub>
$H_2(e)$	4,074 <sub>0,013</sub>	4,082 <sub>0,015</sub>	4,082 <sub>0,015</sub>	4,083 <sub>0,016</sub>	<b>4,073</b> <sub>0,014</sub>	4,079 <sub>0,016</sub>	4,096 <sub>0,010</sub>
Tiempo	0,050 <sub>0,019</sub>	0,673 <sub>0,231</sub>	0,673 <sub>0,231</sub>	0,673 <sub>0,231</sub>	271,608 <sub>107,645</sub>	<b>0,050</b> <sub>0,014</sub>	291,182 <sub>25,402</sub>
		m = 10	k = 40	n = 100	DT(Y) = 2,823		
$D_{\delta, \hat{\delta}}$	40,395 <sub>8,487</sub>	59,120 <sub>7,494</sub>	58,865 <sub>7,481</sub>	60,922 <sub>7,322</sub>	<b>24,403</b> <sub>8,504</sub>	56,285 <sub>12,434</sub>	89,702 <sub>4,969</sub>
AIC	2009,825 <sub>537,522</sub>	1848,683 <sub>519,646</sub>	<b>1848,229</b> <sub>519,631</sub>	1853,089 <sub>518,847</sub>	2376,169 <sub>557,670</sub>	2297,605 <sub>576,486</sub>	4382,030 <sub>160,610</sub>
$H_2(e)$	4,072 <sub>0,014</sub>	4,080 <sub>0,013</sub>	4,080 <sub>0,013</sub>	4,082 <sub>0,013</sub>	<b>4,072</b> <sub>0,013</sub>	4,086 <sub>0,016</sub>	4,090 <sub>0,011</sub>
Tiempo	0,144 <sub>0,054</sub>	1,035 <sub>0,445</sub>	1,035 <sub>0,445</sub>	1,035 <sub>0,445</sub>	284,650 <sub>345,150</sub>	<b>0,137</b> <sub>0,089</sub>	499,047 <sub>65,318</sub>
		m = 20	k = 60	n = 100	DT(Y) = 3,059		
$D_{\delta, \hat{\delta}}$	114,691 <sub>6,804</sub>	139,646 <sub>4,288</sub>	140,166 <sub>4,341</sub>	147,319 <sub>3,616</sub>	<b>92,084</b> <sub>12,045</sub>	0,000 <sub>0,000</sub>	163,504 <sub>4,960</sub>
AIC	3823,781 <sub>1075,918</sub>	<b>3580,740</b> <sub>1074,163</sub>	3582,173 <sub>1074,093</sub>	3668,646 <sub>1074,236</sub>	4583,164 <sub>1173,347</sub>	0,000 <sub>0,000</sub>	9802,854 <sub>255,752</sub>
$H_2(e)$	4,083 <sub>0,007</sub>	4,085 <sub>0,008</sub>	4,085 <sub>0,008</sub>	4,086 <sub>0,009</sub>	<b>4,079</b> <sub>0,008</sub>	<i>nan<sub>nan</sub></i>	4,085 <sub>0,008</sub>
Tiempo	<b>0,313</b> <sub>0,054</sub>	2,577 <sub>0,445</sub>	2,577 <sub>0,445</sub>	2,577 <sub>0,445</sub>	745,608 <sub>345,150</sub>	0,426 <sub>0,089</sub>	1418,851 <sub>65,318</sub>
		m = 10	k = 20	n = 400	DT(Y) = 2,353		
$D_{\delta, \hat{\delta}}$	14,425 <sub>4,582</sub>	27,162 <sub>6,033</sub>	27,247 <sub>6,049</sub>	27,490 <sub>6,071</sub>	<b>9,583</b> <sub>3,449</sub>	20,753 <sub>14,505</sub>	71,746 <sub>13,390</sub>
AIC	14831,122 <sub>3852,753</sub>	13376,760 <sub>3704,566</sub>	13368,223 <sub>3703,172</sub>	<b>13342,259</b> <sub>3699,937</sub>	15463,140 <sub>3868,162</sub>	15323,514 <sub>4123,471</sub>	22063,045 <sub>1566,243</sub>
$H_2(e)$	5,586 <sub>0,007</sub>	5,588 <sub>0,006</sub>	5,588 <sub>0,006</sub>	5,588 <sub>0,006</sub>	<b>5,586</b> <sub>0,007</sub>	5,595 <sub>0,020</sub>	5,609 <sub>0,009</sub>
Tiempo	0,125 <sub>0,039</sub>	4,008 <sub>0,895</sub>	4,008 <sub>0,895</sub>	4,008 <sub>0,895</sub>	2679,831 <sub>1160,780</sub>	<b>0,104</b> <sub>0,031</sub>	327,002 <sub>17,686</sub>
		m = 10	k = 40	n = 400	DT(Y) = 3,080		
$D_{\delta, \hat{\delta}}$	14,312 <sub>4,490</sub>	28,909 <sub>6,813</sub>	28,979 <sub>6,823</sub>	29,335 <sub>6,872</sub>	<b>8,286</b> <sub>4,033</sub>	21,751 <sub>15,277</sub>	90,063 <sub>5,761</sub>
AIC	13016,811 <sub>3505,913</sub>	12219,061 <sub>3242,267</sub>	12216,385 <sub>3241,410</sub>	<b>12201,891</b> <sub>3235,858</sub>	13608,205 <sub>3613,720</sub>	13602,937 <sub>3604,000</sub>	21382,121 <sub>708,481</sub>
$H_2(e)$	5,595 <sub>0,007</sub>	5,594 <sub>0,007</sub>	5,594 <sub>0,007</sub>	<b>5,594</b> <sub>0,007</sub>	5,595 <sub>0,007</sub>	5,598 <sub>0,012</sub>	5,601 <sub>0,006</sub>
Tiempo	0,227 <sub>0,037</sub>	4,690 <sub>0,812</sub>	4,690 <sub>0,812</sub>	4,690 <sub>0,812</sub>	2352,238 <sub>1009,621</sub>	<b>0,218</b> <sub>0,045</sub>	493,534 <sub>34,199</sub>
		m = 20	k = 100	n = 1000	DT(Y) = 3,808		
$D_{\delta, \hat{\delta}}$	43,484 <sub>5,740</sub>	88,873 <sub>8,501</sub>	88,820 <sub>8,504</sub>	89,549 <sub>8,522</sub>	<b>19,893</b> <sub>4,297</sub>	0,000 <sub>0,000</sub>	198,504 <sub>4,178</sub>
AIC	85343,667 <sub>8236,814</sub>	<b>78921,898</b> <sub>7831,563</sub>	78927,862 <sub>7831,632</sub>	78843,776 <sub>7827,253</sub>	90572,563 <sub>8783,688</sub>	0,000 <sub>0,000</sub>	125125,805 <sub>1442,854</sub>
$H_2(e)$	6,559 <sub>0,002</sub>	6,559 <sub>0,002</sub>	6,559 <sub>0,002</sub>	6,559 <sub>0,002</sub>	<b>6,559</b> <sub>0,002</sub>	<i>nan<sub>nan</sub></i>	6,564 <sub>0,002</sub>
Tiempo	1,597 <sub>0,081</sub>	51,019 <sub>3,166</sub>	51,019 <sub>3,166</sub>	51,019 <sub>3,166</sub>	21575,520 <sub>8650,111</sub>	<b>1,353</b> <sub>0,118</sub>	2640,258 <sub>87,250</sub>

Tabla 6.3: Media y desviación típica de la norma, AIC, entropía y tiempo de ejecución de 30 simulaciones (III). Desviación media total de las variables endógenas 4,191.

Medida	MC2E	BMOM			$Bmom_{OPT}$	$Bayes_{2E}$	MCMC
		Bondad ajuste	Precisión estimación	Tradicional			
		m = 10	k = 20	n = 100	DT(Y) = 3,972		
$D_{\delta,\delta}$	64,243 <sub>4,592</sub>	65,812 <sub>3,864</sub>	65,711 <sub>3,880</sub>	65,826 <sub>3,868</sub>	<b>62,303</b> <sub>4,772</sub>	63,981 <sub>5,653</sub>	84,628 <sub>43,130</sub>
AIC	3389,415 <sub>905,619</sub>	<b>3245,597</b> <sub>914,478</sub>	3230,032 <sub>916,655</sub>	3233,181 <sub>916,539</sub>	3894,438 <sub>938,251</sub>	3479,163 <sub>940,086</sub>	4904,369 <sub>430,502</sub>
$H_2(e)$	4,058 <sub>0,010</sub>	4,069 <sub>0,011</sub>	4,071 <sub>0,014</sub>	4,071 <sub>0,014</sub>	<b>4,054</b> <sub>0,014</sub>	4,063 <sub>0,017</sub>	4,083 <sub>0,016</sub>
Tiempo	<b>0,047</b> <sub>0,011</sub>	0,712 <sub>0,238</sub>	0,712 <sub>0,238</sub>	0,712 <sub>0,238</sub>	327,569	0,054 <sub>124,536</sub>	292,828 <sub>16,768</sub>
		m = 10	k = 40	n = 100	DT(Y) = 4,536		
$D_{\delta,\delta}$	83,503 <sub>5,985</sub>	85,800 <sub>4,952</sub>	85,333 <sub>5,012</sub>	85,525 <sub>4,978</sub>	<b>77,058</b> <sub>8,590</sub>	82,921 <sub>5,715</sub>	94,550 <sub>11,234</sub>
AIC	3335,222 <sub>756,664</sub>	3334,452 <sub>766,319</sub>	<b>3323,311</b> <sub>765,301</sub>	3332,931 <sub>766,050</sub>	4072,035 <sub>916,311</sub>	3368,906 <sub>763,631</sub>	4639,055 <sub>261,408</sub>
$H_2(e)$	4,065 <sub>0,017</sub>	4,072 <sub>0,015</sub>	4,071 <sub>0,016</sub>	4,072 <sub>0,016</sub>	<b>4,059</b> <sub>0,016</sub>	4,065 <sub>0,016</sub>	4,085 <sub>0,014</sub>
Tiempo	<b>0,139</b> <sub>0,052</sub>	1,050 <sub>0,430</sub>	1,050 <sub>0,430</sub>	1,050 <sub>0,430</sub>	331,036 <sub>302,939</sub>	0,142 <sub>0,060</sub>	483,945 <sub>70,635</sub>
		m = 20	k = 60	n = 100	DT(Y) = 4,418		
$D_{\delta,\delta}$	157,976 <sub>4,312</sub>	158,340 <sub>4,211</sub>	159,119 <sub>4,256</sub>	159,527 <sub>4,257</sub>	<b>156,305</b> <sub>4,729</sub>	0,000 <sub>0,000</sub>	169,204 <sub>12,479</sub>
AIC	<b>5923,506</b> <sub>61488,941</sub>	5974,900 <sub>1496,556</sub>	5955,546 <sub>1490,238</sub>	6095,253 <sub>1496,796</sub>	6186,213 <sub>1570,472</sub>	0,000 <sub>0,000</sub>	10176,159 <sub>299,723</sub>
$H_2(e)$	4,084 <sub>0,008</sub>	4,084 <sub>0,008</sub>	<b>4,082</b> <sub>0,007</sub>	4,084 <sub>0,007</sub>	4,084 <sub>0,009</sub>	<i>nan</i> <sub><i>nan</i></sub>	4,086 <sub>0,008</sub>
Tiempo	<b>0,319</b> <sub>0,052</sub>	2,505 <sub>0,430</sub>	2,505 <sub>0,430</sub>	2,505 <sub>0,430</sub>	439,737 <sub>302,939</sub>	0,415 <sub>0,060</sub>	1390,733 <sub>70,635</sub>
		m = 10	k = 20	n = 400	DT(Y) = 3,166		
$D_{\delta,\delta}$	57,727 <sub>5,872</sub>	61,223 <sub>4,344</sub>	61,396 <sub>4,336</sub>	61,458 <sub>4,326</sub>	<b>52,494</b> <sub>8,571</sub>	55,637 <sub>5,684</sub>	75,588 <sub>18,005</sub>
AIC	20596,386 <sub>3503,758</sub>	18472,471 <sub>3488,608</sub>	18466,514 <sub>3496,757</sub>	<b>18455,702</b> <sub>3498,762</sub>	23873,170 <sub>4223,949</sub>	20621,834 <sub>3584,533</sub>	24427,720 <sub>1452,321</sub>
$H_2(e)$	5,557 <sub>0,007</sub>	5,563 <sub>0,006</sub>	5,563 <sub>0,006</sub>	5,563 <sub>0,006</sub>	<b>5,553</b> <sub>0,006</sub>	5,558 <sub>0,010</sub>	5,590 <sub>0,012</sub>
Tiempo	0,114 <sub>0,023</sub>	4,042 <sub>0,720</sub>	4,042 <sub>0,720</sub>	4,042 <sub>0,720</sub>	3203,778 <sub>1009,648</sub>	<b>0,101</b> <sub>0,029</sub>	316,958 <sub>20,624</sub>
		m = 10	k = 40	n = 400	DT(Y) = 4,203		
$D_{\delta,\delta}$	75,404 <sub>7,037</sub>	81,800 <sub>5,701</sub>	81,998 <sub>5,707</sub>	82,094 <sub>5,678</sub>	<b>64,009</b> <sub>10,438</sub>	76,075 <sub>7,571</sub>	94,743 <sub>13,298</sub>
AIC	18755,204 <sub>3414,020</sub>	17417,064 <sub>3297,987</sub>	17400,546 <sub>3296,650</sub>	<b>17391,812</b> <sub>3296,950</sub>	22630,862 <sub>4240,619</sub>	18940,266 <sub>3388,913</sub>	22764,141 <sub>1110,199</sub>
$H_2(e)$	5,577 <sub>0,005</sub>	5,581 <sub>0,005</sub>	5,581 <sub>0,005</sub>	5,582 <sub>0,005</sub>	<b>5,575</b> <sub>0,006</sub>	5,579 <sub>0,007</sub>	5,596 <sub>0,008</sub>
Tiempo	0,242 <sub>0,045</sub>	4,538 <sub>0,863</sub>	4,538 <sub>0,863</sub>	4,538 <sub>0,863</sub>	3968,580 <sub>1490,813</sub>	<b>0,227</b> <sub>0,041</sub>	514,619 <sub>42,887</sub>
		m = 20	k = 100	n = 1000	DT(Y) = 4,848		
$D_{\delta,\delta}$	180,287 <sub>6,136</sub>	190,691 <sub>4,463</sub>	190,646 <sub>4,466</sub>	190,735 <sub>4,451</sub>	<b>159,196</b> <sub>11,157</sub>	180,287 <sub>6,852</sub>	202,415 <sub>6,844</sub>
AIC	97999,498 <sub>10710,142</sub>	92340,942 <sub>10683,055</sub>	92345,481 <sub>10683,908</sub>	<b>92318,846</b> <sub>10682,283</sub>	118914,540 <sub>13403,792</sub>	98592,443 <sub>11085,038</sub>	129176,009 <sub>1934,991</sub>
$H_2(e)$	6,549 <sub>0,003</sub>	6,553 <sub>0,003</sub>	6,553 <sub>0,003</sub>	6,553 <sub>0,003</sub>	<b>6,546</b> <sub>0,003</sub>	6,549 <sub>0,004</sub>	6,563 <sub>0,003</sub>
Tiempo	<b>1,632</b> <sub>0,113</sub>	51,529 <sub>3,407</sub>	51,529 <sub>3,407</sub>	51,529 <sub>3,407</sub>	34335,860 <sub>10226,573</sub>	1,643 <sub>0,913</sub>	2652,106 <sub>77,169</sub>

Tabla 6.4: Media y desviación típica de la norma, AIC, entropía y tiempo de ejecución de 30 simulaciones (IV). Desviación media total de las variables endógenas 11,840.

Medida	MC2E	BMOM			<i>Bmom<sub>OPT</sub></i>	<i>Bayes<sub>2E</sub></i>	MCMC
		Bondad ajuste	Precisión estimación	Tradicional			
		m = 10	k = 20	n = 100			
					DT(Y) = 12,511		
$D_{\delta, \hat{\delta}}$	83,084 <sub>12,748</sub>	<b>76,477</b> <sub>6,039</sub>	76,643 <sub>6,184</sub>	76,699 <sub>6,279</sub>	92,806 <sub>22,442</sub>	82,445 <sub>8,761</sub>	91,919 <sub>14,065</sub>
AIC	4489,501 <sub>202,329</sub>	4533,026 <sub>177,548</sub>	<b>4480,348</b> <sub>172,137</sub>	4483,763 <sub>170,444</sub>	4859,169 <sub>689,871</sub>	4526,854 <sub>224,714</sub>	5003,599 <sub>140,176</sub>
$H_2(e)$	4,085 <sub>0,013</sub>	<b>4,083</b> <sub>0,012</sub>	4,080 <sub>0,015</sub>	4,080 <sub>0,015</sub>	4,087 <sub>0,014</sub>	4,085 <sub>0,014</sub>	4,086 <sub>0,012</sub>
Tiempo	<b>0,046</b> <sub>0,013</sub>	0,736 <sub>0,217</sub>	0,736 <sub>0,217</sub>	0,736 <sub>0,217</sub>	354,003 <sub>276,866</sub>	0,051 <sub>0,014</sub>	290,513 <sub>22,513</sub>
					DT(Y) = 10,536		
$D_{\delta, \hat{\delta}}$	100,856 <sub>8,532</sub>	<b>99,273</b> <sub>9,204</sub>	99,587 <sub>9,154</sub>	99,698 <sub>9,218</sub>	107,580 <sub>22,358</sub>	100,254 <sub>8,258</sub>	108,654 <sub>15,244</sub>
AIC	4542,194 <sub>325,482</sub>	4587,587 <sub>327,329</sub>	<b>4560,707</b> <sub>328,502</sub>	4565,456 <sub>325,411</sub>	4729,987 <sub>425,810</sub>	4566,148 <sub>338,258</sub>	4966,476 <sub>140,644</sub>
$H_2(e)$	4,090 <sub>0,012</sub>	4,092 <sub>0,011</sub>	4,093 <sub>0,009</sub>	4,093 <sub>0,009</sub>	4,089 <sub>0,010</sub>	<b>4,088</b> <sub>0,011</sub>	4,089 <sub>0,013</sub>
Tiempo	0,165 <sub>0,038</sub>	1,063 <sub>0,510</sub>	1,063 <sub>0,510</sub>	1,063 <sub>0,510</sub>	239,829 <sub>144,697</sub>	<b>0,153</b> <sub>0,060</sub>	498,617 <sub>58,595</sub>
					DT(Y) = 11,825		
$D_{\delta, \hat{\delta}}$	187,152 <sub>4,944</sub>	<b>180,498</b> <sub>5,544</sub>	181,677 <sub>5,274</sub>	182,914 <sub>5,559</sub>	183,813 <sub>7,732</sub>	185,855 <sub>4,872</sub>	205,569 <sub>14,605</sub>
AIC	9442,222 <sub>385,935</sub>	9572,833 <sub>378,670</sub>	9479,122 <sub>375,399</sub>	9542,330 <sub>359,403</sub>	9538,956 <sub>393,251</sub>	<b>9439,217</b> <sub>390,780</sub>	10895,023 <sub>179,560</sub>
$H_2(e)$	4,077 <sub>0,007</sub>	4,077 <sub>0,008</sub>	<b>4,074</b> <sub>0,009</sub>	4,075 <sub>0,008</sub>	4,077 <sub>0,009</sub>	4,077 <sub>0,007</sub>	4,081 <sub>0,009</sub>
Tiempo	<b>0,303</b> <sub>0,038</sub>	2,547 <sub>0,510</sub>	2,547 <sub>0,510</sub>	2,547 <sub>0,510</sub>	346,855 <sub>144,697</sub>	0,420 <sub>0,060</sub>	1425,652 <sub>58,595</sub>
					DT(Y) = 11,041		
$D_{\delta, \hat{\delta}}$	70,070 <sub>5,847</sub>	<b>68,182</b> <sub>4,222</sub>	68,381 <sub>4,224</sub>	68,396 <sub>4,235</sub>	77,933 <sub>15,048</sub>	70,220 <sub>6,458</sub>	84,575 <sub>19,170</sub>
AIC	22988,527 <sub>1064,347</sub>	22907,002 <sub>1100,056</sub>	<b>22892,270</b> <sub>1094,051</sub>	22896,558 <sub>1093,126</sub>	26370,651 <sub>3598,783</sub>	23048,610 <sub>1086,191</sub>	25280,252 <sub>871,622</sub>
$H_2(e)$	5,583 <sub>0,007</sub>	5,588 <sub>0,007</sub>	5,588 <sub>0,007</sub>	5,588 <sub>0,007</sub>	5,588 <sub>0,006</sub>	<b>5,583</b> <sub>0,007</sub>	5,585 <sub>0,008</sub>
Tiempo	0,113 <sub>0,023</sub>	4,119 <sub>0,659</sub>	4,119 <sub>0,659</sub>	4,119 <sub>0,659</sub>	3483,394 <sub>1495,80</sub>	<b>0,101</b> <sub>0,027</sub>	323,998 <sub>18,624</sub>
					DT(Y) = 12,153		
$D_{\delta, \hat{\delta}}$	91,457 <sub>5,470</sub>	91,628 <sub>5,126</sub>	<b>91,137</b> <sub>5,203</sub>	91,149 <sub>5,206</sub>	104,975 <sub>17,365</sub>	91,409 <sub>5,437</sub>	100,542 <sub>11,608</sub>
AIC	22961,128 <sub>805,790</sub>	23234,376 <sub>795,089</sub>	23181,079 <sub>795,44</sub>	23185,416 <sub>793,832</sub>	26358,952 <sub>3440,265</sub>	<b>22979,676</b> <sub>831,064</sub>	24295,830 <sub>326,455</sub>
$H_2(e)$	5,573 <sub>0,004</sub>	5,576 <sub>0,004</sub>	5,575 <sub>0,004</sub>	5,575 <sub>0,004</sub>	<b>5,571</b> <sub>0,006</sub>	5,574 <sub>0,004</sub>	5,576 <sub>0,006</sub>
Tiempo	0,226 <sub>0,022</sub>	4,364 <sub>0,800</sub>	4,364 <sub>0,800</sub>	4,364 <sub>0,800</sub>	2941,702 <sub>946,765</sub>	<b>0,204</b> <sub>0,023</sub>	498,964 <sub>37,740</sub>
					DT(Y) = 12,972		
$D_{\delta, \hat{\delta}}$	201,168 <sub>2,378</sub>	201,138 <sub>2,521</sub>	<b>201,100</b> <sub>2,560</sub>	201,108 <sub>2,565</sub>	204,053 <sub>6,275</sub>	0,000 <sub>0,000</sub>	217,002 <sub>16,618</sub>
AIC	130375,200 <sub>5522,081</sub>	131510,812 <sub>5740,622</sub>	<b>131449,818</b> <sub>5731,829</sub>	131466,196 <sub>5731,970</sub>	134941,445 <sub>10868,319</sub>	0,000 <sub>0,000</sub>	139485,114 <sub>740,127</sub>
$H_2(e)$	6,556 <sub>0,003</sub>	6,555 <sub>0,002</sub>	6,555 <sub>0,002</sub>	6,555 <sub>0,002</sub>	<b>6,555</b> <sub>0,003</sub>	<i>nan<sub>nan</sub></i>	6,556 <sub>0,003</sub>
Tiempo	<b>1,633</b> <sub>0,225</sub>	50,813 <sub>3,104</sub>	50,813 <sub>3,104</sub>	50,813 <sub>3,104</sub>	22192,915 <sub>14299,055</sub>	1,465 <sub>0,294</sub>	2634,509 <sub>119,424</sub>

La tabla 6.5 muestra resultados del experimento completo: la primera fila es la desviación media de las 30 simulaciones que en conjunto tienen las  $m$  variables endógenas del MES. A continuación y por bloques, uno para cada medida, se muestra el nombre del método que ha obtenido el valor medio mínimo en esa medida, apareciendo sombreado el método *Bmom<sub>OPT</sub>*. En total, se han generado MES con catorce variabilidades diferentes, siendo

la máxima de estas variabilidades media igual a seis veces la mínima.

El primer bloque muestra los métodos que han obtenido el valor medio mínimo de la norma euclídea,  $D_{\delta, \hat{\delta}}$ . Se puede observar que en simulaciones con desviación mínima y aumentándola hasta casi un 200%,  $Bmom_{OPT}$  es el método que obtiene el valor medio mínimo en un gran número de MES. Si la variabilidad aumenta por encima del 200%, son los métodos BMOM los que, en la mayoría de MES, ofrecen valores mínimos de norma.

Analizando el bloque del AIC, son los métodos BMOM los que proporcionan valores medios mínimos, seguido por el método MC2E, para todos los niveles de desviación estudiados.

En el bloque de la entropía se repite el patrón de la norma euclídea, no encontrando un valor mínimo para MES aumentando la desviación típica por encima de 150%.

Y por último, atendiendo al tiempo de ejecución, se llega a la misma conclusión que en todos los experimentos llevados a cabo, siendo el método  $Bmom_{OPT}$  el que más tiempo precisa en la estimación. Los valores mínimos medios se alcanzan con los métodos  $Bayes_{2E}$  y MC2E.

---

Tabla 6.5: Métodos con valores mínimos en norma, AIC, entropía y tiempo en segundos, según variabilidad de los datos. 30 iteraciones.

		DT(Y)	1,81	2,07	2,13	2,92	2,97	3,22	4,19	4,63	5,19	5,20	5,34	7,79	10,82	11,84
$D_{\delta, \hat{\delta}}$	10 20 100	$Bmom_{OPT}$	Precisión	Precisión	Bondad											
	10 40 100	$Bmom_{OPT}$	Precisión	Bondad	Bondad											
	20 60 100	$Bayes_{2E}$	$Bmom_{OPT}$	$Bmom_{OPT}$	$Bmom_{OPT}$	$Bmom_{OPT}$	$Bmom_{OPT}$	$Bayes_{2E}$	$Bmom_{OPT}$	$Bayes_{2E}$	$Bmom_{OPT}$	$Bayes_{2E}$	$Bayes_{2E}$	Bondad	Bondad	Bondad
	10 20 400	$Bmom_{OPT}$	Bondad	Bondad												
	10 40 400	$Bmom_{OPT}$	MC2E	Bondad	Precisión											
	20 100 1000	$Bmom_{OPT}$		MC2E	Precisión											
AIC	10 20 100	Precisión	Tradic.	Tradic.	Tradic.	Tradic.	Precisión	Precisión	Precisión	Precisión	Tradic.	Precisión	Precisión	Precisión	Precisión	Precisión
	10 40 100	Precisión	Tradic.	Precisión	Precisión	Tradic.	Precisión	Precisión	MC2E	Precisión	Precisión	MC2E	MC2E	$Bayes_{2E}$	MC2E	
	20 60 100	MC2E	Precisión	Bondad	Bondad	Precisión	MC2E	MC2E	MC2E	Precisión	MC2E	MC2E	Precisión	MC2E	$Bayes_{2E}$	
	10 20 400	Tradic.	Bondad	Tradic.	Tradic.	Bondad	MC2E	$Bayes_{2E}$	Precisión							
	10 40 400	Tradic.	Bondad	Tradic.	Tradic.	Bondad	MC2E	MC2E	MC2E							
	20 100 1000	Tradic.	Precisión	Tradic.	Tradic.			MC2E	MC2E							
$H_2(e)$	10 20 100	MCMC	$Bmom_{OPT}$	$Bmom_{OPT}$	$Bmom_{OPT}$	$Bmom_{OPT}$	$Bmom_{OPT}$	$Bmom_{OPT}$	Bondad	$Bmom_{OPT}$	Precisión	Precisión	Bondad	$Bmom_{OPT}$	Precisión	
	10 40 100	$Bmom_{OPT}$	Precisión	MC2E	$Bmom_{OPT}$	MC2E	Precisión	$Bayes_{2E}$	$Bayes_{2E}$							
	20 60 100	$Bayes_{2E}$	$Bayes_{2E}$	$Bmom_{OPT}$	$Bmom_{OPT}$	$Bmom_{OPT}$	$Bmom_{OPT}$	Bondad	Precisión	$Bayes_{2E}$	MC2E	MC2E	Tradic.	$Bmom_{OPT}$	MCMC	
	10 20 400	$Bmom_{OPT}$	Precisión	Tradic.	MC2E	MC2E	Tradic.	$Bmom_{OPT}$	MC2E							
	10 40 400	$Bmom_{OPT}$	$Bmom_{OPT}$	MC2E	Precisión	$Bmom_{OPT}$	MC2E	$Bmom_{OPT}$	Tradic.	$Bmom_{OPT}$	$Bmom_{OPT}$	$Bmom_{OPT}$	$Bmom_{OPT}$	Tradic.	$Bmom_{OPT}$	
	20 100 1000	$Bmom_{OPT}$	$Bmom_{OPT}$	MC2E	MC2E	$Bmom_{OPT}$		$Bmom_{OPT}$	$Bmom_{OPT}$							
Tiempo	10 20 100	MC2E	MC2E	MC2E	$Bayes_{2E}$	MC2E	MC2E	MC2E	MC2E	$Bayes_{2E}$	$Bayes_{2E}$	MC2E	MC2E	$Bayes_{2E}$	MC2E	
	10 40 100	$Bayes_{2E}$	$Bayes_{2E}$	MC2E	$Bayes_{2E}$	$Bayes_{2E}$	MC2E	MC2E	$Bayes_{2E}$	MC2E	$Bayes_{2E}$	MC2E	MC2E	$Bayes_{2E}$	$Bayes_{2E}$	
	20 60 100	MC2E														
	10 20 400	$Bayes_{2E}$														
	10 40 400	$Bayes_{2E}$	MC2E	$Bayes_{2E}$	$Bayes_{2E}$	$Bayes_{2E}$										
	20 100 1000	$Bayes_{2E}$	MC2E	$Bayes_{2E}$	$Bayes_{2E}$	$Bayes_{2E}$		$Bayes_{2E}$	$Bayes_{2E}$							

La tabla 6.6 se muestra el porcentaje de modelos en los que  $D_{\delta, \hat{\delta}}$  y  $H_2(e)$  coinciden, para diferentes desviaciones típicas de las variables endógenas, es decir, muestra el porcentaje de veces en los que siendo el método que ofrece una estimación con menor norma, es el criterio  $H_2(e)$  el que seleccionaría ese método. De esta tabla se desprende que al aumentar la desviación típica hasta un 130 %, la entropía  $H_2(e)$  selecciona el método en el que  $D_{\delta, \hat{\delta}}$  es mínima en torno a un 80 % de los casos.

Tabla 6.6: Porcentaje de coincidencia entre  $D_{\delta, \hat{\delta}}$  y  $H_2(e)$

DT(Y)	1,81	2,07	2,13	2,92	2,97	3,22	4,19	4,63	5,19	5,20	5,34	7,79	10,82	11,84
%	83,33 %	83,33 %	66,67 %	66,67 %	100,00 %	66,67 %	83,33 %	33,33 %	50,00 %	50,00 %	33,33 %	33,33 %	0,00 %	0,00 %

Finalmente, en la tabla 6.7 se muestra un resumen con los porcentajes de modelos en los que  $Bmom_{OPT}$  es el mejor en términos de menor norma euclídea y con la frecuencia en que  $H_2(e)$  lo seleccionaría como tal.

Tabla 6.7: Porcentaje de MES con  $Bmom_{OPT}$  mejor método de estimación y es seleccionado por  $H_2(e)$ .

DT(Y)	Número de modelos	$D_{\delta, \hat{\delta}}$	$H_2(e)$
1,00 - 2,99	30	96,67 %	76,67 %
3,00 - 6,00	35	88,57 %	58,62 %
> 6,00	17	0,00 %	41,18 %

### 6.3. Conclusiones

Se ha llevado a cabo un estudio de estimación de MES que contemplan escenarios con diferentes variabilidades con el objetivo de encontrar una relación entre los métodos y el error de estimación, y el criterio de selección basado en la entropía. Como principales conclusiones se extrae que, en MES con variables endógenas con variabilidad comprendida entre (1,00 - 6,00), el método de estimación  $Bmom_{OPT}$  es el que obtiene menor norma euclídea frente a los métodos MC2E, los BMOM y Bayesianos. Tal y como se definió en el capítulo 5, la norma euclídea es una medida de la distancia entre los parámetros del modelo y sus estimaciones, por lo que  $Bmom_{OPT}$  se le puede atribuir la propiedad de método con mínimo error de estimación de parámetros bajo esas condiciones de variabilidad.

Otro resultado que se extrae del experimento es la consideración de la medida basada en entropía,  $H_2(e)$ , como alternativa al AIC como criterio para seleccionar el método de estimación con menor error. Además, los resultados obtenidos con este criterio obtienen un alto nivel de coincidencia con el método que alcanza la menor  $D_{\delta, \hat{\delta}}$ .

Y como era previsible, cuando la variabilidad toma valores elevados no hay un método cuya bondad de estimación sea mejor que el resto y de la misma manera, ni  $H_2(e)$  ni AIC son buenas medidas como criterio de selección del método de estimación.

Por otro lado, el método  $Bmom_{OPT}$  tiene mayor coste computacional que el resto de métodos estudiados en este trabajo. Dependiendo del tipo de problema a resolver, debe plantearse si es preferible la estimación con un resultado óptimo (menor error) u obtener

una estimación con mayor error pero en un menor tiempo de cómputo. En el ámbito de la salud, por ejemplo, generalmente prioriza una estimación con error mínimo ante la minimización del tiempo, por lo que  $Bmom_{OPT}$  sería la mejor opción en esos casos.

---



# CAPÍTULO 7

---

## **Resolución de un Modelo de Ecuaciones Simultáneas mediante una Red Neuronal Artificial**

---

En este capítulo se lleva a cabo la aplicación de redes neuronales artificiales para resolver modelos de ecuaciones simultáneas. Concretamente se han estudiado dos casos, en primer lugar se estima un MES de dimensión reducida tanto en tamaño muestral como en el número de variables endógenas y exógenas, y en segundo lugar un MES de mayor dimensión. El objetivo en ambos casos es obtener predicciones de las variables endógenas del modelo a través del entrenamiento de una RNA.

### **7.1. Resolución de un MES mediante una RNA: caso 1**

En esta sección se ha generado un MES con dos variables endógenas, cuatro exógenas y un total de 19 observaciones. Las relaciones entre las variables quedan reflejadas en el

modelo (7.1):

$$\begin{aligned} y_1 &= \beta_{1,2}y_2 + \gamma_{1,1}x_1 + \gamma_{1,2}x_2 + u_1 \\ y_2 &= \beta_{2,1}y_1 + \gamma_{2,2}x_2 + \gamma_{2,3}x_3 + \gamma_{2,4}x_4 + u_2 \end{aligned} \quad (7.1)$$

Los datos, generados de manera similar a los experimentos que fueron descritos en los capítulos 5 y 6, se presentan en la tabla 7.1:

Tabla 7.1: Datos generados para la resolución de un MES mediante RNA: caso 1.

obs.	$y_1$	$y_2$	$x_1$	$x_2$	$x_3$	$x_4$	cte.
1	14,70	27,13	7	3	0	10	1
2	9,90	41,80	3	4	10	3	1
3	17,10	41,87	9	3	8	8	1
4	3,40	30,60	0	2	10	0	1
5	22,80	52,00	10	7	8	7	1
6	13,20	38,13	6	3	9	4	1
7	12,80	33,53	7	1	7	8	1
8	4,60	26,80	0	4	0	8	1
9	11,70	53,90	2	8	5	10	1
10	12,60	56,37	2	9	7	7	1
11	5,50	31,67	1	3	3	9	1
12	5,10	15,67	3	0	0	9	1
13	14,20	48,20	3	9	8	1	1
14	8,50	28,87	3	3	1	10	1
15	21,10	33,00	9	7	3	2	1
16	8,00	29,10	1	6	3	2	1
17	14,80	29,93	7	3	2	9	1
18	10,00	18,53	3	4	1	2	1
19	19,60	37,07	7	8	3	3	1

La red neuronal se ha implementado en  a través del paquete *neuralnet* [62]. Entre las opciones disponibles en este paquete, cabe destacar la selección del número de capas ocultas de la red, de la función de activación y de la función para el cálculo del error, así como la representación gráfica de la red resultante. En este problema se ha entrenado una RNA sin capas ocultas, función de activación logística (7.1) y algoritmo backpropagation.

$$f(x) = \frac{1}{1 - e^{-x}}$$

Figura 7.1: Función de activación logística

Esta red se ha configurado de tal manera que el 70% de la muestra se ha utilizado para entrenamiento y el 30% restante para validación. La red resultante se muestra en la figura 7.2 y las predicciones para cada variable endógena y el error cometido como diferencia con el valor real correspondiente, y el valor del error cuadrático medio (ECM) se muestran en la tabla 7.2:

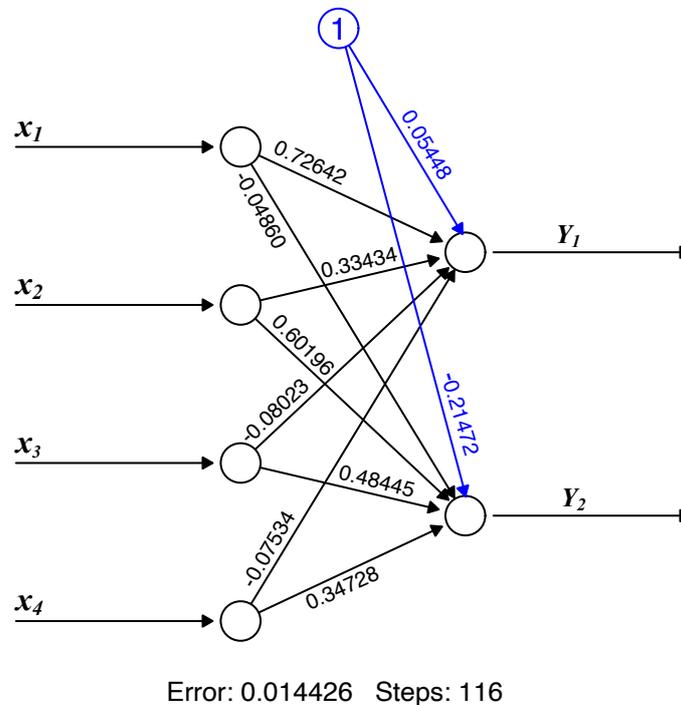


Figura 7.2: Red neuronal artificial con 4 inputs, 2 outputs y sin capa oculta.

Además de la RNA, se han estimado los parámetros del modelo para comparar las predicciones obtenidas con ambas técnicas. En concreto, se han utilizado los métodos vistos en el capítulo 5, con los que se han obtenido resultados muy similares entre sí. La tabla 7.3 muestran las predicciones obtenidas con las estimaciones del método  $Bmom_{OPT}$ :

Tabla 7.2: Predicciones y ECM con la RNA

obs.	$\hat{y}_1$	$\hat{u}_1$	$\hat{y}_2$	$\hat{u}_2$
5	21,326	-1,47	58,199	6,20
6	13,0891	-0,11	37,7342	-0,40
8	6,17039	1,57	21,0687	-5,73
12	7,36924	2,27	23,0195	7,35
14	9,22949	0,73	28,4343	-0,43
18	11,1195	1,12	31,6912	13,16
ECM=25,87				

Tabla 7.3: Predicciones y ECM con  $Bmom_{OPT}$ 

obs	$\hat{y}_1$	$\hat{u}_1$	$\hat{y}_2$	$\hat{u}_2$
1	27,07	35,40678	-12,37	-8,27
2	29,40	46,89616	-19,50	-5,10
3	37,43	52,33174	-20,33	-10,47
4	17,30	31,58916	-13,90	-0,99
5	48,00	66,44262	-25,20	-14,44
6	31,07	45,41478	-17,87	-7,28
7	28,27	41,14882	-15,47	-7,62
8	17,40	28,38344	-12,80	-1,58
9	37,95	60,09008	-26,25	-6,19
10	40,18	63,35354	-27,58	-6,99
11	20,33	34,6467	-14,83	-2,98
12	12,33	18,37944	-7,23	-2,71
13	37,60	56,74618	-23,40	-8,55
14	21,93	33,6335	-13,43	-4,77
15	37,00	45,31454	-15,90	-12,31
16	22,05	33,6366	-14,05	-4,54
17	28,47	38,47342	-13,67	-8,54
18	17,77	24,2932	-7,77	-5,76
19	37,03	48,98824	-17,43	-11,92
ECM=153,10				

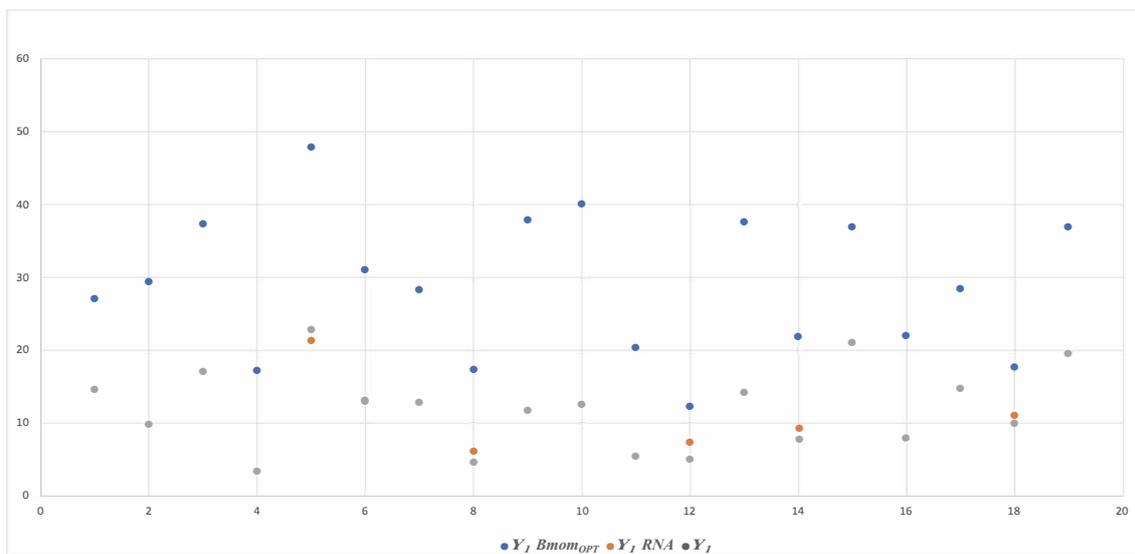


Figura 7.3: Comparativa de los valores  $Y_1$  frente a  $\hat{Y}_1$  obtenidos por  $Bmom_{OPT}$  y la RNA

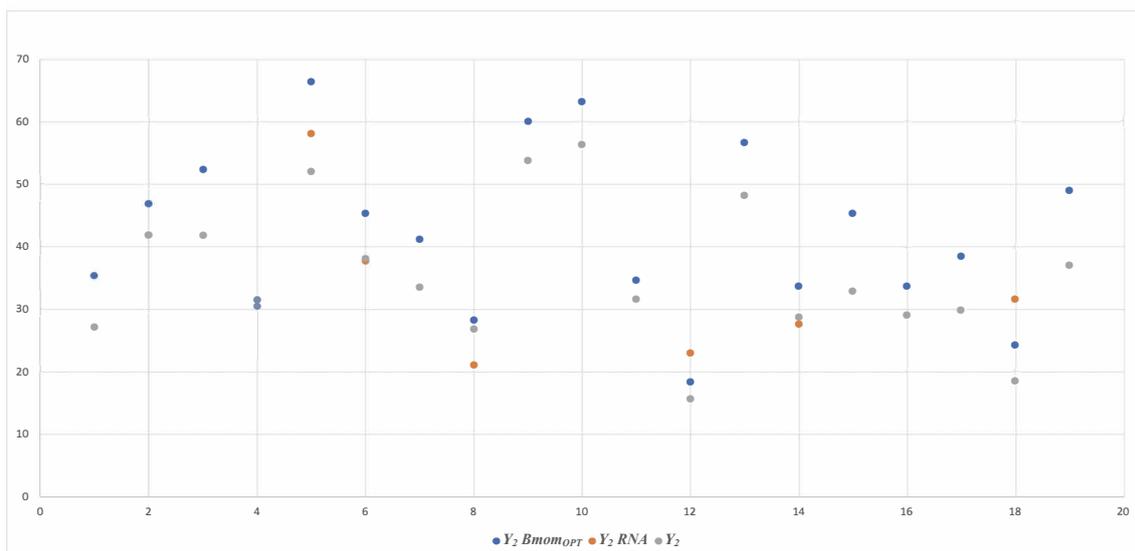


Figura 7.4: Comparativa de los valores  $Y_2$  frente a  $\hat{Y}_2$  obtenidos por  $Bmom_{OPT}$  y la RNA

Y a pesar de que las predicciones con la RNA solo se han obtenido para una parte del conjunto de datos, lo que se denomina muestra de testeo, en los gráficos 7.3 - 7.4 se observa que las predicciones de la variable  $Y_1$  son mejores que las de  $Y_2$ , aunque, en su conjunto, tal y como se expuso en las tablas 7.2 - 7.3 se obtiene mejor predicción con la RNA que con el método  $Bmom_{OPT}$ .

Comparando los resultados obtenidos por ambas técnicas y, aunque hay un número diferente de predicciones, el EMC que se obtiene con la RNA es menor que el obtenido con  $Bmom_{OPT}$ .

## 7.2. Resolución de un MES mediante una RNA: caso 2

En esta sección se ha resuelto otro MES de mayor dimensión, concretamente con 10 variables endógenas, 20 exógenas y un tamaño muestral igual a 2000. Igual que en el ejemplo anterior, los datos han sido generados de manera similar a los experimentos que fueron descritos en los capítulos 5 y 6, no presentándolos en este caso por el gran tamaño muestral. De manera similar, esta red se ha entrenado de nuevo en  a través del paquete *neuralnet*, sin capas ocultas, función de activación logística (7.1) y algoritmo backpropagation, utilizando el 70 % de la muestra para su entrenamiento y el 30 % restante para su validación.

Como resultados se presentan los errores cuadráticos medios obtenidos con la red para cada output 7.4, así como el ECM de la red completa y el obtenido por la estimación con el método  $Bmom_{OPT}$  7.5.

A la vista de los resultados, se puede concluir que el error obtenido es menor con la red neuronal que con el método  $Bmom_{OPT}$ , al igual que ocurría en el ejemplo 1. En este caso, dada la magnitud de la red, no se muestra el gráfico de la RNA.

---

Tabla 7.4: Error cuadrático medio para cada output de la RNA

Output	SE
$y_1$	17081,77
$y_2$	88029,55
$y_3$	2498,205
$y_4$	1535,839
$y_5$	20557,99
$y_6$	29749,55
$y_7$	1725,62
$y_8$	36722,94
$y_9$	224800,2
$y_{10}$	4460,649

Tabla 7.5: Error cuadrático medio de la RNA y método  $Bmom_{OPT}$ 

RNA	$Bmom_{OPT}$
17081,77	72400,30

### 7.3. Conclusiones

El objetivo de este capítulo es introducir la resolución de MES mediante en entrenamiento de una RNA. Concretamente, se han resuelto dos MES en los que la generación de los datos, método de estimación y tipo de red neuronal es similar en ambos casos, variando únicamente la dimensión del problema. Se ha entrenado una red neuronal *clásica* de tipo backpropagation, sin capas ocultas y sin introducir en su diseño la arquitectura propia que posee la característica de simultaneidad de un MES.

Con estos ejemplos, queda mostrada una alternativa para resolver estos problemas mediante la aplicación de las redes neuronales artificiales sin necesidad de un modelado previo, al estar basadas en un concepto de entrenamiento-aprendizaje.

Y, finalmente, si el problema planteado solo requiere un modelo para predecir valores de las variables endógenas, una alternativa con menor ECM es entrenar una RNA. Por el

contrario, si el problema requiere la estimación de los parámetros del modelo, la solución habría que plantearla aplicando métodos de inferencia estadística.

---

# CAPÍTULO 8

---

## Ejemplos

---

El objetivo de este capítulo es aplicar los métodos de estimación y criterios de información estudiados en el capítulo 5 así como llevar a cabo la comparación de los resultados obtenidos con los diferentes métodos y medidas en ejemplos. El capítulo se ha dividido en dos secciones, presentando en cada una de ellas un problema modelado a través de un MES. Ambas aplicaciones tienen en común un alto interés para la sociedad al abordar temas como la economía y el bienestar social o la salud.

La primera aplicación trata un problema clásico dentro del marco de la microeconomía, del que hay disponibles estimaciones llevadas a cabo en la literatura.

Por el contrario, del segundo problema no se dispone de ningún modelado previo, siendo necesaria una primera etapa para determinar las relaciones existentes entre las variables, para posteriormente estimar los coeficientes.

## 8.1. Modelo de la Oferta de mano de obra de trabajadoras casadas

El modelo descrito en *The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions*, de Mroz (1987) [5], relaciona el salario percibido por mujeres casadas trabajadoras con las horas dedicadas al trabajo junto otros factores socio-familiares y laborales. El MES, que relaciona de manera simultánea estas variables, se presenta en (8.1), siendo la primera ecuación la función de oferta de trabajo de las mujeres casadas trabajadoras y la segunda ecuación, la función de oferta salarial:

$$\begin{aligned} \text{hours} &= \gamma_{11} + \beta_{12} \text{lwage} + \gamma_{12} \text{educ} + \gamma_{13} \text{age} + \gamma_{14} \text{kidslt6} + \gamma_{15} \text{kidsge6} + \gamma_{16} \text{nwifeinc} + u_1 \\ \text{lwage} &= \gamma_{21} + \beta_{21} \text{hours} + \gamma_{22} \text{educ} + \gamma_{23} \text{exper} + \gamma_{24} \text{expersq} + u_2 \end{aligned} \quad (8.1)$$

A continuación, se describen las variables endógenas:

- *hours*: número de horas trabajadas en 1975.
- *lwage*: salario percibido por hora, en logaritmos.

y las variables exógenas:

- *kidslt6*: número de hijos menores de 6 años.
  - *kidsge6*: número de hijos entre 6 y 18 años.
  - *age*: edad de la mujer en años.
  - *educ*: años de educación de la mujer.
  - *exper*: experiencia laboral de la mujer.
  - *nwifeinc*: otros ingresos en la familia, incluye el salario del marido.
  - *expersq*: experiencia al cuadrado.
-

Los datos están disponibles en el libro de J.M. Wooldridge [63], son de sección cruzada y corresponden a 428 mujeres casadas que se encontraban trabajando en 1975. Para las variables endógenas se han calculado los estadísticos más importantes que se recogen en la tabla 8.1, que muestra la media, la desviación típica (DT) y el coeficiente de variación (CV).

Tabla 8.1: Media, DT y CV de las variables endógenas del modelo de la Oferta de mano de obra de trabajadoras casadas.

Variable	Media	DT	CV
hours	1302,930	776,274	0,596
lwage	1,190	0,723	0,608

La variable *hours* tiene una DT elevada por lo que también se ha calculado el CV y así disponer de una medida de variabilidad relativa a su media que pueda servir de referencia con las variabilidades utilizadas en el estudio comparativo llevado a cabo en el capítulo 6.

El modelo ha sido estimado con todos los métodos estudiados en esta tesis, a partir de los cuales se han calculado los valores de AIC, entropías, tiempo de ejecución y los parámetros  $k_1$  y  $k_2$  obtenidos por el método  $Bmom_{OPT}$ .

En primer lugar, se muestran los coeficientes estimados en la tabla 8.2, en la que se observa que las estimaciones obtenidas son similares entre métodos, encontrando tan solo un cambio de signo en dos parámetros estimados con MCMC respecto al resto de métodos. En concreto, los efectos de las variables *nwifenic* y *expersq* son positivos en MCMC, mientras que en el resto de métodos, la estimación es negativa. Al comparar los tres métodos BMOM se observan valores en las estimaciones muy próximos entre sí, tanto en magnitud como en signo. El método  $Bayes_{2E}$  obtiene estimaciones muy próximas a las obtenidas por MC2E y por último, el método  $Bmom_{OPT}$  también obtiene parámetros estimados en concordancia con el resto de métodos.

Tabla 8.2: Coeficientes estimados del MES de la Oferta de mano de obra de trabajadoras casadas.

hours	lwage	educ	age	kidslt6	kidsge6	nwifenic	exper	expersq	cte
MC2E									
-1	1544,8187	-177,4490	-10,7841	-210,8339	-47,5571	-9,2491	0	0	2432,1980
0,0002	-1	0,1111	0	0	0	0	0,0326	-0,0007	-0,6928
Bondad ajuste									
-1	1085,4371	-129,5167	-9,8860	-249,5525	-67,3951	-7,7779	0	0	2338,8351
0,0001	-1	0,1101	0	0	0	0	0,035	-0,0007	-0,6473
Precisión estimación									
-1	1062,0095	-127,0722	-9,8402	-251,5271	-68,4068	-7,7028	0	0	2334,0738
0,0001	-1	0,1097	0	0	0	0	0,036	-0,0007	-0,6284
Tradicional									
-1	1058,7647	-126,7337	-9,8339	-251,8005	-68,5469	-7,6924	0	0	2333,4143
0,0001	-1	0,1097	0	0	0	0	0,036	-0,0007	-0,6280
$Bmom_{OPT}^a$									
-1	103,9948	-27,1121	-7,9673	-332,2725	-109,7778	-4,6346	0	0	2139,3707
0,0001	-1	0,109	0	0	0	0	0,0379	-0,0008	-0,5928
$Bayes_{2E}$									
-1	1526,8106	-174,6931	-10,5548	-217,2800	-47,8460	-9,2155	0	0	2409,4337
0,0001	-1	0,1089	0	0	0	0	0,0337	-0,0007	-0,6375
MCMC									
-1	19,5946	-12,5101	-22,5204	-335,9012	-84,6159	0,3292	0	0	2387,8466
0,0000	-1	0,2013	0	0	0	0	0,0430	0,0014	-0,6250

<sup>a</sup> Estimación con  $Bmom_{OPT}$  obtenida con los valores  $k_1=-1,054$  y  $k_2=3$ .

En cada ecuación, el valor -1 indica la variable dependiente, y el 0, que la variable no está presente en la ecuación.

En la tabla 8.3 se muestran el AIC, entropías y tiempo de ejecución. Cabe resaltar que en todos los casos, el método de estimación que ofrece menor AIC y entropías es el método MC2E.

Tabla 8.3: AIC, entropías, tiempo de ejecución y  $k_1$  y  $k_2$  para el modelo de Oferta de mano de obra de trabajadoras casadas.

Medida	MC2E	BMOM			$Bmom_{OPT}$	$Bayes_{2E}$	MCMC
		Ajuste	Estimación	Tradicional			
AIC	4850,069	4971,789	4992,206	4992,954	5080,592	4899,667	5970,783
$H(e)$	243,987	249,001	249,503	249,547	273,362	244,398	321,355
$H_2(e)$	5,497	5,517	5,519	5,520	5,606	5,499	5,772
$H_{1/2}(e)$	336,220	339,320	339,575	339,612	355,550	336,203	385,317
Segundos	0,120	0,138	0,138	0,138	33,750	0,118	6,384
$k_1$	1,000	0,983	0,983	0,983	0,429		
$k_2$	1,000	1,000	0,983	0,983	2,000		

También se han obtenido esas mismas medidas al estimar con  $Bmom_{OPT}$  con diferentes valores de los parámetros  $k_1$  y  $k_2$ . y los resultados se muestran en la tabla 8.4.

Tabla 8.4: AIC, entropías y tiempo de ejecución obtenidos con  $Bmom_{OPT}$  con diferentes valores de  $k_1$  y  $k_2$  para el modelo de Oferta de mano de obra de trabajadoras casadas.

AIC	5941,970	5935,515	5101,005	5101,621	5102,967	5106,625	5100,620	5100,633	5107,209
$H(e)$	308,593	307,882	275,793	275,977	276,383	277,516	275,703	275,728	277,605
$H_2(e)$	5,731	5,729	5,615	5,616	5,617	5,621	5,615	5,615	5,621
$M_{1/2}$	374,038	373,470	356,186	356,375	356,761	357,602	356,205	356,331	357,619
Segundos	15,053	13,591	21,914	24,436	22,699	16,405	28,806	23,686	18,448
$k_1$	-3,000	-3,000	-1,054	-1,149	-1,406	-2,929	-3,444	-11,580	-36,000
$k_2$	1,104	1,005	3,000	3,000	3,000	3,000	6,000	16,000	21,489
AIC	5110,441	5102,708	5080,592	5109,655	5100,568	5083,719	6024,233	5104,910	5379,323
$H(e)$	280,231	276,279	273,362	279,658	275,664	270,111	330,277	276,901	298,440
$H_2(e)$	5,631	5,617	5,606	5,629	5,615	5,596	5,799	5,619	5,698
$M_{1/2}$	359,817	356,523	355,550	359,405	356,045	353,469	388,333	356,996	371,863
Segundos	19,552	30,185	33,750	14,951	26,395	39,306	1,646	23,462	84,817
$k_1$	-32,373	-0,391	0,429	-2,000	-0,993	0,882	0,000	-1,975	1,011
$k_2$	-3,726	2,000	2,000	0,761	3,000	0,998	0,000	3,000	0,351

De la estimación se puede concluir que los diferentes métodos proporcionan resultados similares, encontrado una pequeña diferencia en el signo de dos parámetros con MCMC.

La desviación típica en términos absolutos de una de las variables endógenas es muy elevada, aunque en términos relativos es pequeña. Según lo obtenido en el capítulo 6, el método que proporciona menor entropía debería ser  $Bmom_{OPT}$ , y aunque los valores de AIC y las entropías son muy similares entre sí, es el método MC2E el que obtiene el valor mínimo.

## **8.2. MES para un problema de relación entre hormonas y otras variables**

Las hormonas sexuales no sólo son importantes en el aspecto sexual del individuo, sino que también participan en muy diversas funciones de prácticamente todos los tejidos y están relacionadas con diversas enfermedades. Para la segunda aplicación se dispone de una muestra de 19 individuos con variables relacionadas con niveles hormonales, parámetros seminales, morfología espermática y calidad de vida. Este conjunto de datos procede del grupo de investigación Salud Pública y Epidemiología del Hospital Virgen de la Arrixaca, Murcia (España), y uno de los principales objetivos es estimar los efectos que producen unas variables sobre otras, permitiendo una relación de simultaneidad. Concretamente, se quieren estimar los efectos que el resto de variables tienen sobre los niveles hormonales a la vez que se relacionan estos indicadores entre sí.

El conjunto de variables puede dividirse en tres grupos según la naturaleza de las mismas, las cuales están descritas en la tabla 8.5:

- i. Variables indicadoras de niveles hormonales,
  - ii. Variables que caracterizan el semen de los individuos, y
  - iii. Variables recogidas a través de un cuestionario sobre el estilo de vida.
-

Tabla 8.5: Variables utilizadas en el modelo del problema de hormonas

VARIABLES	DESCRIPCIÓN
$y_1$	FSH: niveles de la hormona foliculoestimulante
$y_2$	LH: niveles de la hormona luteinizante
$y_3$	Estradiol
$y_4$	Testosterona total en suero
$y_5$	Hormona sérica unión gonadotropinas
$x_1$	Distancia dedos d2 medida 1
$x_2$	Distancia dedos d2 medida 2
$x_3$	Distancia dedos d4 medida 1
$x_4$	Distancia dedos d4 medida 2
$x_5$	Talla en cm
$x_6$	Peso en kilos
$x_7$	Índice masa corporal
$x_8$	Tamaño testículo derecho (mL)
$x_9$	Tamaño testículo izquierdo (mL)
$x_{10}$	Distancia ano - pene
$x_{11}$	Distancia ano - escroto
$x_{12}$	Peso de la muestra
$x_{13}$	Media Concentración espermática
$x_{14}$	Media movimiento rápido
$x_{15}$	Media movimiento lento
$x_{16}$	Media movimiento no progresivo
$x_{17}$	Media espermatozoides inmóviles
$x_{18}$	% Espermatozoides normales

Al no disponer de un modelo con las relaciones entre las variables definidas previamente, es imprescindible en primer lugar obtener el MES que describe esas relaciones simultáneas del conjunto de variables. Para la búsqueda del modelo se ha recurrido a un algoritmo que, de entre todos los modelos posibles, elige como mejor solución aquél con AIC mínimo y sujeto a las condiciones de orden y rango.

El modelo resultante queda reflejado en la tabla 8.6, en la cuál, el -1 indica la variable dependiente, el 1, que la variable está en la ecuación y el 0, que no lo está. Este MES está formado por 5 variables endógenas (mismo número de ecuaciones) y 18 variables exógenas.

Tabla 8.6: MES propuesto para el problema de hormonas.

$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$	$x_{13}$	$x_{14}$	$x_{15}$	$x_{16}$	$x_{17}$	$x_{18}$
-1	1	1	1	1	1	0	1	0	1	1	1	1	0	0	1	0	1	1	1	1	1	1
0	-1	1	0	0	0	1	1	1	1	0	1	0	1	1	1	1	1	0	1	1	0	0
1	0	-1	0	1	1	1	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0
1	1	0	-1	0	1	1	1	1	0	0	0	0	1	1	1	0	1	1	0	1	1	1
1	0	0	1	-1	0	1	0	1	1	0	1	1	0	1	0	1	0	1	1	0	0	1

El modelo propuesto, expresado en forma de sistema de ecuaciones, se muestra a continuación:

$$\begin{aligned}
y_1 &= \beta_{1,2} y_2 + \dots + \beta_{1,5} y_5 + \gamma_{1,1}x_1 + \gamma_{1,3}x_3 + \gamma_{1,5}x_5 + \dots + \gamma_{1,8}x_8 + \gamma_{1,11}x_{11} + \gamma_{1,13}x_{13} + \\
&\quad + \dots + \gamma_{1,18}x_{18} + u_1 \\
y_2 &= \beta_{2,3} y_3 + \gamma_{2,1}x_2 + \gamma_{2,3}x_3 + \dots + \gamma_{2,5}x_5 + \gamma_{2,7}x_7 + \gamma_{2,9}x_9 + \dots + \gamma_{2,13}x_{13} + \gamma_{2,15}x_{15} + \\
&\quad + \gamma_{2,16}x_{16} + u_2 \\
y_3 &= \beta_{3,1} y_1 + \beta_{3,5} y_5 + \gamma_{3,1}x_1 + \dots + \gamma_{3,3}x_3 + \gamma_{3,5}x_5 + \gamma_{3,12}x_{12} + \gamma_{3,14}x_{14} + \gamma_{2,16}x_{16} + u_3 \\
y_4 &= \beta_{4,1} y_1 + \beta_{4,2} y_2 + \gamma_{4,1}x_1 + \dots + \gamma_{4,3}x_3 + \gamma_{4,9}x_9 + \dots + \gamma_{4,11}x_{11} + \gamma_{4,13}x_{13} + \gamma_{4,14}x_{14} + \\
&\quad + \gamma_{4,16}x_{16} + \dots + \gamma_{4,18}x_{18} + u_4 \\
y_5 &= \beta_{5,1} y_1 + \beta_{5,4} y_4 + \gamma_{5,1}x_2 + \gamma_{5,4}x_4 + \gamma_{5,5}x_5 + \gamma_{5,5}x_5 + \gamma_{5,8}x_8 + \gamma_{5,10}x_{10} + \gamma_{5,12}x_{12} + \\
&\quad + \gamma_{5,14}x_{14} + \gamma_{5,15}x_{15} + \gamma_{5,18}x_{18} + u_5
\end{aligned} \tag{8.2}$$

Los datos utilizados para la estimación del MES se muestran en la tabla 8.7, de la que se puede extraer que la variabilidad de los datos, medida como la raíz cuadrada de la media de las desviaciones cuadráticas de cada variable respecto a su media es 5,84. Los coeficientes estimados por los diferentes métodos, se muestran en la tabla 8.8, en la cual, se puede observar que, en general, los métodos utilizados proporcionan valores similares en magnitud y signo, encontrando diferencias más notables en las estimaciones obtenidas con MCMC, sobre todo en la magnitud de algunos coeficientes. Pero, dado que el MES tiene un número elevado de variables, es difícil controlar todas excepto una, para medir el efecto que ésta causa en el variable dependiente, salvo que un especialista en el tema pueda determinar sin lugar a duda, al menos, el signo del efecto.

Tabla 8.7: Datos utilizados para estimar el MES del problema de hormonas.

$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$	$x_{13}$	$x_{14}$	$x_{15}$	$x_{16}$	$x_{17}$	$x_{18}$
3,24	4,02	16,96	1,85	17,60	102,90	102,80	102,20	102,30	180	78	24,07	23	22	127,57	30,35	4,4	9,12	8,2	55,3	14,7	21,8	25
3,16	3,31	30,71	4,89	27,30	99,82	99,80	99,70	99,73	175	77	25,14	23	21	66,22	147,38	3,4	5,50	14,6	34,8	13,5	37	17
1,60	3,05	16,42	3,71	30,90	106,01	106,09	105,37	105,48	180	90	27,78	24	23	121,87	37,96	3,5	4,60	3,5	58,7	8,6	29,2	14
2,53	2,89	37,19	4,50	25,90	111,57	111,56	122,50	122,70	186	112	32,37	22	21	153,42	68,01	2,2	1,32	1,5	31,8	19,3	47,3	10
4,40	2,83	20,04	3,39	24,20	112,18	112,00	109,96	109,90	184	80	23,63	22	20	117,56	35,39	2,7	9,40	13,8	42,0	8,8	35,3	21
3,70	2,74	33,16	1,94	23,80	109,56	109,67	112,73	112,80	192	95	25,77	22	20	138,31	54,75	5,3	26,60	26,7	29,2	8,2	36	16
2,78	2,61	35,62	3,76	29,20	108,10	108,12	109,11	109,50	189	94	26,32	24	23	138,28	57,13	4,5	22,00	39,8	18,8	6,8	34,5	11
1,19	3,25	68,75	5,05	33,20	96,52	96,57	99,43	99,40	172	70	23,66	23	22	128,28	40,76	3,6	21,20	35,2	38,0	7,0	19,8	25
2,45	2,85	19,66	5,17	41,60	95,86	95,90	95,53	95,57	166	55	19,96	22	20	107,12	38,33	6,0	52,00	24,5	43,7	8,3	23,5	20
4,53	2,92	26,90	4,02	22,90	96,85	96,80	97,90	97,89	163	71	26,72	20	20	116,07	23,97	4,6	10,00	21,3	52,0	6,5	20,2	20
4,20	5,06	24,70	3,51	22,10	116,15	116,10	116,10	116,14	188	85	24,05	25	24	112,42	29,83	5,2	16,00	6,1	56,5	11,2	26,2	22
3,20	2,69	33,74	5,22	36,80	109,47	109,52	114,91	114,98	180	80	24,69	23	22	128,22	36,40	4,5	8,07	16,5	32,8	16,5	34,16	28
2,49	3,99	17,88	3,86	36,50	106,96	106,95	106,80	106,99	188	76	21,50	23	22	120,85	31,31	3,4	11,59	30,0	30,8	12,0	27,16	32
5,65	3,93	18,75	2,52	34,30	109,54	109,54	111,54	111,64	180	80	24,69	21	20	118,34	41,13	4,5	10,60	10,2	36,3	18,3	35,2	19
7,54	7,46	34,00	8,28	70,30	108,60	108,70	113,21	113,20	178	82	25,88	24	22	124,43	55,90	5,0	28,60	15,3	40,2	9,7	34,8	15
2,74	1,78	24,51	1,48	23,90	108,05	108,06	111,88	111,95	171	70	23,94	17	15	118,49	51,02	2,9	9,05	19,0	26,8	8,3	45,83	17
3,34	1,83	20,00	3,47	43,40	98,10	98,12	103,42	103,40	169	67	23,46	20	20	119,25	42,80	4,5	25,00	40,2	18,2	10,8	30,83	32
3,91	3,21	20,34	2,09	34,60	105,00	105,30	112,00	112,50	176	88	28,41	25	24	135,14	62,13	4,9	15,47	28,2	32,8	10,2	28,83	22
14,38	8,39	38,23	4,29	35,00	102,82	102,90	110,07	110,10	175	72	23,51	20	19	117,93	46,73	4,8	11,30	41,7	23,0	9,5	25,83	35

Tabla 8.8: Coeficientes estimados del MES del problema de hormonas.

	$x_{31}$	$x_{32}$	$x_{33}$	$x_{34}$	$x_{35}$	$x_{36}$	$x_{37}$	$x_{38}$	$x_{39}$	$x_{40}$	$x_{41}$	$x_{42}$	$x_{43}$	$x_{44}$	$x_{45}$	$x_{46}$	$x_{47}$	$x_{48}$				
<b>MC2E</b>																						
-1	3.3396	0.0691	-1.3186	0.0406	1.8592	0	-1.8414	0	-2.3071	2.1611	-4.8082	0.3672	0	0	-0.0501	0	0.0983	3.2667	3.086	3.3709	3.7544	0.5823
0	-1	0.0482	0	0	-0.4778	-1.8101	2.2952	0.0814	0	-0.3152	0	-0.0350	-0.0686	-0.0165	0.5501	-0.0455	0	0.0714	-0.0907	0	0	0
0.7484	0	-1	0.2321	44.7585	-48.8418	2.6354	0	0.9551	0	0	0	1.1895	0	-2.4406	0	-0.1163	0	-1.6660	0	-1.6660	0	0
-0.0645	0.9353	0	-1	0	2.8194	-3.0669	5.9859	-6.0091	0	0	0	0.9948	-0.0094	-0.031	0.0291	0.0222	0	-0.0592	0.3619	0.005	0.005	0.005
-1.3043	0	0	5.6852	-1	0	1.3937	0	-0.4330	-0.6722	0	0.4518	0.1682	0	0.0793	0	4.6436	0	-0.0729	-0.4425	0	0	0.4358
<b>Bondad ajuste</b>																						
-1	0.6638	-0.0384	1.1757	-0.2048	0.2328	0	0.1914	0	1.4128	-1.6921	5.8994	-0.9921	0	0	0.0381	0	0.2068	-2.7954	-2.8769	-2.7167	-3.0527	0.0902
0	-1	0.0229	0	0	-0.5037	-0.1071	0.6183	0.0704	0	-0.4146	0	0.0473	-0.042	-0.0018	0.5703	-0.0581	0	0.0735	-0.1282	-0.1282	0	0
-0.0728	0	-1	0.1040	47.4870	-51.5325	2.8865	0	0.7811	0	0	0	1.1808	0	-1.2341	0	-0.0267	0	-1.6004	0	-1.6004	0	0
-0.2923	0.4948	0	-1	0	2.2125	-2.5871	7.3016	-6.9645	0	0	0	0.6385	-0.0645	-0.0290	0.0188	0.0630	0	0.0044	0.1028	0.1028	-0.0964	0.0964
-0.7552	0	0	5.6657	-1	0	1.5796	0	-0.713	-0.6644	0	0.5031	0.4243	0	0.1180	0	4.2684	0	-0.1157	-0.4878	0	0	0.4155
<b>Precisión estimación</b>																						
-1	1.8888	0.0108	0.0339	-0.0924	0.9773	0	-0.7392	0	-0.2901	0.0719	0.9976	-0.5698	0	0	-0.0023	0	0.1571	-0.0202	-0.1471	0.0702	0.0635	0.3155
0	-1	-0.1561	0	0	-0.6871	11.9479	-11.2516	-0.0078	0	-1.1177	0	0.6298	0.1463	0.102	0.7136	-0.1477	0	0.0883	-0.3933	-0.3933	0	0
-1.2172	0	-1	0	-0.0745	51.289	-55.2819	3.2364	0	0.5385	0	0	1.1686	0	0.447	0.0981	0	0.0981	0	-1.5089	0	-1.5089	0
-0.2504	0.5758	0	-1	0	2.3242	-2.6753	7.0596	-6.7888	0	0	0	0.7040	-0.0544	-0.0294	0.0207	0.0555	0	-0.0073	0.1504	0.1504	-0.0778	0.0778
1.7564	0	0	5.5762	-1	0	2.4297	0	-1.9937	-0.6285	0	0.7379	1.5956	0	0.2953	0	2.5526	0	-0.3114	-0.6950	0	0	0.323
<b>Tradicional</b>																						
-1	1.8149	0.0078	0.1028	-0.0992	0.9324	0	-0.6831	0	-0.1874	-0.0346	1.2934	-0.4074	0	0	0.0002	0	0.1601	-0.1877	-0.3119	-0.0980	-0.1245	0.3019
0	-1	0.3566	0	0	-0.1617	-22.5822	22.7485	0.2161	0	0.8963	0	-1.0389	-0.3931	-0.1954	0.3032	0.1088	0	0.0459	0.366	0.366	0	0
-4.1736	0	-1	0	-0.5356	61.1108	-64.9678	4.1401	0	-0.0882	0	0	1.1371	0	0.4207	0	0.4207	0	-1.2727	0	-1.2727	0	0
-0.3395	0.4035	0	-1	0	2.0868	-2.4877	7.5742	-7.1624	0	0	0	0.5647	-0.0760	-0.0286	0.0167	0.0714	0	0.0175	0.0491	0.0491	-0.1174	0.1174
9.1864	0	0	5.3117	-1	0	4.9446	0	-5.7824	-0.5223	0	1.4323	5.0605	0	0.8196	0	-2.5231	0	-0.8902	-1.3080	-1.3080	0	0.0492
<b>Bmonopt</b>																						
-1	2.3465	0.0292	-0.3928	-0.0505	1.2555	0	-1.0869	0	-0.9264	0.7310	-0.8340	-0.1373	0	0	-0.0174	0	0.1386	1.0167	0.8728	1.1115	1.2279	0.3997
0	-1	0.1504	0	0	-0.3730	-8.6940	9.0734	0.1261	0	0.0863	0	-0.3677	-0.1762	-0.0758	0.4683	0.0057	0	0.0629	0.0606	0.0606	0	0
7.3720	0	-1	0	1.2651	22.7531	-27.1409	0.6105	0	2.3592	0	0	1.2600	0	-12.1710	0	-0.8390	0	-2.1952	0	-2.1952	0	0
-0.7113	-0.3155	0	-1	0	1.0963	-1.7045	9.7218	-8.7219	0	0	0	-0.0170	-0.1659	-0.0253	0	-0.0001	0.1379	0	0.1212	0.1212	-0.3738	-0.2829
-8.0808	0	0	5.9265	-1	0	-0.9000	0	3.0224	-0.7690	0	-0.1816	-2.9919	0	-0.3990	0	9.2728	0	0.4550	0.1165	0	0	0.6855
<b>Bdays2e</b>																						
-1	3.7479	0.0422	-0.7638	-0.0718	1.4248	0	-1.7237	0	-0.7177	0.4505	0.6625	0.5128	0	0	-0.0878	0	0.1985	0.6598	0.4883	0.8215	1.3942	0.5458
0	-1	0.0634	0	0	-0.3287	-0.3688	0.7397	0.0384	0	-0.2913	0	0.0129	-0.0523	-0.0117	0.7144	-0.0559	0	0.0532	-0.0593	-0.0593	0	0
1.2581	0	-1	0	0.2764	19.2300	-22.7722	2.0087	0	1.0449	0	0	0.9316	0	-3.7400	0	-0.1436	0	-1.4811	0	-1.4811	0	0
-0.1184	1.0933	0	-1	0	2.7657	-2.9816	4.0951	-4.1540	0	0	0	0.9266	-0.0043	-0.0282	0.0348	0.0092	0	-0.0801	0.3752	0.3752	0.0366	0.0366
-0.8638	0	0	7.3488	-1	0	2.2598	0	-1.4678	-0.6683	0	1.0606	-0.2758	0	0.1873	0	4.9044	0	-0.2521	-0.6489	0	0	0.6022
<b>MCMC</b>																						
-1	3.6119	0.0787	-1.4512	-0.5524	-386.0422	0	6.3128	0	-9.7605	5.1071	-5.2186	20.4997	0	0	-0.4884	0	-0.2409	14.5396	11.1873	11.7600	10.9321	-1301.3694
0	-1	-0.0002	0	0	-6.1420	0.4527	0.3196	0.1041	0	-0.5146	0	0.0990	-0.0492	0.0060	0.6777	0.0631	0	0.1305	-0.1498	-0.1498	0	0
0.0410	0	-1	0	-0.0079	16.8599	-54.9278	4.0948	0	0.6020	0	0	1.7582	0	-1.7230	0	-0.1679	0	-1.5894	0	-1.5894	0	0
0.2235	-0.1989	0	-1	0	-19.9744	-5.0350	7.9913	-5.4836	0	0	0	-0.1017	0.7805	-0.0227	0	-0.2410	0	-0.0525	0	-0.0525	0	8.6829
0.5617	0	0	-0.6056	-1	0	23.7927	0	4.4040	-3.3587	0	-0.9241	1.6889	0	2.1121	0	-2.7920	0	0.1692	-0.4017	-0.4017	0	46.0579

En cada ecuación, el valor -1 indica la variable dependiente, y el 0, que la variable no ha entrado en la ecuación.

Tabla 8.9: Valores de AIC, entropía, tiempo de estimación y parámetros  $k_1$  y  $k_2$  para el problema de hormonas.

Medida	MC2E	BMOM			$Bmom_{OPT}$	$Bayes_{2E}$	MCMC
		Ajuste	Estimación	Tradicional			
AIC	487,35	487,84	488,50	489,88	<b>486,42</b>	487,23	713,42
$H(e)$	9,64	9,54	9,75	9,62	<b>9,27</b>	9,79	12,75
$H_2(e)$	2,26	2,25	2,28	2,26	<b>2,23</b>	2,28	2,55
$M_{1/2}(e)$	16,18	16,07	16,19	16,03	<b>15,76</b>	16,42	18,97
Segundos	0,06	0,09	0,09	0,09	10,78	<b>0,02</b>	252,00
$k_1$	1,00	-7,50	-7,50	4,40	2,21		
$k_2$	1,00	1,00	-7,50	4,40	2,77		

Además, se han obtenido los valores de AIC, entropía, tiempo de ejecución y parámetros  $k_1$  y  $k_2$  para estos modelos, resultados que se muestran en la tabla 8.9. Se observa que el valor mínimo de  $H_2(e)$ , al igual que en el resto de entropías, se alcanza con  $Bmom_{OPT}$ .

De acuerdo al estudio de la variabilidad del capítulo 6 y a las conclusiones del estudio experimental del capítulo 5, el SEM estimado por el método  $Bmom_{OPT}$  puede considerarse como una primera aproximación a tener en cuenta por los especialistas en la materia. Por lo tanto, este modelo permite a los médicos, en base a datos fáciles de observar en un paciente, tener valores aproximados de sus niveles hormonales. Además, a mayor número de observaciones en el modelo y menor variabilidad de los datos, esta aproximación será más confiable.

### 8.3. Conclusiones

En este capítulo se han utilizado tanto el método de estimación propuesto como el parameter criteria para la selección del mejor método, ambos desarrollados en esta tesis en dos problemas diferentes con datos reales. El primero es un problema económico, el Modelo de Oferta de mano de obra de trabajadoras casadas, que se estima y calculan las medidas de entropía para, de entre el conjunto de métodos, elegir el que menor

error obtiene. En este caso se observa que el método que obtiene menor entropía y AIC es MC2E, mejorando por tanto a las estimaciones del nuevo método propuesto. En el segundo problema se propone un modelo para un conjunto de variables para las cuales no se había modelado ningún MES previamente, siendo por tanto la principal aportación de esta aplicación, proveer a los especialistas del tema un modelo y unas estimaciones con las que iniciar la investigación.

---

# CAPÍTULO 9

---

## Conclusiones y Trabajos futuros

---

En este último capítulo se resumen las principales conclusiones que se extraen del trabajo realizado con el fin de tener una visión general de todos los resultados obtenidos así como las vías futuras de investigación.

### 9.1. Conclusiones

En el proceso de investigación de los modelos de ecuaciones simultáneas y de los métodos de estimación se han considerado las metodologías frecuentista y bayesiana, y lejos de comparar cuál de estos enfoques es el mejor, en este trabajo se ha querido integrar ambas como opciones en la búsqueda de un método de estimación óptimo y con mejoras de los ya existentes. Siguiendo esta línea, se ha elegido el Método de Momentos Bayesiano, método ya estudiado previamente en el que se proponen valores a dos parámetros según se busque un menor error de estimación, mayor bondad de ajuste o una combinación entre ambos. Estos valores óptimos son funciones del número de variables endógenas, exógenas y del tamaño de la muestra. Partiendo de este método función de dos parámetros, en esta tesis se ha investigado y se propone un método que obtiene unos valores óptimos de los

parámetros que, según los estudios experimentales, minimizan el error como diferencia entre los coeficientes del modelo y sus estimaciones. Esto se ha llevado a cabo con la simulación de MES a través de un software generador de modelos aleatorios sujeto a varios parámetros como el número de variables endógenas, exógenas o la variabilidad de los datos y la búsqueda de los valores óptimos se ha llevado a cabo con una función de  elegida tras un estudio entre las diferentes opciones disponibles.

Una vez obtenido el método de estimación, se han investigado diferentes escenarios para determinar criterios en los que este método ofrece mejores resultados, lo cual ha motivado a seguir la investigación controlando los rangos de variabilidad y número de los datos así como el número de variables que forman el modelo.

De manera paralela, otro objetivo principal de esta investigación es el estudio y propuesta de un criterio de información indicador de la bondad de un método frente a otro. Para ello se ha investigado en el campo de las entropías, proponiendo nuevos criterios adaptados a los MES.

Se han llevado a cabo varios estudios experimentales para comparar el método optimizado propuesto con otros métodos. Para ello se han obtenido la norma euclídea de los coeficientes reales y los estimados, una variación del Criterio de información de Akaike que usa los errores reales entre las variables endógenas y las estimadas por los diferentes métodos, el AIC para MES y medidas basadas en entropía que permiten determinar la elección del mejor método en una situación no experimental. Estas medidas muestran que, el método propuesto en esta tesis es el que menor norma euclídea produce y que la medida de entropía estudiada puede ser una alternativa al AIC como criterio de selección del método de estimación de un MES.

Esta tesis está centrada en modelos lineales, pero se ha llevado a cabo una breve introducción a los MES No Lineales revisando la literatura y abriendo una nueva línea de trabajo con el uso de Redes Neuronales Artificiales. Los resultados, a pesar de llevar a cabo un pequeño estudio experimental, muestran que este tipo de aplicación puede aportar buenas soluciones a los MES cuando las relaciones no necesariamente sean lineales.

Y con el fin de aportar soluciones a problemas tanto de tipo social, económico o de sa-

---

lud, se exponen dos aplicaciones prácticas con datos reales. La primera, es un problema clásico en economía modelado a través de un MES de dos ecuaciones del que ya se tienen estimaciones previas en la literatura. La segunda aplicación, también con datos reales, es un problema original del campo de la medicina del que no consta ninguna solución previa como la que se aporta en esta tesis.

## 9.2. Trabajos futuros

Como líneas futuras de investigación y a la vista de los resultados obtenidos se proponen las siguientes, algunas de ellas relacionadas y otras complementarias con el trabajo realizado en esta tesis, para llevar a cabo a corto y medio plazo:

- Resolución de MES: tal y como se ha comprobado en el capítulo 5 donde se presenta un nuevo método de estimación, el método propuesto es una alternativa para estimar el modelo con menor error en base a una medida de entropía. Se plantea como línea a corto plazo la reducción del coste de ejecución en la obtención de los valores óptimos de los parámetros.
  - Criterios de información: en esta tesis se ha propuesto una medida alternativa para medir la calidad de un modelo y ha quedado en evidencia la necesidad de seguir investigando en este tipo de parameter criteria.
  - Influencia de la variabilidad de los datos en la bondad de los diferentes métodos de estimación: continuar la investigación ampliando el rango de las variabilidades de los datos.
  - Modelos de Ecuaciones Simultáneas No Lineales: ante la necesidad de proponer como solución MES no lineales y como línea de trabajo a corto plazo queda abierto el desarrollo de RNA para este tipo de modelos.
-



---

# Índice de tablas

---

2.1. Valores de los parámetros $K_1$ y $K_2$ propuestos para minimizar la función de pérdida. . . . .	15
5.1. Medidas de entropía como criterio de parámetro para la selección del método de estimación . . . . .	53
5.2. Media y desviación estándar de 50 simulaciones de $D_{\delta, \hat{\delta}}$ . Sigma 0,1. . . .	59
5.3. Media y desviación estándar de 50 simulaciones de $AIC_{real}$ . Sigma 0,1. . .	59
5.4. Media y desviación estándar de 50 simulaciones de $AIC$ . Sigma 0,1. . . .	59
5.5. Media y desviación estándar de 50 simulaciones de $AIC_{MC2E}$ . Sigma 0,1.	60
5.6. Media y desviación estándar de 50 simulaciones de $H(e)$ . Sigma 0,1. . . .	60
5.7. Media y desviación estándar de 50 simulaciones de $H_2(e)$ . Sigma 0,1. . .	60
5.8. Media y desviación estándar de 50 simulaciones de $M_{1/2}(e)$ . Sigma 0,1. .	61
5.9. Media y desviación estándar de 50 simulaciones del Tiempo de ejecución. Sigma 0,1. . . . .	61
6.1. Media y desviación típica de la norma, AIC, entropía y tiempo de ejecución de 30 simulaciones (I). Desviación media total de las variables endógenas 2,133. . . . .	66

6.2. Media y desviación típica de la norma, AIC, entropía y tiempo de ejecución de 30 simulaciones (II). Desviación media total de las variables endógenas 2,921. . . . .	67
6.3. Media y desviación típica de la norma, AIC, entropía y tiempo de ejecución de 30 simulaciones (III). Desviación media total de las variables endógenas 4,191. . . . .	68
6.4. Media y desviación típica de la norma, AIC, entropía y tiempo de ejecución de 30 simulaciones (IV). Desviación media total de las variables endógenas 11,840. . . . .	69
6.5. Métodos con valores mínimos en norma, AIC, entropía y tiempo en segundos, según variabilidad de los datos. 30 iteraciones. . . . .	71
6.6. Porcentaje de coincidencia entre $D_{\delta, \hat{\delta}}$ y $H_2(e)$ . . . . .	71
6.7. Porcentaje de MES con $Bmom_{OPT}$ mejor método de estimación y es seleccionado por $H_2(e)$ . . . . .	72
7.1. Datos generados para la resolución de un MES mediante RNA: caso 1. . . . .	76
7.2. Predicciones y ECM con la RNA . . . . .	78
7.3. Predicciones y ECM con $Bmom_{OPT}$ . . . . .	78
7.4. Error cuadrático medio para cada output de la RNA . . . . .	81
7.5. Error cuadrático medio de la RNA y método $Bmom_{OPT}$ . . . . .	81
8.1. Media, DT y CV de las variables endógenas del modelo de la Oferta de mano de obra de trabajadoras casadas. . . . .	85
8.2. Coeficientes estimados del MES de la Oferta de mano de obra de trabajadoras casadas. . . . .	86
8.3. AIC, entropías, tiempo de ejecución y $k_1$ y $k_2$ para el modelo de Oferta de mano de obra de trabajadoras casadas. . . . .	87

---

8.4. AIC, entropías y tiempo de ejecución obtenidos con $Bmom_{OPT}$ con diferentes valores de $k_1$ y $k_2$ para el modelo de Oferta de mano de obra de trabajadoras casadas. . . . .	87
8.5. Variables utilizadas en el modelo del problema de hormonas . . . . .	89
8.6. MES propuesto para el problema de hormonas. . . . .	90
8.7. Datos utilizados para estimar el MES del problema de hormonas. . . . .	91
8.8. Coeficientes estimados del MES del problema de hormonas. . . . .	92
8.9. Valores de AIC, entropía, tiempo de estimación y parámetros $k_1$ y $k_2$ para el problema de hormonas. . . . .	93

---



---

# Índice de figuras

---

4.1. Inteligencia Artificial, Machine Learning y Deep Learning . . . . .	38
4.2. Similitudes entre una neurona biológica y una artificial . . . . .	39
4.3. Ejemplo de estructura típica de una red neuronal artificial . . . . .	40
7.1. Función de activación logística . . . . .	77
7.2. Red neuronal artificial con 4 inputs, 2 outputs y sin capa oculta. . . . .	77
7.3. Comparativa de los valores $Y_1$ frente a $\hat{Y}_1$ obtenidos por $Bmom_{OPT}$ y la RNA	79
7.4. Comparativa de los valores $Y_2$ frente a $\hat{Y}_2$ obtenidos por $Bmom_{OPT}$ y la RNA	79



---

## Siglas y acrónimos

---

MES	→	Modelo de ecuaciones simultáneas
MCO	→	Mínimos cuadrados ordinarios
MC2E	→	Mínimos cuadrados en dos etapas
MC3E	→	Mínimos cuadrados en tres etapas
MCI	→	Mínimos cuadrados indirectos
MVIL	→	Máxima verosimilitud con información limitada
MVIC	→	Máxima verosimilitud con información completa
MCONL	→	Mínimos cuadrados ordinarios no lineales
MC2ENL	→	Mínimos cuadrados en dos etapas no lineal
MC3ENL	→	Mínimos cuadrados en tres etapas no lineal
MVICNL	→	Máxima verosimilitud con información completa no lineal
TRUNFIML	→	Máxima verosimilitud con información completa no lineal computacionalmente eficiente
$Bayes_{2E}$	→	Bayesiano en dos etapas
BMOM	→	Método de momentos Bayesiano
$Bmom_{OPT}$	→	Método de momentos Bayesiano Optimizado
MCMC	→	Markov Chain Monte Carlo
NIG	→	Distribución Normal Inversa Gamma
AIC	→	Criterio de información de Akaike
BIC	→	Criterio de información de Bayesiano
$H(e)$	→	Entropía de Shannon
$H_2(e)$	→	Medida de Entropía
$M_{1/2}(e)$	→	Medida de Entropía
$D_{\delta, \hat{\delta}}$	→	Norma Euclídea

DT	→	Desviación típica
CV	→	Coficiente de variación
ECM	→	Error cuadrático medio
IA	→	Inteligencia Artificial
ML	→	Machine Learning
DL	→	Deep Learning
RNA	→	Red neuronal artificial
RNBR	→	Red neuronal de base radial
RNN	→	Red neuronal recurrente

---

---

# Anexo I

---

Se adjunta en este anexo el artículo publicado en *Mathematics* en marzo de 2021 sobre el que se ha basado y desarrollado el capítulo 5.



Article

# A New Computational Method for Estimating Simultaneous Equations Models Using Entropy as a Parameter Criteria

Belén Pérez-Sánchez <sup>1,\*</sup>, Martín González <sup>2</sup>, Carmen Perea <sup>2</sup> and Jose J. López-Espín <sup>2</sup><sup>1</sup> Department of Statistics, Mathematics and Informatics. Miguel Hernández University, 03202 Elche, Spain<sup>2</sup> Center of Operations Research, Miguel Hernández University, 03202 Elche, Spain; mgonzaleze@umh.es (M.G.); perea@umh.es (C.P.); jlopez@umh.es (J.J.L.-E.)

\* Correspondence: m.perezs@umh.es

**Abstract:** Simultaneous Equations Models (SEM) is a statistical technique widely used in economic science to model the simultaneity relationship between variables. In the past years, this technique has also been used in other fields such as psychology or medicine. Thus, the development of new estimating methods is an important line of research. In fact, if we want to apply the SEM to medical problems with the main goal being to obtain the best approximation between the parameters of model and their estimations. This paper shows a computational study between different methods for estimating simultaneous equations models as well as a new method which allows the estimation of those parameters based on the optimization of the Bayesian Method of Moments and minimizing the Akaike Information Criteria. In addition, an entropy measure has been calculated as a parameter criteria to compare the estimation methods studied. The comparison between those methods is performed through an experimental study using randomly generated models. The experimental study compares the estimations obtained by the different methods as well as the efficiency when comparing solutions by Akaike Information Criteria and Entropy Measure. The study shows that the proposed estimation method offered better approximations and the entropy measured results more efficiently than the rest.

**Keywords:** simultaneous equations models; bayesian method of moments; markov chain monte carlo; akaike information criteria; entropy; computational statistics



**Citation:** Pérez-Sánchez, B.; González, M.; Perea, C.; López-Espín, J.J. A New Computational Method for Estimating Simultaneous Equations Models Using Entropy as a Parameter Criteria. *Mathematics* **2021**, *9*, 700. <https://doi.org/10.3390/math9070700>

Academic Editor: Tzong-Ru Tsai

Received: 26 February 2021

Accepted: 22 March 2021

Published: 24 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Simultaneous Equations Models (SEM) [1] is statistical model formed by a set of regression equations that reflect the simultaneity between the set of dependent and independent variables of the model. SEM is used when there is a bidirectional influence relationship between both types of variables. The estimation of the coefficients of a SEM can be made by methods based on either the classical statistical approach or the Bayesian approach.

Regarding classic inference, the estimation of a SEM can be made by limited information and full information methods. Limited information methods estimate each of the equations of the structural form [1] without making use of the information contained in the detailed specification of the rest of the model, only considering both the endogenous and exogenous variables that are included in this equation. Ordinary Least Squares (OLS), Indirect Least Squares (ILS), and the Two Stage Least Squares (2SLS) are examples of limited information methods [1]. Full information methods consider joint estimation of the whole model in the structural form. These methods require the specification all equations, and all of them have to be identified. In general, they are more asymptotically efficient than the others since they incorporate all the information of the system, but, with the drawback that if any equation is incorrectly specified, estimates that are inconsistent with the other equations may be generated. Examples of these kinds of methods are Full Information Maximum Likelihood (FIML) or Three Stage Least Squares (3SLS) [1].

On the other hand, the Bayesian inference which is based on a given set of data, does not use sampling assumptions, but introduces a high degree of complexity due to the prior specification of the distribution, as well as the obtaining of the posterior distribution. Some estimation techniques in the Bayesian approach are the Bayesian Method of Moments (BMOM) [2] or the methods used by Chao and Phillips [3], Geweke [4], and Kleibergen and Van Dijk [5]. The recent development of the Markov Chain Monte Carlo method has been key in making the computation of large models that require integrations over hundreds or even thousands of unknown parameters possible. The Metropolis–Hastings algorithm and the Gibbs Sampling [6] are examples of them.

Regarding model selection, the literature is limited to a comparison between Bayesian and classical estimators, concluding that Bayesian methods perform better in the case of a small sample [2].

Applications of SEMs can be found mainly in the economic framework, although there are some applications of SEMs in other fields. For instance, in tax research, the effects of fiscal decentralization on regional income inequality in Indonesia have been studied using provincial-level data over the period 2001–2014 [7], or finding a SEM that relates employment to mental health [8], others have studied the impact of foreign trade on energy efficiency in China’s textile industry [9] or biomass energy consumption, economic growth, and carbon emission in West Africa [10]. And other studies have concentrated on comparing the results obtained by SEMs versus linear regression modeling of complex phenotypes [11], a model for knowing the forecast demand on the facilities provided airside at airports [12] or a SEM for modeling prescriptions in primary care [13]. The Bayesian SEM approach has been used in agricultural science [14], or for studying the impact of product information on third-party websites on the feedback mechanism between internal word-of-mouth and retail sales on [Download.com](#) and [Amazon.com](#) (accessed on 1 March 2020) [15]. It can be useful for analyzing the interdependence of a television program viewership between spouses [16], for exploring peer effects in casino gambling behavior [17], for modeling the interaction between people’s health risk perception and betel chewing habits in Taiwan [18], and for studying the effects of repetitive iodine thyroid blocking on the development of the foetal brain and thyroid in rats [19].

When several estimate models are available, it is necessary to have a selection parameter criteria. There are a lot of useful information parameter criteria for comparing SEMs, such as Akaike Information Criteria (AIC) [20,21], its corrected version (AICc) [22], Schwarz Information Criteria (SIC) [23], Bayesian Information Criteria (BIC) [24], Hannan and Quinn (HQ) [25], and Model Selection Criterion based on Kullback–Leibler’s Symmetric Divergence [26].

Entropy was initially introduced in thermodynamics, where it was used to provide the basis for the second law of thermodynamics. Subsequently, mechanical statistics provided a connection between the macroscopic properties of entropy and the states of the system, and from a mathematical point of view, are non-negative functions defined in probability distributions with multiple applications such as using information theory for measuring a system stability [27]. Applications of entropy have been used in the fields of finance, [28], environmental and water engineering [29], urban systems [30], and applications for customer satisfaction surveys [31].

In our paper, a new method for the SEM estimation is developed and compared with other methods through the AIC and an entropy measure developed by Amigó [32], which allows us to select the estimation method with the highest homogeneity in the estimation errors.

The organization of the paper is as follows: In Section 2, the model is set up and several classical and Bayesian methods for estimating SEM are briefly reviewed. Section 3 describes the proposed estimation method. In Section 4, the entropy is shown and a new version is obtained as information criteria for selecting the method of estimation with minimum error. The experimental design and the results are shown in Section 5, and, finally, the conclusions and future lines of work are presented in Section 6.

## 2. Definition of the Model and Methods for Estimating a SEM Problem

### 2.1. Definition of the Model

Consider  $m$  interdependent or endogenous variables which depend on  $k$  independent or exogenous variables. Suppose that each endogenous variable can be expressed as a linear combination of the other endogenous variables, the exogenous ones, and white noise that represents stochastic interference. Thus, a SEM as a matrix form is [1]:

$$\begin{aligned}
 y_1 &= B_{1,2}y_2 + B_{1,3}y_3 + \dots + B_{1,m}y_m + \Gamma_{1,1}x_1 + \dots + \Gamma_{1,k}x_k + u_1 \\
 y_2 &= B_{2,1}y_1 + B_{2,3}y_3 + \dots + B_{2,m}y_m + \Gamma_{2,1}x_1 + \dots + \Gamma_{2,k}x_k + u_2 \\
 &\vdots \\
 y_m &= B_{m,1}y_1 + B_{m,2}y_2 + \dots + B_{m,m-1}y_{m-1} + \Gamma_{k,1}x_1 + \dots + \Gamma_{m,k}x_k + u_m.
 \end{aligned} \tag{1}$$

The equations can be represented in matrix form as:

$$YB^T + X\Gamma^T + U = 0 \tag{2}$$

where  $B \in \mathbb{R}^{m \times m}$  and  $\Gamma \in \mathbb{R}^{m \times k}$  are matrices of coefficients,  $Y \in \mathbb{R}^{n \times m}$  is the matrix of endogenous variables,  $X \in \mathbb{R}^{n \times k}$  is the matrix of exogenous variables, and  $U \in \mathbb{R}^{n \times m}$  is the matrix of white noise variables, being  $n$  the sample size. Some coefficients of  $B$  and  $\Gamma$  are zero, and are known a priori. The number of endogenous and exogenous variables in the  $i$ th equation of (1) is denoted by  $m_i$  and  $k_i$ . An equation is identified if the number of variables (endogenous and exogenous) in the equation is lower than or equal to  $k + 1$ , that is  $m_i - 1 \leq k - k_i$  (order condition (1)). When  $m_i - 1 = k - k_i$ , the equation is exactly identified and when  $m_i - 1 < k - k_i$  it is over-identified. Only identified equations can be solved. Solving the model is equivalent to obtaining a estimation of  $B$  and  $\Gamma$  in (2) from a representative sample of the model (a set of values of the data variables  $X$  and  $Y$ ), in order to explain a well-known matrix equation that represents the relationship between both sets of variables.

### 2.2. Methods for Estimating an SEM Problem

There are different techniques for estimating SEM parameters. In the classical approach, examples of these are 3SLS, 2SLS, OLS, K-class estimators [33], etc., and in the Bayesian approach, MCMC algorithms, and several conjugate distributions are some examples. In this section, a brief review of those estimators that have been used in our work is shown.

#### 2.2.1. Two Stage Least Squares (2SLS)

2SLS is the most common estimation method for a SEM [1], developed independently by Theil (1953) [34] and Basmann (1957) [35]. This method is called two stage, because in its mathematical expression, Ordinary Least Square (OLS) is applied twice. In the first step, new variables, called proxy, are calculated by Least Squares using the exogenous variables of the model as independent variables, and in the second step, the endogenous variables are substituted by the *proxys*, and then Least Squares is applied once more. Both, the 2SLS and OLS method are single K-class estimators [33], expressed in (3) when  $K_1 = K_2 = 0$  (OLS), and when  $K_1 = K_2 = 1$  (2SLS).

#### 2.2.2. Bayesian Method of Moments (BMOM)

This method, proposed by Zellner in 1998 [2], applies the principle of maximum entropy and generates optimal estimation evaluated by double K-class estimators shown in Table 1. When there is not enough information available to obtain the likelihood function, allowing for data analysis without specifying a probability function and sampling assumption.

Considering, for example, the first structural equation,  $y_1 = Y_1\beta_1 + X_1\gamma_1 + u_1$ , being  $Y_1 \in \mathbb{R}^{n \times m}$  and  $X_1 \in \mathbb{R}^{n \times k}$  the matrix of endogenous and exogenous variables and

$u_1 \in \mathbb{R}^{n \times 1}$ , the white noise vector. The parameters  $\delta_1 = (\beta_1 \ \gamma_1)' \in \mathbb{R}^{1 \times (m+k)}$ , are estimated by BMOM, that minimize the loss functions are given by:

$$\hat{\delta}_1(K_1, K_2) = \begin{bmatrix} Y_1'Y_1 - K_1\hat{V}_1'\hat{V}_1 & Y_1'Z_1 \\ Z_1'Y_1 & Z_1'Z_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_1 - K_2\hat{V}_1)' y_1 \\ Z_1'y_1 \end{bmatrix} \tag{3}$$

with  $Z_1 = (X\Pi_1 \ X_1)'$ , where  $\Pi_1$  is the coefficient matrix of the reduced form equation for  $Y_1$ , being  $Y_1 = X_1\Pi_1 + V_1$ . Those coefficients can be calculated by LS obtaining  $\hat{\Pi}_1 = (X'X)^{-1}X'Y_1$  and  $\hat{V}_1 = Y_1 - X\hat{\Pi}_1$ .

Table 1 shows the  $K_1$  and  $K_2$  parameters for two loss functions using the BMOM approach, being  $n$  and  $k$ , the sample size and the number of exogenous variables:

**Table 1.** K1 and K2 parameters proposed to minimize loss function. Bayesian Method of Moments (BMOM).

Loss Function		BMOM Approach
Goodness of fit	$L_g = (y_1 - Z_1\hat{\delta}_1)'(y_1 - Z_1\hat{\delta}_1)$	$K_1 = 1 - k/(n - k), K_2 = 1$
Precision of estimation	$L_p = (\delta_1 - \hat{\delta}_1)'Z_1'Z_1(\delta_1 - \hat{\delta}_1)$	$K_1 = K_2 = 1 - k/(n - k)$

### 2.2.3. Bayesian Approach in Two Stages (*Bayes<sub>2S</sub>*)

In Bayesian inference, a pragmatic solution for choosing the prior distribution is to select a member of the distribution family so that it remains in the same family as the prior distribution. If the prior is conjugate, the posterior distribution after the first observation belongs, by definition, to the same type and is used as the new prior distribution in the next observation. Incorporating this second observation, the new posterior distribution, also belongs to the class of conjugation. This sequential process only updates the value of parameters of the distribution [6]. In this work, the *Bayes<sub>2S</sub>* method uses a Normal-Inverse Gamma prior to obtaining the exact analytic expressions for the posterior distribution of the structural  $B$  and  $\Gamma$  coefficients of the SEM. Basically, it has been applied in two stages as the 2SLS method, used in the two-steps Bayesian Least Squares instead of Ordinary Least Squares.

### 2.2.4. Markov Chain Monte Carlo (MCMC)

In the Bayesian approach, it is essential to select prior distribution, however, there are situations where this selection is somewhat difficult due to the absence of previous model information. The MCMC methodology provides a wide scope for statistical modeling, and is widely used to summarize complicated posterior distributions in econometrics models. In particular, Bayesian methods need to integrate the posterior distribution of model parameters for that reason and MCMC draws samples from these posterior distributions. There are many ways for constructing these chains like Gibbs sampler or special cases of the general framework of Metropolis and Hastings [6]. In particular, in this work, Gibbs Sampling has been used for the simulation of posterior distribution, calculating the average to estimate the model parameters.

### 3. The Proposed Estimation Method: Optimized BMOM Method (*Bmom<sub>OPT</sub>*)

A variation of BMOM (Section 2.2.2) is proposed obtaining  $K_1$  and  $K_2$  by the optimization of different parameter criteria, instead of setting them by the proposed values in (1). Concretely, the optimization of  $K_1$  and  $K_2$  parameters that minimize the AIC it is proposed, which is a quality measure of statistical models [20] based on sample fit to estimate the likelihood of a model. Thus, given a collection of model-based estimates for data, AIC obtains the quality of each model with respect to other models, providing a way for a model selection.

The expression for this measure is:

$$AIC = n \ln |\hat{\Sigma}_e| + 2 \sum_{i=1}^m (m_i + k_i - 1) + m(m + 1) \tag{4}$$

where  $n$  is the sample size,  $m$  is the number of equations,  $m_i$  and  $k_i$  are the number of endogenous and exogenous variables in  $i$ -equation and  $\hat{\Sigma}_e$ , the variance-covariance matrix of the errors  $e_j = Y_j - \hat{Y}_j, j=1, \dots, m$ .

For the experimental study, a large number of SEMs models have been randomly generated through a model generator tool, called real models. Thus, new errors, denoted by  $e_j^r$ , have been obtained as the difference between the values generated by the real models,  $Y_j^r$ , and the estimated values by each method described in the previous section,  $\hat{Y}_j$ . These errors are substituted in (4), and, a new measure, denoted  $AIC_{real}$  is calculated. This value is a measure of the error and we propose it as a reference parameter, which could be an indicator of the goodness of the estimated model. It can only be calculated if the real coefficients are known, that is, in an experimental study. To reach the minimum  $AIC_{real}$  value, an algorithm based on the Quasi-Newton method is used, so that, in each iteration the algorithm looks for an approximation of the inverse of the Hessian matrix. So, the algorithm guarantees approximation at every step of the process.

#### 4. Entropy as an Information Parameter Criteria

Another motivation in this work is to obtain an alternative measure to AIC which is closely related to the BIC. In this section, it presents a parameter criteria of a SEM quality based on entropy,  $H(e)$ . Although entropy was introduced by Clausius [36] in thermodynamics to measure the amount of energy in a system that cannot produce work, this concept appears in many contexts (statistical mechanics, information theory, etc.) as disorder, uncertainty, randomness, complexity, etc. Claude Shannon [37] in 1948 built his theory of information and communication, being generalized by some authors as Tsallis [38]. The expression for Shannon entropy [39],  $H_S(x)$ , is:

$$H_S(x) = - \sum_{i=1}^n p(x) \log p(x). \tag{5}$$

On the other hand, a new entropy measure, developed by Amigó [32] as a variation of generalized entropy, has been developed to allow having small  $p(x)$  values in the distribution. The expression for this entropy,  $H_A(p)$ , is:

$$H_A(p) = \prod_{i=1}^n (2 - (p_i)^{p_i}). \tag{6}$$

The model SEM has  $m$  equations, the same number as endogenous variables. Then, applying this entropy to each equation in the estimated model obtains:

$$H_j(e) = \prod_{i=1}^n (2 - (p_{ij})^{p_{ij}}) \quad j = 1, 2, \dots, m \tag{7}$$

where  $n$  is the sample size and  $m$  the number of endogenous variables, and the  $p_{ij}$  values for each endogenous variable have been obtained as follows:

$$p_{ij} = \frac{e_{ij}}{\sum_{i=i}^n e_{ij}} \quad j = 1, 2, \dots, m, \quad \text{where } e_{ij} = Y_{ij} - \hat{Y}_{ij} \tag{8}$$

where  $p_{ij}$  is the error mass in each endogenous variable and has been calculated from the error matrix as the difference between endogenous variables and its estimation through

each method. Finally, for each method and estimated model, the average of logarithms of (7) is calculated:

$$H(e) = \frac{\sum_{j=1}^m \ln H_j(e)}{m}. \tag{9}$$

The minimum value of  $H(e)$  is reached when  $e_{ij}$  are homogeneous, so, the  $\hat{Y}_{ij}$  values are more well-balanced.

### 5. Experimental Design and Results

#### 5.1. Experimental Design

In the experimental study a large number of SEMs are generated (that is, generating the matrices  $X, Y, B$  y  $\Gamma$  of each model) and then are estimated through the methods presented in Sections 2 and 3 and finally, the models are compared to their estimations.

The SEMs have been generated as follows: The values of matrices  $B$  and  $\Gamma$  are generated randomly following a Uniform distribution in  $[0, 10]$ , matrix  $X$  a multivariate normal distribution, and finally, matrix  $Y$  as  $\Pi X$  plus a Normal distribution with mean 0 and sigma 0.1. Two functions of the R package have been used: In MCMC, Markov chains have been simulated through the *MCMCregress* function of *MCMCpack* package without prior information, and, in *BmomOPT* the function *optim* has been used to obtain the optimal values  $K_1$  and  $K_2$ . In *Bayes2S*, a Normal-Inverse Gamma prior has been used, with an average of 5 and precision of 0.2 as the initial parameters.

For the comparison study, measures based on the generated model parameters and measures based on the estimated parameters have been calculated, in order to have the criteria to find the best estimation method. So, it is possible to have criteria to choose the best estimation method. In the first type, the Euclidean distance between  $\delta$  and  $\hat{\delta}$ , denoted by  $D_{\delta, \hat{\delta}}$ , where  $\delta = [B \ \Gamma]$  is the coefficient matrix and  $\hat{\delta}$  its estimation, and  $AIC_{real}$ . In the second type,  $AIC$  and the entropy  $H(e)$  have been calculated. And finally, the execution time has been calculated. Table 2 shows the average and the standard deviation of 50 simulations for each measure, model, and method when the number of variables and sample size of SEMs are varying.

**Table 2.** Average and standard deviation of 50 simulations of  $D_{\delta, \hat{\delta}}$ ,  $AIC_{real}$ ,  $AIC$ ,  $H(e)$ , and execution time in seconds. Markov Chain Monte Carlo (MCMC). Sigma 0.1.

BMOM									
	m	k	n	2SLS	BMOM		<i>BmomOPT</i>	<i>Bayes2S</i> <sup>a</sup>	MCMC <sup>b</sup>
					Goodness of Fit	Precision of Estimation			
$D_{\delta, \hat{\delta}}$	10	20	100	27.670 7.778	40.914 7.538	40.966 7.546	<b>20.647</b> 8.826	33.673 12.084	71.410 10.491
	10	40	100	40.927 9.104	58.635 8.039	58.932 8.100	<b>26.769</b> <b>8.209</b>	56.340 13.721	91.076 4.918
	20	60	100	115.852 8.537	141.029 6.294	140.906 6.250	<b>94.999</b> 12.498	146.640 5.904	163.771 6.092
	10	20	400	16.563 11.257	27.508 6.955	27.534 6.974	<b>10.619</b> 5.868	22.233 17.026	70.449 7.899
	10	40	400	15.199 4.366	30.538 6.467	30.494 6.446	<b>7.923</b> 2.576	26.009 22.856	90.357 5.568
	10	40	1000	7.394 2.830	17.218 5.450	17.229 5.458	<b>5.130</b> 1.944	9.233 7.540	95.210 10.424

Table 2. Cont.

BMOM									
	m	k	n	2SLS	Goodness of Fit	Precision of Estimation	$Bmom_{OPT}$	$Bayes_{2S}^a$	MCMC <sup>b</sup>
AIC <sub>real</sub>	10	20	100	1361.764 781.683	1895.361 801.349	1896.600 800.971	<b>1156.465</b> 792.277	1993.402 831.727	4598.074 395.783
	10	40	100	1915.270 651.931	2345.388 636.213	2351.286 635.582	<b>1718.000</b> 626.334	2531.052 715.689	4547.361 286.781
	20	60	100	5941.844 779.559	6458.904 762.773	6456.303 763.063	<b>5692.839</b> 793.354	6732.681 777.638	10,172.018 206.247
	10	20	400	4438.187 3271.180	7319.586 3449.337	7323.685 3450.560	<b>3409.573</b> 3232.656	7517.898 5184.224	22875.120 1641.293
	10	40	400	4645.919 2358.815	6989.900 2744.681	6983.853 2744.491	<b>3877.930</b> 2299.173	7382.701 4943.642	22,057.334 917.641
	10	40	1000	6824.562 8001.673	11,738.582 8294.696	11,742.499 8294.949	<b>5913.439</b> 8123.421	9046.596 8850.159	63,247.410 2586.679
AIC	10	20	100	2168.030 854.334	1784.122 808.851	<b>1783.602</b> 808.150	2419.415 893.327	2391.887 785.620	4413.688 421.952
	10	40	100	2009.636 639.153	1850.753 626.141	<b>1850.716</b> 626.059	2372.355 692.930	2254.874 728.737	4348.852 256.390
	20	60	100	3866.102 1051.585	3647.221 1056.690	<b>3645.439</b> 1056.018	4543.780 1054.504	4119.416 1128.524	9856.712 241.454
	10	20	400	15,027.648 3347.331	13,448.372 3161.253	<b>13,446.459</b> 3161.039	15,524.231 3459.419	15626.142 3279.198	22,160.759 1516.667
	10	40	400	12,849.587 2699.761	<b>11,990.079</b> 2509.827	11,991.659 2509.734	13,479.303 2744.344	13,606.665 2961.619	21,421.728 858.468
	10	40	1000	37,879.770 9621.598	36,438.251 9459.390	<b>36,437.331</b> 9459.167	38,479.010 9605.284	38,035.942 9483.024	61,720.600 2138.345
H(e)	10	20	100	4.074 0.013	4.081 0.012	4.081 0.012	<b>4.074</b> 0.013	4.084 0.018	4.096 0.016
	10	40	100	4.076 0.012	4.080 0.010	4.080 0.010	<b>4.075</b> 0.014	4.084 0.011	4.087 0.014
	20	60	100	<b>4.086</b> 0.008	4.087 0.009	4.087 0.009	<b>4.086</b> 0.009	<b>4.086</b> 0.009	4.088 0.009
	10	20	400	<b>5.579</b> 0.005	<b>5.579</b> 0.005	5.579 0.005	<b>5.579</b> 0.005	5.587 0.021	5.609 0.008
	10	40	400	<b>5.590</b> 0.005	<b>5.590</b> 0.005	<b>5.590</b> 0.005	<b>5.590</b> 0.005	5.593 0.014	5.602 0.007
	10	40	1000	6.530 0.003	6.531 0.003	6.531 0.003	<b>6.530</b> 0.004	6.532 0.013	6.568 0.004
Time (s)	10	20	100	0.073 0.145	0.732 0.246	0.732 0.246	258.227 98.562	<b>0.056</b> 0.017	294.453 17.949
	10	40	100	0.143 0.047	1.022 0.395	1.022 0.395	274.088 345.929	<b>0.140</b> 0.050	499.554 68.116
	20	60	100	<b>0.314</b> 0.047	2.495 0.395	2.495 0.395	748.504 345.929	0.407 0.145	1435.791 0.145
	10	20	400	0.125 0.031	4.265 0.864	4.265 0.864	2586.186 1005.068	<b>0.109</b> 0.030	328.571 18.174
	10	40	400	0.235 0.037	4.533 0.765	4.533 0.765	2281.255 804.186	<b>0.214</b> 0.032	507.791 38.694
	10	40	1000	0.426 0.066	21.385 1.595	21.385 1.595	14,534.080 21,689.539	<b>0.376</b> 0.107	524.904 26.564

<sup>a</sup> All initial parameters have average 5.0 and precision 0.2. <sup>b</sup> Without prior information. Chain size 10,000.

### 5.2. Experimental Results

Regarding  $D_{\delta, \hat{\delta}}$ , the results shows that the  $Bmom_{OPT}$  method is better in all estimated models. Nevertheless, the high computational cost used by this method could be an

issue in large problems. In such cases it would be more efficient to use *Bayes2S*, since this method obtains good estimation in less time.  $AIC_{real}$  shows similar results than  $D_{\delta,\hat{\delta}}$ . Regarding estimated measures, *AIC* does not offer satisfactory results because its values are not in agreement with the Euclidean distance or with  $AIC_{real}$ , which is proposed as a reformulation that works well. The entropy  $H(e)$  has yielded satisfactory results and it could be considered as a new comparative measurement.

In all methods, when the complexity of model increases,  $D_{\delta,\hat{\delta}}$  increases, being the opposite when the sample size increases. Both BMOMs methods (Goodness of fit and Precision of estimation) show similar results, having a small difference between them. MCMC provides estimations with the biggest average of  $D_{\delta,\hat{\delta}}$ , in which no trend is appreciated when the number of variables and sample size vary. The minimum average of  $AIC_{real}$  is reached by  $Bmom_{OPT}$ , with MCMC being the worst method. Regarding the entropy, the minimum average value has been obtained by  $Bmom_{OPT}$ , except in the smallest case, where *Bayes2S* obtains the minimum value with a small difference. *Bayes2S* and 2SLS require less execution time in all cases, with  $Bmom_{OPT}$  performing the worst.

## 6. Conclusions and Future Work

In this paper, the estimation of simultaneous equations models was studied through the comparison of models, carried out through an experimental study using randomly generated models. A new estimation method was proposed,  $Bmom_{OPT}$ , based on the optimization of some parameter of the Bayesian Method of Moments and minimizing the Akaike Information Criteria. The computational study showed that the proposed method was the best one regarding the minimum  $D_{\delta,\hat{\delta}}$  and entropy. The study also showed that the *AIC* parameter presented deficiencies for selecting the estimation method with a minimum  $D_{\delta,\hat{\delta}}$  value and minimum  $AIC_{real}$ .

The *AIC* is one of the most used parameter criteria to compare different methods of estimation. Nevertheless, in this study, the results showed that using entropy instead of *AIC* in the evaluation of the methods provides values according with the quality of the estimation (similarity with the real value).

In future, the study of information criteria parameters and their application in SEM problems, the study of use other criteria for optimization, as well as studying how to reduce the execution costs can be considered.

**Author Contributions:** Conceptualization, J.J.L.-E. and C.P.; software, M.G.; validation, B.P.-S., J.J.L.-E. and C.P.; formal analysis, B.P.-S. and C.P.; writing—original draft preparation, B.P.-S.; writing—review and editing, J.J.L.-E. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Spanish Ministry for Economy and Competitiveness (Ministerio de Economía, Industria y Competitividad) under grant TIN2016-80565-R.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data can be solicited to author.

**Acknowledgments:** The authors are grateful for the computer resources and assistance provided by the Scientific Computing and Parallel Programming Group of the University of Murcia for the simulation study. Constructive comments from the referees to improve the presentation of the paper are great appreciated.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Gujarati, D.; Porter, D. *Econometría*, 5th ed.; México, D.F., Ed.; McGraw-Hill Interamericana Editores SA: Ciudad de México, México, 2004.

2. Zellner, A. The finite sample properties of simultaneous equations' estimates and estimators Bayesian and non-Bayesian approaches. *J. Econom.* **1998**, *83*, 185–212. [[CrossRef](#)]
3. Chao, J.C.; Phillips, P.C. Jeffreys prior analysis of the Simultaneous Equations Model in the case with  $n+1$  endogenous variables. *J. Econom.* **2002**, *111*, 251–283. [[CrossRef](#)]
4. Geweke, J. Bayesian Reduced Rank Regression in Econometrics. *J. Econ. Trics* **1996**, *75*, 121–146. [[CrossRef](#)]
5. Kleibergen, F.; Dijk, H.V. Bayesian simultaneous equation analysis using reduced rank structures. *Econom. Theory* **1998**, *14*, 701–743. [[CrossRef](#)]
6. Gelman, A.; Carlin, J.B.; Stern, H.S.; Dunson, D.; Vehtari, A.; Rubin, D.B. *Bayesian Data Analysis*; Chapman and Hall: New York, NY, USA, 2015.
7. Siburian, M.E. Fiscal decentralization and regional income inequality: Evidence from Indonesia. *Appl. Econ. Lett.* **2019**, 1–4. [[CrossRef](#)]
8. Steele, F.; French, R.; Bartley, M. Adjusting for Selection Bias in Longitudinal Analysis Using Simultaneous Equations Modeling: The Relationship Between Employment Transitions and Mental Health. *Epidemiology* **2013**, *24*, 703–711. [[CrossRef](#)]
9. Zhao, H.; Lin, B. Impact of foreign trade on energy efficiency in China's textile industry. *J. Clean. Prod.* **2019**, *245*, 118878. [[CrossRef](#)]
10. Adewuyi, A.O.; Awodumi, O.B. Biomass energy consumption, economic growth and carbon emissions: Fresh evidence from West Africa using a simultaneous equation model. *Energy* **2017**, *119*, 453–471. [[CrossRef](#)]
11. King, T. Using simultaneous equation modelling for defining complex phenotypes. *BMC Genet.* **2003**, *4*, S10. [[CrossRef](#)] [[PubMed](#)]
12. Pitfield, D.; Caves, R.; Quddus, M. Airline strategies for aircraft size and airline frequency with changing demand and competition: A simultaneous-equations approach for traffic on the north Atlantic. *J. Air Transp. Manag.* **2009**, *16*, 151–158. [[CrossRef](#)]
13. Olmeda, N.G.; Martinez, I.B. Application of simultaneous equation models to temporary disability prescriptions in primary healthcare centres. *Int. J. Comput. Math.* **2014**, *91*, 252–260. [[CrossRef](#)]
14. Strathe, A.; Jørgensen, H.; Kebreab, E.; Danfær, A. Bayesian simultaneous equation models for the analysis of energy intake and partitioning in growing pigs. *J. Agric. Sci.* **2012**, *150*, 764–774. [[CrossRef](#)]
15. Zhou, W.; Duan, W. An empirical study of how third-party websites influence the feedback mechanism between online Word-of-Mouth and retail sales. *Decis. Support Syst.* **2015**, *76*, 14–23. [[CrossRef](#)]
16. Yang, S.; Narayan, V.; Assael, H. Estimating the Interdependence of Television Program Viewership Between Spouses: A Bayesian Simultaneous Equation Model. *Mark. Sci.* **2006**, *25*, 336–349. [[CrossRef](#)]
17. Park, H.; Manchanda, P. When Harry Bet with Sally: An Empirical Analysis of Multiple Peer Effects in Casino Gambling Behavior. *Mark. Sci.* **2015**, *2*, 179–194. [[CrossRef](#)]
18. Chen, C.; Chang, K.; Lin, L.; Lee, J. Health risk perception and betel chewing behavior: The evidence from Taiwan. *Addict. Behav.* **2013**, *38*, 2714–2717. [[CrossRef](#)]
19. Cohen, D.P.; Benadjaoud, M.A.; Lestaevel, P.; Lebsir, D.; Benderitter, M.; Souidi, M. Effects of repetitive Iodine Thyroid Blocking on the Development of the Foetal Brain and Thyroid in rats: A Systems Biology approach. *bioRxiv* **2019**. [[CrossRef](#)]
20. Akaike, H. *Information Theory and an Extension of the Maximum Likelihood Principle*; Springer: New York, NY, USA, 1998.
21. Keerativibool, W. New Criteria for Selection in Simultaneous Equations Model. *Thail. Stat.* **2012**, *10*, 163–181.
22. Hurvich, C.; Tsai, C. Regression and time series model selection in small samples. *Biometrika* **1989**, *76*, 297–397. [[CrossRef](#)]
23. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
24. Findley, D.F. Counterexamples to parsimony and BIC. *Ann. Inst. Stat. Math.* **1991**, *43*, 505–514. [[CrossRef](#)]
25. Hannan, E.J.; Quinn, B.G. The Determination of the Order of an Autoregression. *J. R. Stat. Soc. Ser. B (Methodol.)* **1979**, *41*, 190–195. [[CrossRef](#)]
26. Keerativibool, W.; Jitthavech, J. Model Selection Criterion Based on Kullback-Leibler's Symmetric Divergence for Simultaneous Equations Model. *Chiang Mai J. Sci.* **2015**, *42*, 761–773.
27. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley and Sons, Inc.: Hoboken, NJ, USA, 1991.
28. Zhou, R.; Cai, R.; Tong, G. Applications of Entropy in Finance: A Review. *Entropy* **2013**, *15*, 4909–4931. [[CrossRef](#)]
29. Cui, H.; Sivakumar, B.; Singh, V. Entropy Applications in Environmental and Water Engineering. *Entropy* **2019**, *20*, 598. [[CrossRef](#)] [[PubMed](#)]
30. Purvis, B.; Mao, Y.; Robinson, D. Entropy and its Application to Urban Systems. *Entropy* **2019**, *21*, 56. [[CrossRef](#)]
31. Oruç, Ö.E.; Kuruoglu, E.; Gündüz, A. Entropy Applications for Customer Satisfaction Survey in Information Theory. *Front. Sci.* **2011**, *1*, 1–4. [[CrossRef](#)]
32. Amigó, J.M.; Balogh, S.G.; Hernández, S. A Brief Review of Generalized Entropies. *Entropy* **2018**, *20*, 813. [[CrossRef](#)]
33. Qayyum, Z.; Hasan, S.S. K-Class estimators—a Review. *Int. J. Math. Trends Technol.* **2017**, *50*, 104–107. [[CrossRef](#)]
34. Theil, H. *Repeated Least Squares Applied to Complete Equation Systems*; Central Planning Bureau: The Hague, The Netherlands, 1953.
35. Basmann, R.L. A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation. *Econometrica* **1957**, *25*, 77–83. [[CrossRef](#)]
36. Clausius, R. *The Mechanical Theory of Heat*; Macmillan: New York, NY, USA, 1879.
37. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
38. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [[CrossRef](#)]
39. Lombardi, O.; Holik, F.; Vanni, L. What is Shannon information? *Synthese* **2016**, *193*, 1983–2012. [[CrossRef](#)]



---

# Bibliografía

---

- [1] Jan Tinbergen. *Statistical Testing of Business Cycle Theories: Part I: A Method and Its Application to Investment Activity*. A, January 1939.
- [2] E. Malinvaud. Econometric methodology at the cowles commission: Rise and maturity. *Econometric Theory*, 4(2):187–209, 1988.
- [3] T.M. King. Using simultaneous equation modelling for defining complex phenotypes. *BMC Genetics*, 4(1):S10, 2003.
- [4] Adeolu O. Adewuyi and Olabanji B. Awodumi. Biomass energy consumption, economic growth and carbon emissions: Fresh evidence from West Africa using a simultaneous equation model. *Energy*, 119:453–471, 2017.
- [5] T.A. Mroz. The sensitivity of an empirical model of married women’s hours of work to economic and statistical assumptions. *Econometrics*, 55:765–799, 1987.
- [6] Domadar Gujarati and Dawn Porter. *Econometría. 5ª edición*. México, DF. McGraw-Hill Interamericana Editores SA, 2004.
- [7] Arnold Zellner. The finite sample properties of simultaneous equations’ estimates and estimators Bayesian and non-Bayesian approaches. *Journal of Econometrics*, 83:185–212, 1998.
- [8] John C. Chao and Peter C.B. Phillips. Jeffreys prior analysis of the simultaneous equations model in the case with  $n+1$  endogenous variables. *Journal of Econometrics*, 111:251–283, 2002.
- [9] J. Geweke. Bayesian reduced rank regression in econometrics. *Journal of Econometrics*, 75(1):121–146, 1996.
- [10] Frank Kleibergen and H. Van Dijk. Bayesian simultaneous equation analysis using reduced rank structures. *Econometric Theory*, 14:701–743, 02 1998.

- 
- [11] A. Gelman, J. B. Carlin, H. S. Stern, D. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall, New York, 2015.
- [12] Matondang Elsa Siburian. Fiscal decentralization and regional income inequality: Evidence from Indonesia. *Applied Economics Letters*, 0:1–4, 2019.
- [13] Fiona Steele, Robert French, and Mel Bartley. Adjusting for selection bias in longitudinal analysis using simultaneous equations modeling. The relationship between employment transitions and mental health. *Epidemiology (Cambridge, Mass.)*, 24:703–711, 07 2013.
- [14] Hongli Zhao and Boqiang Lin. Impact of foreign trade on energy efficiency in China’s textile industry. *Journal of Cleaner Production*, 245:118878, 2019.
- [15] D.E. Pitfield, R.E. Caves, and M.A. Quddus. Airline strategies for aircraft size and airline frequency with changing demand and competition: A simultaneous-equations approach for traffic on the north Atlantic. *Journal of Air Transport Management*, 16(3):151–158, 2009.
- [16] Natividad Guadalajara Olmeda and Isabel Barrachina Martinez. Application of simultaneous equation models to temporary disability prescriptions in primary healthcare centres. *International Journal of Computer Mathematics*, 91(2):252–260, 2014.
- [17] A.B. Strathe, Henry Jørgensen, Ermias Kebreab, and A. Danfær. Bayesian simultaneous equation models for the analysis of energy intake and partitioning in growing pigs. *The Journal of Agricultural Science*, 150:764–774, 12 2012.
- [18] W. Zhou and W. Duan. An empirical study of how third-party websites influence the feedback mechanism between online Word-of-Mouth and retail sales. *Decision Support Systems*, 76:14–23, 2015.
- [19] Sha Yang, Vishal Narayan, and Henry Assael. Estimating the interdependence of television program viewership between spouses: A Bayesian simultaneous equation model. *Marketing Science*, 25:336–349, 07 2006.
- [20] H.M. Park and P. Manchanda. When Harry Bet with Sally: An empirical analysis of multiple peer effects in casino gambling behavior. *Marketing Science*, 2:179–194, 2015.
- [21] CM. Chen, KL Chang, L Lin, and JL Lee. Health risk perception and betel chewing behavior. The evidence from Taiwan. *Addictive behaviors*, 38:2714–2717, 07 2013.
- [22] David P.A. Cohen, Mohamed Amine Benadjaoud, Phillipe Lestaevel, Dalila Lebsir, Marc Benderitter, and Maâmar Souidi. Effects of repetitive iodine thyroid blocking on the development of the foetal brain and thyroid in rats: A systems biology approach. *bioRxiv*, 2019.
-

- 
- [23] Zoha Qayyum and S. Suhaib Hasan. K - Class estimators - a Review. *International Journal of Mathematics Trends and Technology*, 50(2):104–107, 2017.
- [24] W.Green. *Econometric Analysis*. Prentice Hall, 3rd edition, 1998.
- [25] H.Theil. Repeated least squares applied to complete equation systems. 1953.
- [26] R. L. Basman. A generalized classical method of linear estimation of coefficients in a structural equation. *Econometrica*, 25(1):77–83, 1957.
- [27] Thomas M. Cover. *Elements of Information Theory*. 1991.
- [28] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley and Sons, Inc., 1991.
- [29] Rongxi Zhou, Ru Cai, and Guanqun Tong. Applications of entropy in finance: A review. *Entropy*, 15(11):4909–4931, 2013.
- [30] Huijuan Cui, B. Sivakumar, and Vijay Singh. *Entropy Applications in Environmental and Water Engineering*. 2019.
- [31] Ben Purvis, Yong Mao, and Darren Robinson. Entropy and its application to urban systems. *Entropy*, 21(1):56, Jan 2019.
- [32] Özlem Ege Oruç, Emel Kuruoglu, and Aslı Gündüz. Entropy applications for customer satisfaction survey in information theory. *Frontiers in Science*, 1(1):1–4, 2011.
- [33] H. Akaike. A new look at the statistical model identification. *Springer Series in Statistics (Perspectives in Statistics)*, pages 215–222, 1974.
- [34] Warangkhan Keerativibool. New criteria for selection in simultaneous equations model. *Thailand Statistician*, 10(2):163–181, 2012.
- [35] C.M. Hurvich and C.L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76:297–397, 1989.
- [36] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.
- [37] E. J. Hannan and B. G. Quinn. The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(2):190–195, 1979.
- [38] Warangkhan Keerativibool and Jirawan Jithavech. Model selection criterion based on Kullback-Leibler's symmetric divergence for simultaneous equations model. *Chiang Mai J. Sci.*, 42(3):761–773, 2015.
-

- 
- [39] Alexander Gorobets. The optimal prediction simultaneous equations selection. *Economic Bulletin*, 36(3):1–8, 2005.
- [40] Subana Shanmuganathan. *Artificial Neural Network Modelling: An Introduction*, pages 1–14. Springer International Publishing, Cham, 2016.
- [41] Takeshi Amemiya. The nonlinear two-stage least-squares estimator. *Journal of Econometrics*, 2(2):105 – 110, 1974.
- [42] Takeshi Amemiya. The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model. *Econometrica*, 45(4):955–968, 1977.
- [43] David A. Besley. On the computational competitiveness of full-information maximum-likelihood and three-stage least-squares in the estimation of nonlinear, simultaneous-equations models. *Journal of Econometrics*, 9(3):315 – 342, 1979.
- [44] Richard Blundell, Dennis Kristensen, and Rosa L. Matzkin. Control functions and simultaneous equations methods. *The American Economic Review*, 103(3):563–569, 2013.
- [45] Stephen M. Goldfeld and Richard E. Quandt. Nonlinear simultaneous equations: Estimation and prediction. *International Economic Review*, 9(1):113–136, 1968.
- [46] C. R. Winegarden. Women’s fertility, market work and marital status: A test of the new household economics with international data. *Economica*, 51(204):447–456, 1984.
- [47] Alberto Prieto, Beatriz Prieto, Eva Martinez Ortigosa, Eduardo Ros, Francisco Pelayo, Julio Ortega, and Ignacio Rojas. Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing*, 214:242–268, 2016.
- [48] Warren Mcculloch and Walter Pitts. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:127–147, 1943.
- [49] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.
- [50] Laveen N. Kanal. *Perceptron*, page 1383–1385. John Wiley and Sons Ltd., GBR, 2003.
- [51] John C. B. Cooper. Artificial neural networks versus multivariate statistics: An application from economics. *Journal of Applied Statistics*, 26(8):909–921, 1999.
- [52] L.R. Kumar. Estimation of simultaneous econometric equations using neural networks. In *Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences*, volume iv, pages 124–128 vol.4, 1991.
-

- 
- [53] David Brennan and Lawrence Marsh. Comparing a traditional econometric simultaneous equations model with an unstructured neural network simultaneous output model. 03 1992.
- [54] Louis E. Caporaletti, Robert E. Dorsey, John D. Johnson, and William A. Powell. A decision support system for in-sample simultaneous equation systems forecasting using artificial neural systems. *Decision Support Systems*, 11(5):481–495, 1994.
- [55] Yaxue Ma, Zhichao Ba, Yuxiang Zhao, Jin Mao, and Gang Li. Understanding and predicting the dissemination of scientific papers on social media: a two-step simultaneous equation modeling–artificial neural network approach. *Scientometrics*, 126(8):7051–7085, 2021.
- [56] Rudolf Clausius. *The mechanical theory of heat*. Macmillan, 1879.
- [57] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.
- [58] Constantino Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52(1):479–487, 1988.
- [59] Olimpia Lombardi, Federico Holik, and Leonardo Vanni. What is Shannon information? *Synthese*, 193(7):1983–2012, 2016.
- [60] José M. Amigó, Samuel G. Balogh, and Sergio Hernández. A brief review of generalized entropies. *Entropy*, 20(11):813, 2018.
- [61] *The R Project for Statistical Computing*. [www.r-project.org](http://www.r-project.org), Octubre 2018.
- [62] *The R Project for Statistical Computing*. [cran.r-project.org/web/packages/neuralnet](http://cran.r-project.org/web/packages/neuralnet), July 2019.
- [63] Jeffrey M. Wooldridge. *Introductory Econometrics*. South-Western Cengage Learning, 2009.
-

