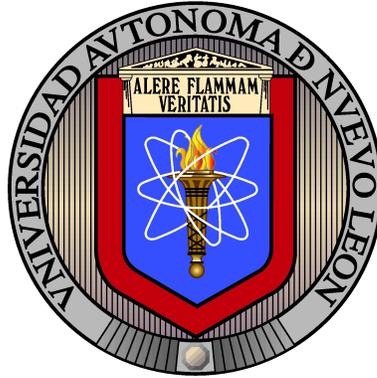


UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN  
FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA  
DIVISIÓN DE ESTUDIOS DE POSGRADO



PRONÓSTICO DE SERIES TEMPORALES CON ESTACIONALIDAD

POR

SERGIO DAVID MADRIGAL ESPINOZA

EN OPCIÓN AL GRADO DE

DOCTOR EN INGENIERÍA

CON ESPECIALIDAD EN

INGENIERÍA DE SISTEMAS

CIUDAD UNIVERSITARIA, ENERO DE 2011

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN  
FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA  
DIVISIÓN DE ESTUDIOS DE POSGRADO



PRONÓSTICO DE SERIES TEMPORALES CON ESTACIONALIDAD

POR

SERGIO DAVID MADRIGAL ESPINOZA

EN OPCIÓN AL GRADO DE

DOCTOR EN INGENIERÍA

CON ESPECIALIDAD EN

INGENIERÍA DE SISTEMAS

CIUDAD UNIVERSITARIA, ENERO DE 2011

**Universidad Autónoma de Nuevo León**  
**Facultad de Ingeniería Mecánica y Eléctrica**  
**División de Estudios de Posgrado**

Los miembros del Comité de Tesis recomendamos que la Tesis «Pronóstico de series temporales con estacionalidad», realizada por el alumno Sergio David Madrigal Espinoza, matrícula 1294374, sea aceptada para su defensa como opción al grado de Doctor en Ingeniería.

El Comité de tesis



Dr. Rodolfo Garza Morales  
Asesor



Dr. César Emilio Villarreal Rodríguez  
Revisor



Dr. Francisco Javier Almaguer Martínez  
Revisor



Dr. José Arturo Berrones Santos  
Revisor



Dr. Pedro Antonio Villezca Becerra  
Revisor

Vo. Bo.



Dr. Moisés Hinojosa Rivera  
Subdirector  
División de Estudios de Posgrado

Ciudad Universitaria, enero de 2011

---

# Dedicatoria

*Con todo mi empeño y con todo mi cariño,  
para la familia Espinoza Návares,  
les deseo lo mejor.*

---

# Agradecimientos

Primeramente, deseo expresar mis mayores agradecimientos a las instituciones que conforman el Fondo Mixto CONACYT-Gobierno del Estado de Nuevo León por darme beca de manutención durante estos años a través del proyecto clave FOMIXNL-2006-C09-32652 titulado «Dirección de tesis entre la Universidad Autónoma de Nuevo León, University of Texas at Austin y/o Instituciones de Educación Superior de la Asociación Nacional de Universidades e Instituciones de Educación Superior».

Toda mi gratitud para la Dra. Deniz Özdemir, quien me asesoró durante la mayor parte del presente trabajo. *Special thanks for you Dra. Deniz!*

Gracias al Dr. Rodolfo Garza Morales por asesorarme durante el desarrollo de la tesis. Sin usted, este trabajo no hubiera sido posible.

Gracias al Dr. César Emilio Villarreal Rodríguez, al Dr. José Arturo Berrones Santos y al Dr. Pedro Antonio Villezca Becerra, por los conocimientos que me han transmitido y cuya aplicación podrán encontrar en este trabajo. ¡Mil gracias!

Muchos compañeros han ido y venido durante la elaboración de este trabajo, pero hay dos que siguen estando ahí. Quiero agradecer en especial al Dr. Francisco Javier Almaguer Martínez y al M.C. Jonás Velasco Álvarez por su lealtad y sinceridad. Lo que me han enseñado, un agradecimiento; lo que me han aguantado, no tiene precio.

Le agradezco mucho al Dr. Oscar Leonel Chacón Mondragón por todo su apoyo, sus consejos y su ayuda. Dr. Chacón, ¡Muchas gracias! Esta tesis fue realizada con programas gratuitos como L<sup>A</sup>T<sub>E</sub>X. Esto se lo debo a la Dra. Satu Elisa Shaeffer, pues fue ella quien me introdujo en el fascinante mundo del *software* libre. ¡Gracias Dra. Elisa!

---

# Resumen

En este trabajo, se estudian diversas metodologías para el pronóstico de series temporales con estacionalidad. En particular, es de interés el pronóstico de series temporales con tendencia lineal y estacionalidad creciente, aunque nuestra propuesta también puede ser extendida para pronosticar series con estacionalidad constante.

No hemos encontrado en la literatura modelos para el pronóstico de series con tendencia lineal y estacionalidad creciente que modelen adecuadamente este tipo de patrones utilizando pocos parámetros o que el mismo modelo sea capaz de pronosticar series con estacionalidad creciente o constante.

Si un modelo usa demasiados parámetros, ocurre el fenómeno conocido como *sobreparametrización* que conduce a malos pronósticos, pues el exceso de parámetros impide la correcta caracterización de la parte de la serie cuya naturaleza es aleatoria. Si el modelo no captura correctamente el componente aleatorio de la información, el modelo fracasará en aislar los patrones que son la esencia del pronóstico; lo que se pronostica son los patrones y no el «ruido».

Desde que se inició el pronóstico de series temporales con estacionalidad, ha sido necesario proponer modelos para el pronóstico de series con estacionalidad constante o creciente. La solución en todas las metodologías ha sido la misma; proponer modelos para uno u otro caso. Sin embargo, esta solución ha conducido a otro problema, ¿Cómo distinguir entre series con estacionalidad constante o creciente? La respuesta ha sido la observación de los datos y el criterio del usuario. Esta solución no está mal cuando uno u otro patrón son muy marcados en la serie. Sin embargo, ¿Qué pasa cuando los patrones de

la serie no son tan marcados y no es fácil distinguir el tipo de estacionalidad de la serie? Para esta pregunta, no hay respuestas. Si elegimos el modelo incorrecto, fallaremos en capturar apropiadamente los patrones de la serie y los pronósticos serán malos.

Para solucionar los problemas anteriores, nosotros proponemos dos modelos para el pronóstico de series temporales con tendencia lineal y estacionalidad creciente que utilizan casi la mitad de los parámetros de la técnica *apta* de la literatura. Lo anterior contribuye a la reducción del problema de sobreparametrización y mejora los pronósticos.

Además, uno de los modelos propuestos tiene la capacidad de pronosticar series con estacionalidad constante o creciente. Esto soluciona el problema de tener que seleccionar un modelo para series con estacionalidad constante o creciente ya que el modelo propuesto pronostica ambos tipos de estacionalidad.

Los modelos propuestos no sólo se limitan al pronóstico de series con tamaño de información regular, sino que también pronostican series con información escasa y con observaciones faltantes.

En este trabajo, se presentan los modelos propuestos, su estimación paramétrica y caracterización estadística así como comparaciones con los modelos *aptos* de la literatura para comprobar su eficacia y las razones por las cuales uno de los modelos propuestos tiene la capacidad de pronosticar series con estacionalidad creciente o constante.

---

# LISTA DE ABREVIACIONES

**AR**( $p$ ) *Auto-Regresive process of order  $p$* . Proceso autorregresivo de orden  $p$ .

**ARIMA**( $p, d, q$ ) *Auto-Regresive Integrated Moving Average process of order  $p, q$* . Proceso de variables autorregresivas de orden  $p$ ; integración de orden  $d$  y de media móvil de orden  $q$ . También se les conoce como los modelos de Box y Jenkins.

**ARIMAS**( $p, d, q$ )( $P, D, Q$ ) *Auto-Regresive Integrated Moving Averages process of order  $p, d, q$  and Seasonal order  $P, D, Q$* . Proceso de variables aleatorias estacionales de orden autoregresivo  $p$  y autoregresivo estacional  $P$ ; orden de integración  $d$  e integración estacional  $D$ ; orden de media móvil  $q$  y media móvil estacional  $Q$ .

**ARMA**( $p, q$ ) *Auto-Regresive Moving Average proces of order  $p, q$* . Proceso de variables autorregresivas de orden  $p$  y de media móvil de orden  $q$ .

**BLSD** *Búsqueda Lineal con Sección Dorada*. Algoritmo diseñado para minimizar la suma de las desviaciones cuadradas del MNoLIN. Este procedimiento saca ventaja de la estructura no lineal del MNoLIN para dar una estimación paramétrica eficiente y eficaz.

**COB** *Cobertura Observada del Intervalo Predictivo*. Es la probabilidad observada para cada valor de  $h$  ( $h = 1, \dots, 12$ ) de que el dato pronosticado ocurra dentro del intervalo predictivo.

**DESV** *Desviación Absoluta Media*. Es la media de las desviaciones absolutas para cada valor de  $h$ .

---

**DESV %** *Desviación Absoluta Media Porcentual*. Es la media  $\forall h$  de las desviaciones absolutas del pronóstico divididas entre el dato pronosticado.

**EST** *Estrechez Media del Intervalo Predictivo*. Es la media de las amplitudes de los intervalos predictivos para cada valor de  $h$ .

**LEV\_MARQ** *Levenberg-Marquardt*. Algoritmo para minimizar la suma de las desviaciones cuadradas para modelos no lineales. El algoritmo lleva este nombre en honor a sus creadores.

**MA( $q$ )** *Moving Average process of order  $q$* . Procesos de media móvil de orden  $q$ .

**MALTDDB** *Modelo Altamente Determinista de [Bowerman, Koehler, y Pack](#)*. Modelo altamente determinista y el más adecuado para comparar el desempeño de los modelos propuestos.

**Matlab/Octave** *Lenguaje de programación compatible con Matlab y Octave*. Los códigos de los diferentes programas implementados en este trabajo fueron escritos en este lenguaje de programación.

**MCNL** *Mínimos Cuadrados No Lineales*. Método para la estimación paramétrica que consiste en minimizar la suma de las desviaciones cuadradas entre un modelo no lineal y una serie temporal.

**MCO** *Mínimos Cuadrados Ordinarios*. Método clásico para la estimación paramétrica que consiste en minimizar las sumas de las desviaciones cuadradas entre un modelo lineal y una serie temporal.

**MCONS** *Método de Rectas para Estacionalidad Constante*. Un método adecuado para el pronóstico de series con estacionalidad constante. Este es el modelo al que tiende el MNoLIN conforme  $\tau$  tiende en magnitud a infinito.

---

**MELI** *Mejor Estimador Lineal Insesgado*. Esta es la propiedad más importante de los estimadores MCO y se cumple cuando las desviaciones siguen un proceso de variables aleatorias independientes e idénticamente distribuidas con media 0 y desviación  $\sigma^2$ .

**MLIN** *Modelo Lineal*. A éste también nos referimos como el primer modelo propuesto o el modelo de rectas con pendiente variable.

**MNoLIN** *Modelo No Lineal*. A éste también nos referimos como el segundo modelo propuesto o el modelo de rectas con intersección en el plano.

**NID(0,  $\sigma^2$ )** *Normal Independent Distributed with Mean 0 and variance  $\sigma^2$* . Proceso de variables aleatorias independientes que siguen una distribución normal con media 0 y varianza  $\sigma^2$ .

**PAR( $p$ )** *Periodic Auto-Regresive model of order  $p$* . Modelo pariódico autorregresivo de orden  $p$ .

**SDC** *Suma de las Desviaciones Cuadradas*. En el contexto de optimización lineal, es la función que al ser minimizada devuelve estimadores MELI si las desviaciones siguen un proceso NID(0,  $\sigma^2$ ) o insesgados y consistentes si las desviaciones siguen un proceso ARMA( $p, q$ ).

**SEC\_DOR** *Sección Dorada*. Algoritmo comúnmente utilizado para la búsqueda del óptimo en regiones unimodales.

**STTLEC** *Series Temporales con Tendencia Lineal y Estacionalidad Creciente*. Hace alusión al tema principal de este trabajo.

**valor-P** *Valor Probabilista*. Es el nivel de significancia más pequeño al que se rechazaría la hipótesis nula.

---

# ÍNDICE GENERAL

<b>Dedicatoria</b>	<b>I</b>
<b>Agradecimientos</b>	<b>II</b>
<b>Resumen</b>	<b>III</b>
<b>Lista de abreviaciones</b>	<b>v</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Descripción del problema a tratar . . . . .	1
1.2. Justificación . . . . .	2
1.2.1. ¿Por qué pronosticar? . . . . .	3
1.2.2. ¿Cuándo podemos esperar la aparición de series con tendencia lineal y estacionalidad? . . . . .	5
1.3. Alcances . . . . .	10
1.4. Organización de la tesis . . . . .	10
<b>2. Antecedentes</b>	<b>14</b>
2.1. Métodos de descomposición . . . . .	14
2.2. Modelos ARIMA con transformación logarítmica . . . . .	16
2.3. Modelos ARIMA con filtros diferenciadores y componentes deterministas	18
2.4. Suavizado exponencial . . . . .	20
2.5. Modelos periódicos . . . . .	23

2.6. Conclusiones . . . . .	24
2.7. Nuestra propuesta . . . . .	26
<b>3. Los modelos propuestos</b>	<b>28</b>
3.1. El primer modelo propuesto . . . . .	28
3.1.1. Interpretación geométrica . . . . .	29
3.1.2. Tres representaciones matemáticas y sus ventajas . . . . .	30
3.1.3. Pronóstico con el primer modelo: caso de ejemplo . . . . .	33
3.2. El segundo modelo propuesto . . . . .	35
3.2.1. Interpretación geométrica . . . . .	35
3.2.2. Tres representaciones matemáticas y sus ventajas . . . . .	37
3.2.3. Pronóstico con el segundo modelo: comparación . . . . .	38
3.3. Conclusiones . . . . .	40
<b>4. Estimación de parámetros</b>	<b>41</b>
4.1. Estimación de parámetros para el primer modelo . . . . .	41
4.2. Estimación de parámetros para el segundo modelo . . . . .	43
4.2.1. El algoritmo de Levenberg-Marquardt . . . . .	43
4.2.2. El algoritmo de estimación diseñado para este modelo . . . . .	45
4.2.3. Experimentación. . . . .	51
4.3. Conclusiones . . . . .	53
<b>5. Caracterización estadística y estimaciones alternativas</b>	<b>55</b>
5.1. Dos teoremas importantes sobre regresión lineal . . . . .	56
5.1.1. Suposiciones de Gauss-Markov . . . . .	56
5.1.2. Suposiciones débiles de Gauss-Markov . . . . .	59
5.2. Caracterización estadística del MLIN . . . . .	60
5.2.1. Las desviaciones siguen un proceso NID(0, $\sigma^2$ ) . . . . .	60

5.2.2. Las desviaciones siguen un proceso ARMA( $p, q$ ) . . . . .	66
5.3. Caracterización estadística del MNoLIN . . . . .	68
5.3.1. Las desviaciones siguen un proceso NID( $0, \sigma^2$ ) y $\tau$ es conocido . . . . .	68
5.3.2. Las desviaciones siguen un proceso ARMA( $p, q$ ) y $\tau$ es conocido . . . . .	73
5.4. Estimadores alternativos para el MNoLIN . . . . .	74
5.4.1. Estimadores analíticos . . . . .	75
5.4.2. Experimentación . . . . .	77
5.5. Conclusiones . . . . .	79
<b>6. Series de ejemplo . . . . .</b>	<b>81</b>
6.1. Descripción de las series . . . . .	81
6.2. El modelo <i>apto</i> . . . . .	85
6.3. Conjuntos de datos o escenarios . . . . .	86
6.4. Conclusiones . . . . .	86
<b>7. Primer escenario: series con observaciones regulares . . . . .</b>	<b>87</b>
7.1. Lineamientos observados . . . . .	87
7.2. Resultados . . . . .	89
7.2.1. Comparación de la cobertura de los intervalos predictivos basada en el estadístico Cob . . . . .	92
7.2.2. Comparación de la estrechez de los intervalos predictivos con base en el estadístico Est . . . . .	94
7.2.3. Comparación del error de pronóstico basada en el estadístico DESV % . . . . .	96
7.3. Conclusiones sobre el pronóstico de series con observaciones regulares . . . . .	98
<b>8. Segundo escenario: series con observaciones escasas . . . . .</b>	<b>100</b>
8.1. Lineamientos observados . . . . .	101
8.2. Resultados . . . . .	102

8.2.1. Comparación de la cobertura de los intervalos predictivos basada en el estadístico COB . . . . .	104
8.2.2. Comparación de la estrechez de los intervalos predictivos basada en el estadístico Est . . . . .	105
8.2.3. Comparación del error de pronóstico basada en el estadístico Desv %	107
8.3. Conclusiones sobre el pronóstico de series con observaciones escasas . . .	108
<b>9. Tercer escenario: series con observaciones faltantes</b>	<b>109</b>
9.1. Estimación cuando hay datos faltantes . . . . .	110
9.2. Lineamientos observados . . . . .	112
9.2.1. Resultados . . . . .	114
9.3. Conclusiones . . . . .	116
<b>10. El modelo no lineal para el pronóstico de series con estacionalidad constante</b>	<b>118</b>
10.1. Comportamiento del modelo no lineal conforme el parámetro $\tau$ tiende en magnitud a infinito . . . . .	118
10.2. El comportamiento de los parámetros . . . . .	121
10.2.1. Comportamiento lineal de $\hat{\beta}_0$ . . . . .	121
10.2.2. Comportamiento constante de las $\hat{\beta}_s$ 's . . . . .	126
10.3. El modelo lineal para estacionalidad constante. . . . .	127
10.4. El MNoLIN para series con estacionalidad constante. . . . .	128
10.5. Caso de ejemplo . . . . .	129
10.6. Conclusiones . . . . .	131
<b>11. Resultados y contribuciones</b>	<b>132</b>
11.1. Modelos propuestos . . . . .	132
11.2. Estimación de parámetros . . . . .	133
11.3. Caracterización estadística . . . . .	134

11.4. Códigos computacionales . . . . .	136
11.5. Experimentación: casos de ejemplo . . . . .	136
11.6. Innovación, un modelo para el pronóstico de series con estacionalidad constante o creciente . . . . .	138
<b>12. Conclusiones y trabajos futuros.</b>	<b>139</b>
<b>Apéndices</b>	<b>142</b>
<b>A. Optimalidad de los puntos críticos</b>	<b>142</b>
<b>B. El valor del coeficiente en un peor caso</b>	<b>145</b>
<b>C. Resultados de pronosticar series con información regular</b>	<b>146</b>
C.1. Resultados con el $MLIN$ . . . . .	146
C.2. Resultados con el $MNoLIN$ . . . . .	148
C.3. Resultados con el $MALTDDB$ . . . . .	149
<b>D. Resultados obtenidos al pronosticar series con información escasa</b>	<b>151</b>
D.1. Resultados obtenidos con el $MLIN$ . . . . .	151
D.2. Resultados obtenidos con el $MNoLIN$ . . . . .	153
D.3. Resultados obtenidos con el $MALTDDB$ . . . . .	154
<b>E. Resultados de pronosticar series con información faltante</b>	<b>156</b>
<b>Bibliografía</b>	<b>158</b>
<b>Biografía</b>	<b>164</b>

---

# ÍNDICE DE FIGURAS

1.1. Exportaciones trimestrales (en miles de francos) de una compañía francesa. . . . .	2
1.2. Etapas del ciclo de vida del producto. . . . .	6
1.3. Ventas nacionales de vehículos subcompactos por mes. . . . .	8
3.1. Serie simulada con el MLIN y con intersección en (0,0) . . . . .	30
3.2. Serie simulada con el MLIN y con intersección en (0,2) . . . . .	30
3.3. Ajuste y pronóstico con el MLIN . . . . .	34
3.4. Serie generada con el segundo modelo y punto de intersección (-4,2) . . . .	36
3.5. Serie generada con el segundo modelo y punto de intersección (-12,-2) . .	36
3.6. Ajuste y pronóstico con el MNoLIN. . . . .	38
3.7. Pronósticos del MLIN y del MNoLIN en un período de prueba. . . . .	39
4.1. Datos de habitaciones de un hotel. . . . .	46
4.2. La SDC como función de $\tilde{\tau}$ ( $SDC_{\min}(\tilde{\tau})$ ). . . . .	47
4.3. La no convexidad de $SDC_{\min}(\tilde{\tau})$ para $\tilde{\tau} \in [-900, -300]$ . . . . .	48
4.4. Coeficiente de la SDC con cada algoritmo. . . . .	53
4.5. Histograma de frecuencias de la razón $SDC_{\text{BLSD}}/SDC_{\text{LEV\_MARQ}}$ . . . . .	53
5.1. Los cocientes $\ \beta_s - \hat{\beta}_s\  / \ \beta_s - \bar{\beta}_s\ $ . . . . .	78
5.2. Histograma de frecuencias de la razón $\ \beta - \hat{\beta}_s\  / \ \beta - \bar{\beta}_s\ $ . . . . .	79
6.1. Serie de ejemplo 1. . . . .	83
6.2. Serie de ejemplo 2. . . . .	83

---

6.3. Serie de ejemplo 3. . . . .	83
6.4. Serie de ejemplo 4. . . . .	84
6.5. Serie de ejemplo 5. . . . .	84
6.6. Serie de ejemplo 6. . . . .	84
9.1. Resultados de la función $\delta(\text{MDL}, \text{SR} = 1, f), \forall \text{MDL}, \forall f$ . . . . .	115
10.1. El MNoLIN con intersección en $(\tau = -100, \beta_0 = -6)$ . . . . .	119
10.2. Demanda de gasolina en Ontario y el ajuste del MNoLIN. . . . .	120
10.3. Comportamiento de $\hat{\beta}_0$ para los datos de la demanda de gasolina. . . . .	122
10.4. Comportamiento de $\hat{\beta}_1, \dots, \hat{\beta}_4$ para los datos del cuadro 10.1. . . . .	126

---

# ÍNDICE DE CUADROS

3.1. Datos tabulados en función de períodos y estaciones. . . . .	33
3.2. Exportaciones trimestrales de una compañía francesa . . . . .	34
3.3. Pronósticos con los modelos propuestos. . . . .	39
6.1. Resumen de las series de ejemplo. . . . .	82
7.1. Esc. 1; serie 1; MLIN; AR(1). . . . .	89
7.2. Esc. 1; serie 1; MNoLIN; AR(12). . . . .	89
7.3. Esc. 1; serie 1; MAltDB; AR(1, 3). . . . .	90
7.4. Comparación del estadístico COB ( $COB_{dist}$ ) para obs. regulares. . . . .	93
7.5. Comparación del estadístico EST para obs. regulares. . . . .	95
7.6. Comparación del estadístico DESV % para obs. regulares. . . . .	97
8.1. Esc. 2; serie 1; MLIN; NID( $0, \sigma^2$ ). . . . .	102
8.2. Esc. 1; Series 1; MLIN; AR(1). . . . .	102
8.3. Comparación del estadístico COB ( $COB_{dist}$ ) para obs. escasas. . . . .	104
8.4. Comparación del estadístico EST para obs. escasas. . . . .	106
8.5. Comparación del estadístico DESV % para obs. escasas. . . . .	107
9.1. $\delta(MDL, SR = 1, f); \forall MDL; \forall f$ . . . . .	114
9.2. Comparación de los 18 valores de $\delta(MDL, SR, f)$ , para cada par de modelos con cada serie. . . . .	116
10.1. Demanda de gasolina de Ontario de 1960 a 1966. . . . .	120

10.2. Resultados de pronosticar la demanda de gasolina de Ontario de 1960 a 1966. 130

C.1. Serie 1; MLIN; AR(1). . . . . 146

C.2. Serie 2; MLIN; AR(1, 3, 7). . . . . 146

C.3. Serie 3; MLIN; AR(1, 2, 3). . . . . 147

C.4. Serie 4; MLIN; AR(1, 3)\*. . . . . 147

C.5. Serie 5; MLIN; AR(1). . . . . 147

C.6. Serie 6; MLIN; NID(0,  $\sigma^2$ ). . . . . 147

C.7. Serie 1; MNoLIN; AR(12). . . . . 148

C.8. Serie 2; MNoLIN; AR(1, 3, 7). . . . . 148

C.9. Serie 3; MNoLIN; AR(1, 2). . . . . 148

C.10. Serie 4; MNoLIN; AR(1, 2, 3, 4). . . . . 148

C.11. Serie 5; MNoLIN; AR(1). . . . . 149

C.12. Serie 6; MNoLIN; NID(0,  $\sigma^2$ ). . . . . 149

C.13. Serie 1; MAltDB; AR(1, 3). . . . . 149

C.14. Serie 2; MAltDB; AR(3, 7, 12). . . . . 149

C.15. Serie 3; MAltDB; AR(1, 2, 3, 12, 14). . . . . 150

C.16. Serie 4; MAltDB; AR(1, 3). . . . . 150

C.17. Serie 5; MAltDB; AR(1). . . . . 150

C.18. Serie 6; MAltDB; NID(0,  $\sigma^2$ ). . . . . 150

D.1. Serie 1; MLIN; NID(0,  $\sigma^2$ ). . . . . 151

D.2. Serie 2; MLIN; NID(0,  $\sigma^2$ ). . . . . 151

D.3. Serie 3; MLIN; NID(0,  $\sigma^2$ ). . . . . 152

D.4. Serie 4; MLIN; NID(0,  $\sigma^2$ ). . . . . 152

D.5. Serie 5; MLIN; NID(0,  $\sigma^2$ ). . . . . 152

D.6. Serie 6; MLIN; NID(0,  $\sigma^2$ ). . . . . 152

D.7. Serie 1; MNoLIN; NID(0,  $\sigma^2$ ). . . . . 153

D.8. Serie 2; MNoLin; NID(0, $\sigma^2$ ).	153
D.9. Serie 3; MNoLin; NID(0, $\sigma^2$ ).	153
D.10. Serie 4; MNoLin; NID(0, $\sigma^2$ ).	153
D.11. Serie 5; MNoLin; NID(0, $\sigma^2$ ).	154
D.12. Serie 6; MNoLin; NID(0, $\sigma^2$ ).	154
D.13. Serie 1; MAltDB; NID(0, $\sigma^2$ ).	154
D.14. Serie 2; MAltDB; NID(0, $\sigma^2$ ).	154
D.15. Serie 3; MAltDB; NID(0, $\sigma^2$ ).	155
D.16. Serie 4; MAltDB; NID(0, $\sigma^2$ ).	155
D.17. Serie 5; MAltDB; NID(0, $\sigma^2$ ).	155
D.18. Serie 6; MAltDB; NID(0, $\sigma^2$ ).	155
E.1. $\delta(\text{MDL}, \text{SR} = 1, f); \forall \text{MDL}; \forall f$ .	156
E.2. $\delta(\text{MDL}, \text{SR} = 2, f); \forall \text{MDL}; \forall f$ .	156
E.3. $\delta(\text{MDL}, \text{SR} = 3, f); \forall \text{MDL}; \forall f$ .	157
E.4. $\delta(\text{MDL}, \text{SR} = 4, f); \forall \text{MDL}; \forall f$ .	157
E.5. $\delta(\text{MDL}, \text{SR} = 5, f); \forall \text{MDL}; \forall f$ .	157
E.6. $\delta(\text{MDL}, \text{SR} = 6, f); \forall \text{MDL}; \forall f$ .	157

---

---

# CAPÍTULO 1

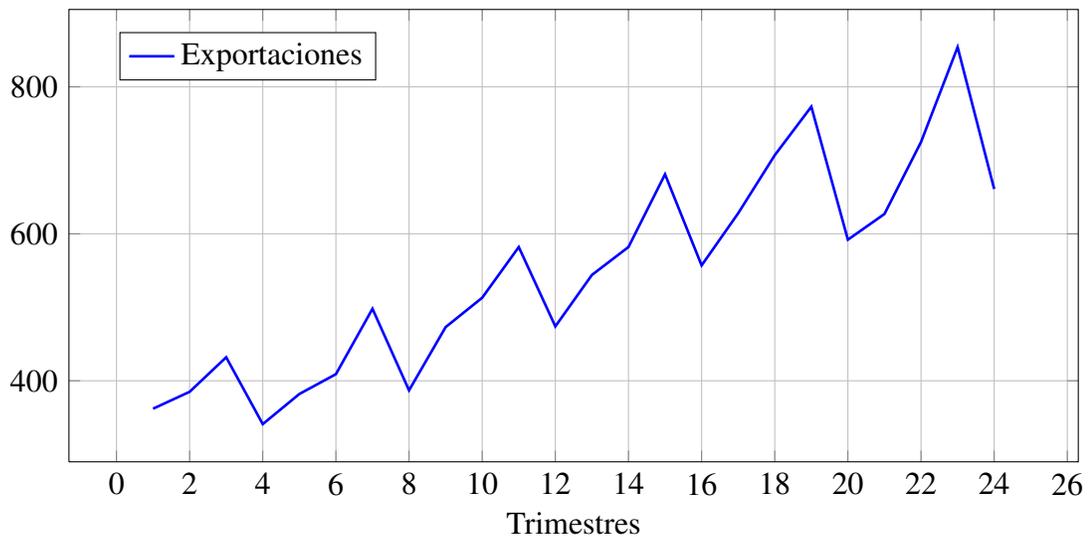
---

## Introducción

Existen muchas razones por las cuales es importante pronosticar datos de cualquier tipo. En particular, nuestro trabajo se centra en el pronóstico de datos conocidos en la literatura como *series temporales con tendencia lineal y estacionalidad creciente*. Sin embargo, nuestra propuesta también puede ser extendida para el caso de *estacionalidad constante*. A continuación, se describirán las características de este tipo de datos, la importancia de su pronóstico y el tipo de situaciones en los que se espera su aparición. Después, se discutirán los alcances de este trabajo así como la organización del mismo.

### 1.1. Descripción del problema a tratar

Una *serie temporal* es una sucesión ordenada de valores de una variable (generalmente aleatoria) observados en intervalos de tiempo igualmente espaciados. Ejemplos de series temporales pueden ser: las exportaciones trimestrales de una compañía manufacturera, las ventas nacionales de vehículos subcompactos por mes, el número mensual de pasajeros de una compañía aeronáutica, etc. *Pronóstico* es la predicción o estimación de valores futuros de una serie temporal tomando como base valores pasados de la serie en cuestión o de otras variables. Si los datos observados en una serie temporal se ven influenciados por los meses, bimestres, trimestres, etc., diremos que la serie es *estacional* y diremos que las estaciones son los meses, bimestres o trimestres que afectan a la serie. Se dice que hay *tendencia lineal* cuando se supone que las observaciones «oscilan» alrededor de una línea recta. Si



**Figura 1.1** – Exportaciones trimestrales (en miles de francos) de una compañía francesa.

además se supone que la amplitud de las estaciones crece a través del tiempo entonces, se dice que la serie posee *estacionalidad creciente*. En la figura 1.1 se muestra un ejemplo típico de una serie de este tipo. En dicha figura se puede apreciar que los datos poseen tendencia lineal y estacionalidad creciente. Los datos de esta figura fueron tomados de [Makridakis et al. \(1998, pág. 162\)](#).

A lo largo de este trabajo, nos referiremos a las Series Temporales con Tendencia Lineal y Estacionalidad Creciente por sus siglas STTLEC. El interés central de este trabajo es la propuesta de técnicas para el pronóstico de STTLEC. Una de nuestras propuestas puede extenderse al caso de series con estacionalidad constante, que son aquellas cuyas amplitudes estacionales *no* cambian a través del tiempo. En el capítulo 10 se abundará sobre el tema de estacionalidad constante.

## 1.2. Justificación

A continuación, se explicará detalladamente la relevancia del pronóstico en la industria y después, cuando es que podemos esperar la aparición de las series de nuestro interés.

### **1.2.1. ¿Por qué pronosticar?**

En la industria, el estudio y pronóstico de series temporales es de utilidad entre otros, para los empresarios que desean conocer la demanda futura de su producto y con ello tomar decisiones tácticas como la planificación de la producción de tal forma que se minimice el costo total de producción e inventarios. Desde un punto de vista estratégico, el empresario puede tomar la decisión de expandir o recortar la capacidad de producción de la planta basándose entre otras cosas, en el pronóstico. Todas estas razones hacen que el pronóstico sea muy importante desde un punto de vista táctico y estratégico.

Normalmente, existe un tiempo de espera entre el conocimiento de un evento y la ocurrencia de éste. Dicho tiempo de espera es la razón principal por la cual existe la planificación y el pronóstico. Si la espera es nula o muy pequeña entonces, no existirá la necesidad de planear. Si el tiempo de espera es muy largo y los resultados del evento final dependen de factores identificables entonces, la planificación podría desempeñar un papel muy importante. En este tipo de situaciones, el pronóstico es necesario para determinar cuando es que un evento ocurrirá de tal manera que se puedan tomar medidas adecuadas.

En administración, la necesidad de planear es grande debido a que el tiempo de espera para la toma de decisiones va del orden de los años (planificación estratégica), pasando por el el orden de los días (para decisiones operativas) hasta llegar inclusive al orden de los segundos (encaminado de telecomunicaciones, planificación dinámica, administración en tiempo real y demanda de electricidad). El pronóstico es de gran ayuda en una planificación eficiente y eficaz.

El pronóstico forma parte importante en la toma de decisiones del administrador. Las organizaciones establecen objetivos, intentan predecir factores del entorno y luego, toman decisiones de tal manera que se logren las metas perseguidas. La necesidad de pronosticar se incrementa tanto como el decisor desee reducir su dependencia de la fortuna y quiera tomar sus decisiones con bases más científicas. Debido a que cada área de una organización

esta relacionada con todas las demás, un buen o mal pronóstico puede afectar a todas ellas. Algunas de las áreas en las que el pronóstico juega un papel importante son:

1. **Programación:** El uso eficiente de recursos requiere de programar la producción, transportación, efectivo, personal, etc. El pronóstico de la demanda, material, trabajo, financiero y de servicios son en esencia la base de esta programación.
2. **Adquisición de recursos:** El tiempo de espera entre la orden y la recepción de materia prima, contratación de personal o la compra de maquinaria y equipos nuevos, puede variar entre algunos días o algunos años. El pronóstico se necesita para la previsión de estas situaciones.
3. **Determinar el requerimiento de recursos:** Toda organización debe determinar cuales recursos quiere a largo plazo. Tales decisiones dependen de las oportunidades del mercado, el entorno y del desarrollo interno de sus finanzas, personal, productos y recursos tecnológicos. Determinar todo lo anterior requiere de un buen pronóstico y de administradores capaces de interpretar las predicciones y tomar decisiones apropiadas.

A pesar de que existen muchas áreas en una organización que requieren del pronóstico, las tres que han sido mencionadas ilustran el pronóstico a corto, mediano y largo plazo. Estos plazos requieren que las compañías desarrollen múltiples métodos para predecir eventos inciertos y que desarrollen un sistema de pronóstico. Para que esto sea posible, se necesita que la organización posea conocimientos y habilidades que cubran al menos los siguientes cuatro puntos:

### **Puntos de un sistema de pronóstico organizacional**

1. Identificación y definición del problema de pronóstico.
2. La aplicación de un amplio número de técnicas de pronóstico

3. Criterios para la selección de un método de pronóstico adecuado.

4. El apoyo de la organización para la correcta aplicación.

En este proyecto, nos enfocamos en el punto 2 ya que investigamos el uso de diversos modelos para el pronóstico de series temporales con estacionalidad y además, proponemos nuevos modelos para este tipo de datos.

### **1.2.2. ¿Cuándo podemos esperar la aparición de series con tendencia lineal y estacionalidad?**

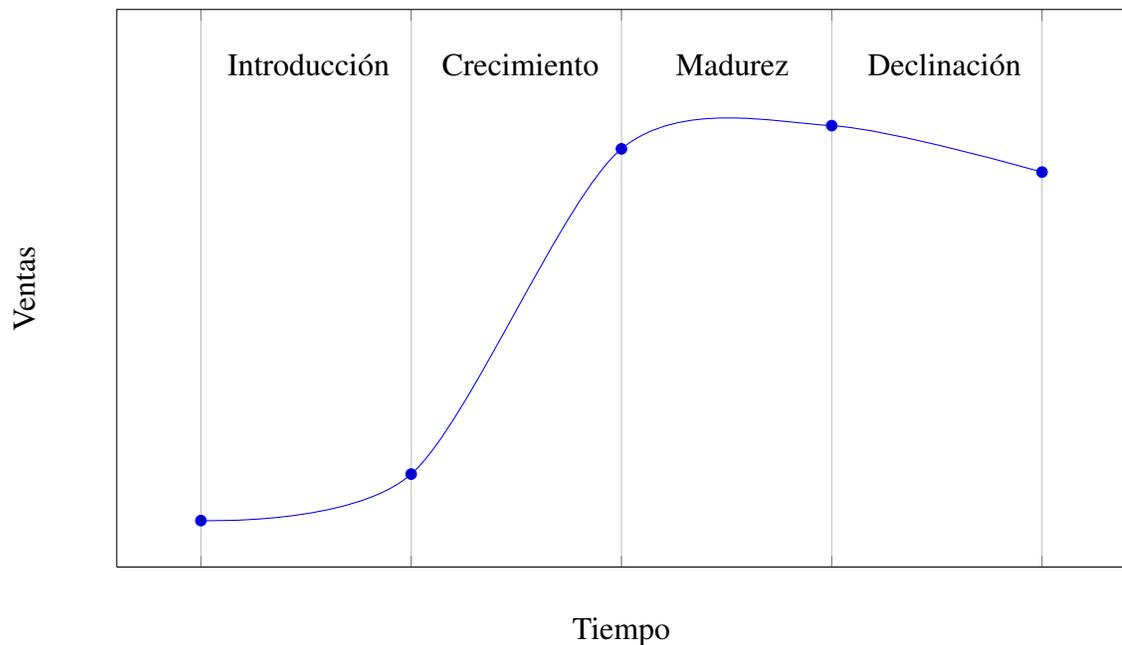
La pregunta que motiva esta subsección puede a su vez dividirse en las dos siguientes preguntas:

#### **¿Cuándo se espera la aparición de series temporales con estacionalidad?**

Con frecuencia, la demanda de un producto, bien o servicio, es afectada por las estaciones del año, el trimestre, el mes, la semana, el día, etc. Por ejemplo, la demanda de bebidas frías aumenta en verano y disminuye en invierno; el consumo de artículos costosos aumenta en diciembre por que en este mes muchas personas reciben su aguinaldo y tienen dinero para adquirir este tipo de bienes. Hay una infinidad de ejemplos en las que resulta obvia la influencia estacional y este fenómeno se repite periódicamente sin distinciones.

#### **¿Cuándo se espera la aparición de series temporales con tendencia lineal?**

En un sector industrial, cada producto pasa por varias fases o etapas —introducción, crecimiento, madurez y declinación— como se aprecia en la figura 1.2. A este concepto se le conoce como *ciclo de vida del producto*. Las etapas las definen los puntos donde se modifica la tasa de crecimiento de las ventas. El crecimiento sigue una forma de S a causa del proceso de innovación y difusión de un producto nuevo. La fase plana de la introducción refleja la dificultad de superar la inercia de los compradores y de estimular



**Figura 1.2** – Etapas del ciclo de vida del producto.

la prueba del nuevo producto. Se registra un crecimiento rápido a medida que un gran número de compradores entran presurosos al mercado, una vez demostrada la efectividad del producto. Finalmente, se logra penetrar en los consumidores potenciales; entonces el crecimiento se detiene y alcanza el nivel básico de crecimiento del grupo relevante de compradores. Por último, el crecimiento se estanca a medida que aparecen productos sustitutos.

El ciclo de vida del producto ha suscitado algunas críticas válidas:

1. La duración de las etapas varía dependiendo del sector industrial al que pertenezca; a menudo no se sabe con certeza en cual etapa se encuentra un sector. Este problema aminora la utilidad del concepto como herramienta de planificación.
2. Algunas industria no siempre muestran el patrón de forma de S. Algunas veces, omiten la etapa de madurez, pues pasan directamente del crecimiento a la declinación; ocurre comúnmente que el crecimiento se revitaliza tras un período de decaimiento o derrumbe; en otras ocasiones, parece que se omite por completo el arranque lento de la fase introductoria.

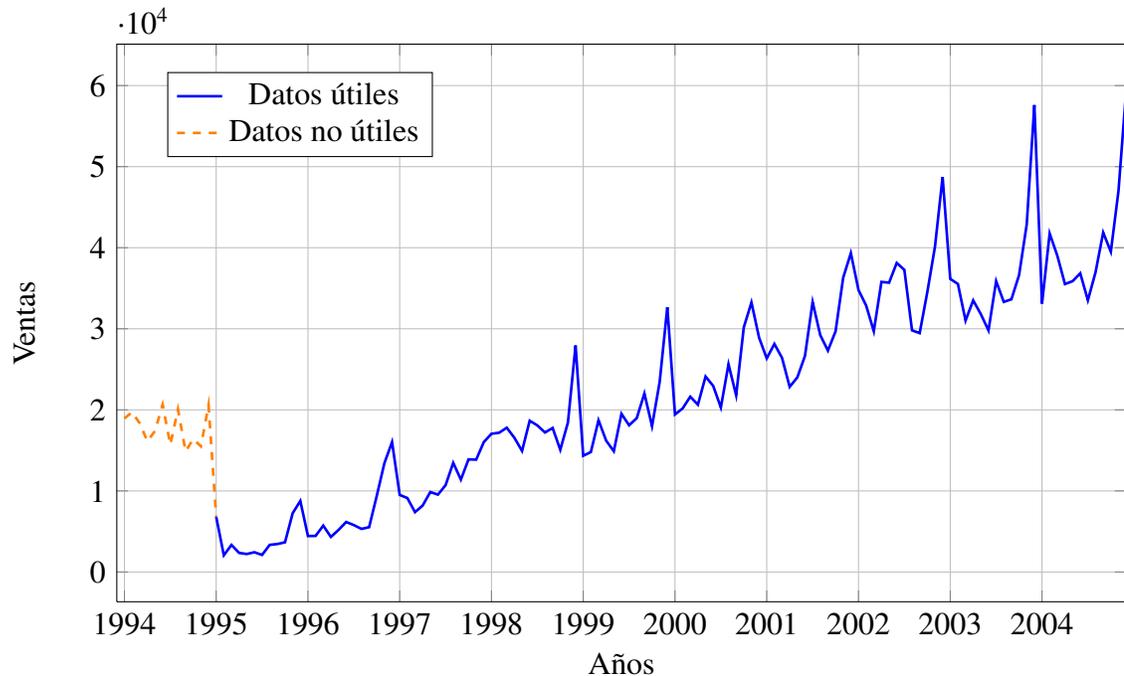
3. Las compañías pueden influir en la forma de la curva del crecimiento innovando, reposicionando y extendiendo sus productos en varias formas. Si una compañía se limita a seguir el ciclo, éste se convertirá en una especie de camisa de fuerza.
4. La índole de la competencia en cada etapa del ciclo de vida es diferente en varias industrias. Por ejemplo, algunas nacen muy concentradas y conservan siempre esta característica. Otras, como la de los cajeros automáticos, se concentran durante un período largo y luego van desconcentrándose. Otras más comienzan sumamente fragmentadas; algunas de ellas se consolidan y algunas no. Los mismos patrones divergentes los encontramos en la publicidad, en la inversión destinada a investigación y desarrollo, en la competencia de precios y en la mayoría de sus características. Estos patrones diferentes ponen seriamente en tela de juicio las consecuencias estratégicas que se atribuyen al ciclo de vida.

De los puntos mencionados anteriormente, en este trabajo nos enfocamos en productos o servicios, cuya demanda sigue un ciclo de vida cuyo crecimiento exhibe patrones estacionales y similar a las críticas realistas del segundo punto. Son de especial interés las series estacionales que cumplen con lo siguiente:

1. Cuando el proceso parece omitir por completo el arranque lento de la fase en la que se introduce el nuevo bien o servicio, entrando directamente a la etapa de crecimiento.
2. Cuando el período de crecimiento se revitaliza después de una etapa de derrumbe.

Un ejemplo del primer caso son las exportaciones de la compañía francesa presentadas en la figura 1.1. En dicha figura, se puede observar que la etapa de arranque es nula o demasiado pequeña. Quizás fue necesario sólo el primer trimestre para que el producto se consolidara. Existen varios factores que pueden contribuir a ello por ejemplo, la publicidad, los contactos de la compañía antes de que arrancara la venta del producto, etc.

Un ejemplo del segundo caso son las ventas nacionales de vehículos subcompactos. En la figura 1.3 se muestra la gráfica de estos datos desde el año 1994 hasta 2004. Estos datos fueron recopilados por la Asociación Mexicana de la Industria Automotriz y pueden ser vistos en las revistas [INEGI \(2000\)](#), [INEGI \(2001\)](#) e [INEGI \(2005\)](#). Algunas de las características que pueden observarse en la serie son las siguientes:



**Figura 1.3** – Ventas nacionales de vehículos subcompactos por mes.

1. **Las ventas exhiben estabilidad entre los años 1994 y 1995** Durante ese año el país poseía estabilidad económica que le permitía a los consumidores de subcompactos, mantener un ritmo ciertamente constante de compras pero sin alcanzar a crecer.
2. **El nivel de las ventas se derrumbó en 1995.** Esto se debe a que el país sufrió una severa crisis económica en ese año. Se puede decir que el mercado «volvió a nacer» después de esta crisis. Conforme fueron pasando los años, el país ha ido superando poco a poco este fenómeno económico, razón por la cual se observa un crecimiento en la demanda de subcompactos a través de los años.

3. **Comportamiento estacional cada año.** La demanda de automóviles suele incrementarse o decrementarse en ciertos meses del año. En los meses de noviembre y diciembre por ejemplo, la demanda es muy alta, especialmente en el último mes, mientras que en el resto de los meses la demanda es más o menos regular.
4. **Estacionalidad creciente.** La amplitud de la estacionalidad se incrementa con el paso de los años. Por ejemplo, el pico decembrino que se puede observar en el año 1996 es menor al observado en 1997 y ambos son pequeños en comparación a los observados en los años 2003 y 2004. El incremento es una clara evidencia de estacionalidad creciente.
5. **Tendencia lineal.** El crecimiento en la tendencia es relativamente constante o lineal. Esto significa que el comportamiento de la tendencia es similar a una línea recta.

Debido a que los datos entre los años 1994 y 1995 son muy diferentes al resto de la serie, los datos correspondientes a este año sólo causarían dificultades al incluirlos en un modelo de pronóstico y por eso, es mejor tomar en cuenta sólo los datos a partir del año 1995 y excluir los demás. Proceder de esta manera, conduce a una STTLEC. La serie de subcompactos son evidencia de que el crecimiento constante o lineal no sólo se alcanza cuando el proceso empieza sus operaciones, sino que también ocurre después de que el mercado toca fondo por una crisis económica, como en este caso.

Este trabajo se enfoca en el pronóstico del consumo de productos, bienes o servicios que se encuentran en etapa de crecimiento constante o lineal y que son afectados por las estaciones del año. En este trabajo, se investigarán varios métodos, como se sugiere en la segunda etapa del proceso de pronóstico, para tener más opciones e hipótesis sobre el comportamiento del bien en crecimiento con la finalidad de elegir aquel modelo que pronostique mejor su comportamiento.

### 1.3. Alcances

En este trabajo se proponen dos modelos de regresión para el pronóstico de STTLEC. El primero es un modelo lineal y el segundo es un modelo de naturaleza no lineal. Debido a que el segundo modelo es de regresión no lineal, es necesario utilizar técnicas de regresión no lineal para estimar sus parámetros. Nosotros proponemos dos maneras alternativas a la técnica estándar de la literatura para la estimación de los parámetros de este modelo, una de las cuales resultó mejor que la técnica estándar para realizar esta labor; el algoritmo de Levenberg-Marquardt.

Además, se deducen las propiedades estadísticas de cada modelo propuesto para dar al usuario datos adicionales al pronóstico como lo son los intervalos predictivos y las pruebas de idoneidad de los parámetros entre las que se incluyen insesgadez, consistencia y las condiciones bajo las cuales son los mejores estimadores lineales insesgados.

Se realizan tres experimentos para comparar los modelos propuestos en este trabajo contra la alternativa *apta* de la literatura para el pronóstico de STTLEC. Los experimentos incluyen situaciones comunes al pronosticar series temporales, como lo son series con número de datos regular, pequeño o con datos faltantes.

Por último, se demuestra matemáticamente la capacidad del modelo no lineal para el pronóstico de series temporales con tendencia lineal y estacionalidad constante. La demostración implica que cuando uno de los parámetros del modelo propuesto tiende a infinito, el modelo no lineal tiende a un conocido modelo para el pronóstico de este tipo de series.

### 1.4. Organización de la tesis

La organización de esta tesis será la siguiente:

**Capítulo 1** Introducción.

**Capítulo 2** Se revisa la literatura para conocer las ventajas y desventajas de todas las metodologías existentes para el pronóstico de STTLEC.

**Capítulo 3** Veremos el primer modelo propuesto para el pronóstico de STTLEC. Este es un modelo de regresión lineal que tiene todas las ventajas de los modelos lineales. También veremos el segundo modelo propuesto para el pronóstico de STTLEC. A pesar de que este es un modelo de regresión no lineal, puede ser bien aproximado a un modelo lineal suponiendo el conocimiento de sólo uno de sus parámetros. Al final de este capítulo, veremos los resultados de ambos modelos al pronosticar un caso de ejemplo.

**Capítulo 4** Veremos la manera clásica de estimar los parámetros del primer modelo que es de regresión lineal; *Mínimos Cuadrados Ordinarios*. Para el caso del segundo modelo, estudiaremos dos técnicas para estimar sus parámetros. La primera es considerada la técnica estándar para realizar esta labor; el algoritmo de Levenberg-Marquardt. La segunda, consiste en un algoritmo propuesto en este trabajo y tiene como objetivo minimizar la suma del cuadrado de las desviaciones entre el modelo y la serie de una manera más eficiente y simple de lo que lo hace el algoritmo de Levenberg-Marquardt. Al final, se realizará un experimento para saber cual de los algoritmos realiza la tarea de minimizar la suma de desviaciones cuadradas de la manera más eficiente.

**Capítulo 5** Se desarrollan las propiedades estadísticas de ambos modelos. El primer modelo, al ser de regresión lineal, tiene muchas bondades en este aspecto y el desarrollo de sus bases estadísticas no supone el conocimiento de los parámetros. Sin embargo, el desarrollo de las bases estadísticas del segundo modelo supone que uno de sus parámetros es conocido con la finalidad de aproximar este modelo no lineal a uno lineal y aprovechar las bondades estadísticas de los modelo lineales. También veremos una técnica alternativa para estimar los

parámetros del segundo modelo. La nueva técnica tiene varias ventajas. Por ejemplo, la deducción de sus bases estadísticas no supone el conocimiento de ningún parámetro y los estimadores no necesitan de ningún algoritmo para ser calculados. Al final, veremos un experimento en el que se comparan estos estimadores con los que minimizan la suma de las desviaciones cuadradas.

**Capítulo 6** Se explican las características de seis series que serán utilizadas para comparar los modelos propuesto con el modelo *apto* de la literatura. También se explican los escenarios bajo los cuales se desarrollarán las comparaciones.

**Capítulo 7** Se comparan los modelos utilizando seis series reales suponiendo que hay un número regular de observaciones. Este escenario supone que las series a pronosticar tienen más de cuatro períodos de observaciones y que no hay observaciones faltantes en las series. También se obtendrán resultados sobre el desempeño de los intervalos predictivos de los modelos.

**Capítulo 8** Se compara el desempeño de los modelos bajo la suposición de que la información es escasa. Se tomarán las seis series de ejemplo tomando los primeros dos a cuatro períodos de observaciones para pronosticar.

**Capítulo 9** Se diseña un experimento para comparar el desempeño de los modelos suponiendo que hay información faltante en las seis series de ejemplo. Se tomarán observaciones de manera aleatoria de cada serie y se verá lo que ocurre con el pronóstico cuando el número de observaciones faltantes se incrementa.

**Capítulo 10** Se demuestra matemáticamente, la habilidad del modelo de regresión no lineal de pronosticar series con tendencia lineal y estacionalidad constante. La prueba consiste en que el modelo no lineal tiende a una técnica para el pronóstico de series con estacionalidad constante cuando uno de sus parámetros tiende

a infinito. Se presenta un caso de estudio en el que se compara el modelo no lineal contra la técnica *apta* de la literatura.

**Capítulo 11** Se discuten las contribuciones del presente trabajo.

**Capítulo 12** Se presentan las conclusiones y los trabajos futuros tentativos.

---

---

# CAPÍTULO 2

---

## Antecedentes

Hay una gran variedad de *metodologías* para el pronóstico de datos con *tendencia lineal* (que es un caso particular de la llamada *tendencia polinomial* o *aditiva*) y *estacionalidad constante* (aditiva) o *creciente* (multiplicativa). Los *modelos de pronóstico* de estas metodologías suelen ser utilizados para el pronóstico de datos con *tendencia lineal* y *estacionalidad constante* o *creciente*.

A continuación, se revisará la frontera de desarrollo de cinco metodologías relevantes para el pronóstico de series con *tendencia polinomial* y *estacionalidad constante* o *creciente*. Los métodos revisados serán:

1. Metodología de descomposición.
2. Metodología ARIMA con transformación logarítmica.
3. Metodología ARIMA con filtros diferenciadores y componentes deterministas.
4. Metodología de suavizado exponencial.
5. Metodología periódica autorregresiva.

### 2.1. Métodos de descomposición

Esta metodología se encuentra entre las más antiguas y surgió como una manera de eliminar las correlaciones observadas en las series que se debían a la *tendencia* de los

datos. En el trabajo pionero de [Poynting \(1884\)](#) se muestra como eliminar la tendencia y la estacionalidad de una serie temporal promediando precios a lo largo de varios años. [Hooker \(1901\)](#) siguió los pasos de Poynting pero con maneras más eficaces de eliminar la tendencia. Después, [Spencer \(1904\)](#) y [Anderson y Nochmals \(1914\)](#) generalizaron el procedimiento de eliminación de tendencia mediante el uso de polinomios de alto grado.

Los métodos de descomposición tuvieron un auge con el advenimiento de las computadoras. [Shiskin \(1957\)](#) desarrolló un programa computacional capaz de realizar los tediosos cálculos asociados a este método de manera fácil y rápida.

Se han realizado esfuerzos como los de [Dagum \(1982\)](#) y [Cleveland \(1983\)](#) para crear *bases estadísticas* para estos métodos.

Los métodos de descomposición suponen que la serie observada es consecuencia de patrones identificables e irregularidades observadas en cada etapa de la serie. El primer paso consiste en tratar de identificar los componentes que ocasionan los patrones. Por lo general, se supone que éstos son generados a su vez por dos variables; *ciclicidad-tendencia* y estacionalidad. Una vez identificadas las variables irregulares y sus interacciones, se supone que las irregularidades son las diferencias entre los datos observados y los patrones.

En general, la representación matemática de un modelo perteneciente a la metodología de descomposición es

$$y_t = f(S_t, T_t, E_t) \quad (2.1)$$

donde  $y_t$  es el valor de la serie (al tiempo  $t$ ,  $t = 1, \dots, n$ );  $S_t$  es el componente estacional;  $T_t$  representa la ciclicidad-tendencia; finalmente,  $E_t$  es la irregularidad de la serie. La forma *aditiva* (estacionalidad aditiva e irregularidad aditiva (*homoscedasticidad*)) es

$$y_t = S_t + T_t + E_t \quad (2.2)$$

La formula multiplicativa (estacionalidad multiplicativa e irregularidad multiplicativa

(heteroscedasticidad)) es

$$y_t = S_t T_t E_t. \quad (2.3)$$

A la ecuación (2.3) se le aplica una transformación logarítmica para obtener

$$\log y_t = \log S_t + \log T_t + \log E_t \quad (2.4)$$

Claramente, no hay una manera directa de capturar estacionalidad multiplicativa utilizando métodos de descomposición. La manera en que la metodología hace esta labor es aplicando una transformación logarítmica a los datos para cambiar la estacionalidad multiplicativa por una aditiva.

A pesar de la gran cantidad de literatura relacionada con los métodos de descomposición, en el texto clásico de pronósticos de [Makridakis et al. \(1998\)](#) se enfatiza lo siguiente:

... en la práctica (los métodos de descomposición) raramente funcionan bien. El poco trabajo que se ha hecho sobre sus bases estadísticas suele ser obscurecido por el hecho de que los errores suelen estar correlacionados. En la actualidad, estos métodos se utilizan para introducir a los estudiantes en las cuestiones relacionadas con pronósticos de series estacionales...

## 2.2. Modelos ARIMA con transformación logarítmica

Existen modelos ARIMA para el pronóstico de series temporales con tendencia lineal y estacionalidad aditiva. Al igual que en la metodología de descomposición, la transformación logarítmica tiene el objetivo de transformar series en las que se observa estacionalidad multiplicativa por series con estacionalidad aditiva. El objetivo es utilizar los modelos aditivos preexistentes para series con estacionalidad multiplicativa transformada en aditiva. Por ejemplo, [Chatfield y Prothero \(1973\)](#) proponen una transformación logarítmica

y luego la aplicación de la metodología *ARIMAS* que no es más que la metodología *ARIMA* aplicada en el contexto de series estacionales. La ventaja de los modelos *ARIMA* radica en sus bases estadísticas sólidas (por lo general, cumplen con la suposición de desviaciones  $NID(0, \sigma^2)$ ).

Los modelos *ARIMA* para datos con estacionalidad aditiva se representa por *ARIMAS*- $(p, d, q)(P, D, Q)$  y tienen la forma

$$\phi_p(B)\Phi_P(B^m)(1-B)^d(1-B^m)^D y_t = \theta_q(B)\Theta_Q(B^m)(1-B)^d(1-B^m)^D \varepsilon_t, \quad (2.5)$$

donde

$$\phi_p(B) = (1 - \phi_1 B - \dots - \phi_p B^p), \quad (2.6)$$

$$\phi_P(B^m) = (1 - \Phi_1 B^m - \dots - \Phi_P B^{m+P}), \quad (2.7)$$

$$\theta_q(B) = (1 - \theta_1 B - \dots - \theta_p B^q), \quad (2.8)$$

$$\theta_Q(B^m) = (1 - \Theta_1 B^m - \dots - \Theta_P B^{m+Q}), \quad (2.9)$$

donde  $B$  es el operador de rezagos tal que  $By_t = y_{t-1}$ . El resto de los parámetros se debe estimar de tal forma que se maximice la *función de verosimilitud* asociada a este modelo.

Para el caso de series con estacionalidad multiplicativa, se aplica el modelo anterior sólo que se reemplaza  $y_t$  por  $\log y_t$ .

La desventaja de la metodología *ARIMA* con transformación logarítmica es que produce intervalos de predicción muy grandes (Bowerman et al. (1990)). Los intervalos son entre 150 y 250 % más amplios que los producidos con otros métodos. Además, los intervalos no son simétricos ni tiene la propiedad de ser no decrecientes (aveces crecen, aveces decrecen) y cuando crecen, parecen hacerlo de forma exponencial. Además, en el caso de datos con tendencia lineal y estacionalidad creciente, la transformación logarítmica convierte la tendencia lineal observada en una más compleja (polinomial o incluso exponencial).

## 2.3. Modelos ARIMA con filtros diferenciadores y componentes deterministas

Estos modelos solucionan el problema que conlleva la transformación logarítmica de los datos. Lo logran al incorporar filtros diferenciadores (conocidos como procesos de raíz unitaria) y componentes estacionales deterministas. Con lo anterior, la metodología ARIMA puede pronosticar de manera directa (sin transformaciones logarítmicas) series temporales con tendencia lineal y estacionalidad creciente. Los primeros en sugerir direcciones de investigación en esta área fueron [Akaike \(1973\)](#), [Priestly \(1973\)](#) y [Wilson \(1973\)](#) al discutir el artículo de [Chatfield y Prothero \(1973\)](#) sobre transformación logarítmica. [Bowerman et al. \(1990\)](#) reunieron esta información y propusieron diferentes filtros diferenciadores y modelos deterministas con variables binarias. [Franses \(1996\)](#) agregó más modelos de este tipo para sumar un total de seis. [Franses y Koehler \(1998\)](#) propusieron una estrategia para seleccionar a priori, de entre los seis modelos, el más adecuado para la serie de tiempo que se desea pronosticar.

Una de las cosas destacables de esta metodología es que dentro de ella hay modelos especialmente diseñados para el pronóstico de series con tendencia lineal y estacionalidad constante o creciente. Por ejemplo, los modelos para datos con estacionalidad constante son

$$\Delta_1 \Delta_m y_t = \mu_t, \quad (2.10)$$

$$\Delta_m y_t = \alpha_0 + \mu_t, \quad (2.11)$$

$$\Delta_1 y_t = \sum_{s=1}^m \alpha_s D_{s,t} + \mu_t, \quad (2.12)$$

$$y_t = \sum_{s=1}^m \alpha_s D_{s,t} + \beta_0 t + \mu_t, \quad (2.13)$$

donde  $\Delta_j = (1 - B^j)$  es el llamado filtro diferenciador;  $\alpha_0$ ,  $\alpha_s$  y  $\beta_0$  son constantes; finalmente,

$D_{s,t}$  es una variable binaria que toma el valor de uno si la estación  $s$  corresponde al tiempo  $t$  o cero de otro modo. Los modelos (2.10), (2.11) y (2.12) sirven para modelar tendencias polinomiales. El modelo (2.13) está diseñado para el caso específico de datos con tendencia lineal y estacionalidad constante. Franses y Koehler (1998) mostraron que los modelos (2.10), (2.11), (2.12) y (2.13) pueden ser «extendidos» para el caso de estacionalidad creciente de la siguiente manera

$$\Delta_1 \Delta_m^2 y_t = \mu_t, \quad (2.14)$$

$$\Delta_m^2 y_t = \alpha_0 + \mu_t, \quad (2.15)$$

$$\Delta_1 \Delta_m y_t = \sum_{s=1}^m \alpha_s D_{s,t} + \mu_t, \quad (2.16)$$

$$\Delta_m y_t = \sum_{s=1}^m \alpha_s D_{s,t} + \beta_0 t + \mu_t, \quad (2.17)$$

$$\Delta_1 y_t = \sum_{s=1}^m \alpha_s D_{s,t} + \sum_{s=1}^m \beta_s D_{s,t} t + \mu_t, \quad (2.18)$$

$$y_t = \sum_{s=1}^m \alpha_s D_{s,t} + \sum_{s=1}^m \beta_s D_{s,t} t + \mu_t, \quad (2.19)$$

Los modelos (2.14), (2.15), (2.16), (2.17) y (2.18) son recomendados para datos con tendencia polinomial o estocástica. El modelo (2.19) es conocido como el modelo altamente determinista y está especialmente diseñado para el pronóstico de datos con tendencia lineal y estacionalidad creciente. Franses y Koehler (1998) mostraron empíricamente que el modelo (2.19) es el apropiado para el pronóstico de datos con tendencia lineal y estacionalidad creciente.

La desventaja de esta metodología radica en que requiere el sacrificio de una gran cantidad de datos o la estimación de muchos parámetros. Por ejemplo, el modelo (2.14) reduce una serie de  $n$  observaciones en una de  $n - 2m - 1$  y el modelo (2.19) necesita estimar  $2m$  parámetros. Que se sacrifiquen datos o se estimen demasiados parámetros es dañino ya que entre menos datos tenemos, más difícil es caracterizar la serie temporal y

predecir su comportamiento. Por otra parte, la alternativa de estimar muchos parámetros también es perjudicial pues aumenta la posibilidad de cometer sobreparametrización.

## 2.4. Suavizado exponencial

En 1956, Brown dio una conferencia para la *Operations Research Society of America* que sería la base del libro *Statistical forecasting for inventory control* (Brown (1959)) y que se convertiría en el primer antecedente de la metodología de suavizado exponencial. Brown propuso un método para el pronóstico de series *estacionarias* conocido a la postre como el método de Brown. La innovación del método radicaba en dar un peso exponencialmente mayor a las observaciones más recientes, por eso se les llama métodos de suavizado exponencial. Esta ponderación, sonaba adecuada desde un punto de vista intuitivo. Naturalmente, la practica contribuyo a la comprobación de esta idea.

El método de Brown no era capaz de capturar tendencia de manera directa. En 1957, Holt desarrollo de manera independiente al trabajo de Brown, un método capaz de realizar esta labor. El método de Holt fue documentado en un memorando para la *Office of Naval Research* y no fue publicado sino hasta el año 2004 (Holt (2004b) y Holt (2004a)). Este hecho no impidió a Holt divulgar sus ideas, las cuales fueron muy populares hacia el año 1960. El método capaz de capturar tendencia aditiva fue conocido a la postre como el método de Holt. Sin embargo, los trabajos de Holt no se limitaron a series con tendencia aditiva, sino que llegó a proponer modelos para series con estacionalidad.

En 1960, Winters aplicó las ideas de Holt sobre métodos de suavizado exponencial para series con estacionalidad a casos reales (Winters (1960)). La publicación incluye dos métodos de suavizado exponencial capaces de realizar esta labor. Al primero se le conoce como el Método Aditivo de Holt-Winters para el caso de estacionalidad aditiva. El segundo es conocido como el Método Multiplicativo de Holt-Winters y se utiliza para series con estacionalidad multiplicativa.

[Pegels \(1969\)](#) reunió toda la información disponible para crear la clasificación de modelos de suavizado exponencial en la que se reunieron 9 modelos existentes hasta esa fecha. En la actualidad hay 24 modelos en esa clasificación ([Hyndman et al. \(2002\)](#) y [Gardner \(2006\)](#)).

Un avance importante en cuanto a la estimación de parámetros en los métodos de suavizado exponencial vino de la mano de las hojas de cálculo y especialmente de los optimizadores que las acompañaban. Tal es el caso de Microsoft Excel. Esta hoja de cálculo fue utilizada por [Rasmussen \(2004\)](#) para estimar los valores de los parámetros de tal forma que el optimizador minimizara la suma del cuadrado de las desviaciones. Ahí se demostró que este procedimiento era mejor que el heurístico tradicional propuesto por [Winters \(1960\)](#).

A pesar de la eficacia y popularidad de los métodos de suavizado exponencial, no todos ellos poseían bases estadísticas sólidas. Tal era el caso del método multiplicativo de Holt-Winters y no fue sino hasta el año [1997](#) que se desarrollaron los modelos de espacio de estados con innovación y se demostró que los métodos de Brown, Holt y Holt-Winters eran un caso particular de un modelo estocástico más general propuesto por [Ord et al. \(1997\)](#). Gracias a lo anterior se pudo demostrar que era posible crear una base estadística para los modelos de suavizado exponencial que todavía no contaban con una. [Koehler et al. \(2001\)](#) dieron fórmulas analíticas de la media, la varianza, la verosimilitud y otras que juntas forman la base estadística del método multiplicativo de Holt-Winters. En [Hyndman et al. \(2005\)](#) se extendieron los resultados analíticos para la mayoría de los modelos de suavizado exponencial conocidos hasta entonces. Sin embargo, un año después, [Hyndman y Akram \(2006\)](#) demostraron que las bases estadísticas del método multiplicativo no eran válidas debido a un problema con la estabilidad del modelo.

La versión aditiva (de Holt-Winters) de los modelos de suavizado exponencial es la

siguiente

$$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (2.20)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (2.21)$$

$$s_t = \gamma(y_t - l_t) + (1 - \gamma)s_{t-m} \quad (2.22)$$

$$f_{t+h} = l_t + hb_t + s_{t-m+h} \quad (2.23)$$

donde  $l_t$  representa el nivel de la serie (al tiempo  $t$ );  $b_t$  es la pendiente;  $s_t$  es el factor estacionalidad;  $\alpha$ ,  $\beta$  y  $\gamma$  son los parámetros de suavizado exponencial del nivel, la tendencia y la estacionalidad, respectivamente;  $f_{t+h}$  es el pronóstico hecho en el tiempo  $t$  para  $h$  pasos hacia adelante. Utilizando esta notación, podemos escribir la versión multiplicativa del método como sigue

$$l_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (2.24)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (2.25)$$

$$s_t = \gamma \frac{y_t}{l_t} + (1 - \gamma)s_{t-m} \quad (2.26)$$

$$f_{t+h} = (l_t + hb_t)s_{t-m+h} \quad (2.27)$$

Una de las desventajas de los métodos de suavizado exponencial es que sus bases estadísticas suelen perder su validez ya que están construidas sobre los supuestos de normalidad en las desviaciones e información perfecta que implica que los estimadores son los parámetros del modelo.

A la hora de pronosticar series temporales con tendencia lineal y estacionalidad constante o creciente, los modelos de suavizado exponencial realizarán las innecesarias labores de actualizar el nivel, la tendencia y la estacionalidad en cada etapa. Esto es necesario para las series complejas cuyos patrones varían en cada etapa; el caso de series

con tendencia polinomial o estocástica. Sin embargo, estas labores son innecesarias para el caso de series con tendencia lineal y estaciones crecientes pues en estas series se observa que la tendencia no cambia a través del tiempo o de lo contrario, no estaríamos hablando de tendencia lineal sino polinomial o estocástica. Además, en estas series se observa que el crecimiento en la amplitud de las estaciones guarda una relación de proporcionalidad con la tendencia y de esta manera, si la tendencia no cambia a través del tiempo, el incremento en la estacionalidad tampoco lo hará y resultará innecesario actualizar este patrón en cada etapa.

## 2.5. Modelos periódicos

A grandes rasgos, podemos decir que un modelo Periódico Auto Regresivo (PAR) es similar a un modelo autorregresivo con la diferencia de que sus parámetros varían en cada estación. Los primeros trabajos en esta metodología fueron los de [Pagano \(1978\)](#) y [Troutman \(1979\)](#). Posteriormente, [Mcleod \(1993\)](#) y [Vencchia y Ballerini \(1991\)](#) utilizaron este tipo de modelos para medir variables ambientales. Quienes han desarrollado más la parte estadística y teórica de estos métodos han sido [Osborn \(1988\)](#), [Todd \(1990\)](#) y [Hansen y Sargent \(1993\)](#).

Que los parámetros de los modelos PAR cambien en cada estación, abre una rica variedad de nuevas posibilidades en cuanto a modelación se refiere; los modelos PAR han sido ampliamente utilizados para la descripción de variables económicas. Entre algunas de sus aplicaciones se incluyen [Birchenhall et al. \(1989\)](#), [Flores y Novales \(1997\)](#), [Franses \(1993\)](#), [Herwartz \(1997\)](#) y [Osborn \(1988\)](#).

A pesar de su riqueza y demostrada utilidad, los modelos PAR no han sido ampliamente aplicados en economía hasta la fecha. Ciertamente, sus aplicaciones son menos que aquellas relacionadas con las metodologías anteriormente revisadas. Sin embargo, un gran número de estudios muestran que los modelos PAR pueden aparecer de mane-

ra natural a partir de aplicaciones de la teoría económica al modelar decisiones en un contexto económico y por ello, su rol no debe ser despreciado. Por ejemplo, [Gersovitz y MacKinnon \(1978\)](#) y [Osborn \(1988\)](#) argumentan que este tipo de procesos surgen al modelar las decisiones estacionales de los consumidores; en [Hansen y Sargent \(1993\)](#) se sugiere que también pueden aparecer como consecuencia de las tecnologías estacionales. Una vez admitido que las fuerzas económicas rectoras subyacentes a un proceso económico (como lo pueden ser las preferencias o las tecnologías) pueden variar de una estación a otra entonces, sutiles efectos periódicos-estacionales pueden entrar en juego incluso en aquellos contextos generalmente tratados como no estacionales. Ejemplos de este tipo son [Ghysels \(1991\)](#) y [Ghysels \(1994\)](#) donde se explora la naturaleza periódica de los ciclos de negocios en los Estados Unidos.

Para ilustrar el funcionamiento de los modelos PAR, presentamos el PAR(1) para datos con estaciones aditivas

$$y_{s,i} = \phi_s + y_{s-1,i} + \varepsilon_{s,i} \quad (2.28)$$

y la versión para estacionalidad multiplicativa es

$$y_{s,i} = \phi_s y_{s-1,i} + \varepsilon_{s,i} \quad (2.29)$$

El parámetro  $\phi_s$  cambia en cada estación, así de simple es este modelo.

Estudios empíricos han demostrado que la mayoría de las veces, esta metodología conducirá a pronósticos menos precisos que los modelos más comunes [Wells \(1997\)](#) y [Novales y de-Fruto \(1997\)](#). Este es sin duda el peor defecto de esta metodología.

## 2.6. Conclusiones

La estructura de algunas metodologías presenta propiedades irrelevantes tomando en cuenta la simplicidad de los patrones de interés. Las STTLEC no parecen necesitar

por ejemplo, de una ecuación que actualice su pendiente, pues esta exhibe un comportamiento uniforme a lo largo de las observaciones. Por ejemplo, el método multiplicativo de Holt-Winters no sólo ofrece una *innecesaria* ecuación para actualizar la tendencia, sino que también ofrece ecuaciones para actualizar el nivel y las estaciones. Todas estas actualizaciones son adecuadas para las series con tendencia polinomial o estocástica pero son innecesarias para un caso tan simple como el de las STTLEC. Si hubiera que actualizar la pendiente en cada etapa de una STTLEC, no estaríamos hablando de una STTLEC. Por estas razones, utilizar metodologías de este tipo sólo agrega dificultad no justificada al pronóstico.

Otro punto de las metodologías revisadas es que casi todas ellas ofrecen **modelos que suponen que no hay observaciones faltantes** en la serie. Esto se debe al hecho de que los modelos suponen que lo que se desea es pronosticar una serie temporal. Recordando este concepto, sabremos que una serie temporal consiste de sucesiones de variables aleatorias observadas en intervalos de tiempo *igualmente espaciados*. Una serie temporal con observaciones faltantes no es una serie temporal en el sentido estricto pues sus observaciones no se encuentran en intervalos de tiempo igualmente espaciados. Como la mayoría de los modelos fueron hechos para pronosticar series temporales, no serán aptos para pronosticar series con observaciones faltantes. El problema radica en que de una u otra forma, las observaciones futuras y por tanto el pronóstico, son una función de las observaciones pasadas. Pero no podemos conocer el futuro si no conocemos la historia. En este caso, la historia tiene observaciones faltantes que no permiten la estimación de los datos futuros.

Sin embargo, hay un modelo, sólo uno de acuerdo a la revisión hecha de la literatura, que es capaz de abordar el pronóstico de series con observaciones faltantes y pertenece a la metodología ARIMA con filtros diferenciadores y componentes deterministas. Éste es el modelo (2.19). Su capacidad de pronosticar series con observaciones faltantes se debe a que es un modelo altamente determinista y que no utiliza filtros diferenciadores para eliminar patrones como desfases o tendencias sino que lo hace de manera directa con componentes

deterministas. La razón por la que este modelo funciona es su naturaleza determinista. Por ejemplo, siempre es posible ajustar una recta a un conjunto de datos desordenado y de apariencia «nebulosa». Esto se debe a que las rectas no suponen nada sobre la distribución de los datos a lo largo de un eje y por esta razón, pueden ser ajustadas a cualquier conjunto de datos sin importar que sean una serie temporal. Lo mismo ocurre con el modelo (2.19) y con cualquier otro modelo determinista.

Otra de las características de las metodologías revisadas es que en ellas se ofrecen modelos para el pronóstico de datos con **estacionalidad aditiva o estacionalidad multiplicativa**. Es decir, hay que decidir a priori el tipo de estacionalidad que exhibe la serie y con base en ello, debemos aplicar un modelo para uno u otro tipo de estacionalidad. Desde la aparición del concepto de estacionalidad en pronósticos, los estadísticos se han preocupado por averiguar si la serie en cuestión exhibe patrones estacionales aditivos o multiplicativos. Una vez que se decide el tipo de estacionalidad, se elige el método que modele adecuadamente este patrón.

En resumen, de las metodologías estudiadas podemos observar lo siguiente:

1. Pueden ser más complejas de lo necesario;
2. Suponen que no hay datos faltantes, salvo el modelo (2.19);
3. Suponen que la serie posee estacionalidad constante o creciente.

## 2.7. Nuestra propuesta

Nuestra propuesta consta de dos modelos deterministas para el pronóstico de STTLEC con las siguientes ventajas:

1. Los nuevos modelos tienen interpretaciones geométricas simples que no agregan complejidad al pronóstico.

2. Además, al ser modelos deterministas, pueden manejar el problema de pronosticar series con observaciones faltantes sin ningún contratiempo, pero a diferencia del modelo (2.19), los modelos propuestos utilizan muy pocos parámetros para dar un pronóstico, reduciendo la posibilidad de sobreparametrización. Por ejemplo, si deseamos pronosticar una serie con 12 estaciones, los modelos propuestos estimarán 13 y 14 parámetros mientras que el modelo (2.19) estimará 24 parámetros.
3. Uno de los modelos propuestos, tiene la característica de pronosticar series con estacionalidad creciente tan bien como series con estacionalidad constante. Por esta razón, elimina la necesidad de conocer a priori el tipo de estacionalidad que exhibe la serie y elimina los potenciales riesgos de confundir un tipo de estacionalidad con otro.

Como consecuencia de los puntos anteriores, los modelos propuestos son capaces de pronosticar *mejor*. Esto se demuestra en el capítulo 6 de este trabajo, donde se realizan experimentos para comparar los modelos propuestos con la técnica *apta* de la literatura. Las comparaciones incluyen exhaustivas comparaciones estadísticas. El resto de este trabajo se dedica al desarrollo de los puntos anteriores.

---

---

# CAPÍTULO 3

---

## Los modelos propuestos

Los dos modelos que hemos propuesto para el pronóstico de STTLEC serán el tema de este capítulo. El primer modelo, ajusta una recta a la STTLEC siendo la pendiente de esta recta una función de la estación actual. El segundo, ajusta una recta a cada conjunto de datos que representan una estación con la característica de que todas las rectas se interceptan en un lugar común. Se presentarán tres diferentes notaciones para cada modelo y un caso de ejemplo en el que se compararán ambas propuestas.

### 3.1. El primer modelo propuesto

Para pronosticar una serie temporal, es necesario primero suponer un comportamiento subyacente a ella. Es decir, es necesario suponer que hay un modelo matemático capaz de imitar el proceso que estamos observando. Para el caso de las STTLEC, nuestra primera propuesta consiste en ajustar una recta para simular el efecto de tendencia lineal y cuya pendiente variará en cada estación para simular el efecto de estacionalidad creciente. Esta idea se expresa matemáticamente en la ecuación (3.1).

$$y_t = \beta_0 + \beta_s x_t + \mu_t \tag{3.1}$$

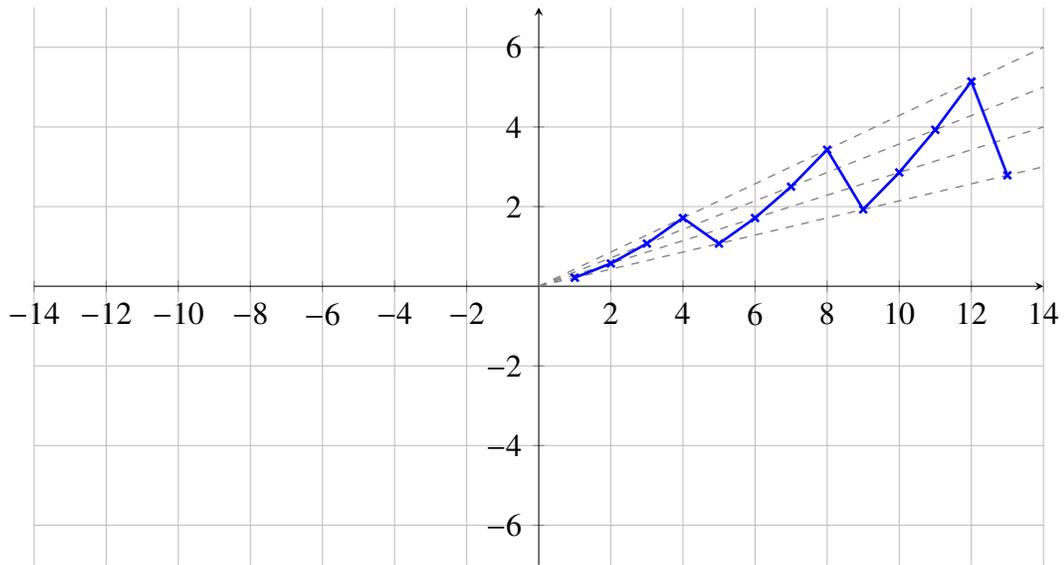
donde  $y_t$  es el dato observado en el tiempo  $t$  con  $t = 1, \dots, n$  (la variable dependiente);  $x_t$  es el valor de la variable independiente en la etapa  $t$ , que es  $x_t = t$  (utilizamos la variable  $x_t$

por que cada observación representa una coordenada cartesiana de la forma  $(x_t, y_t)$ ;  $\beta_0$  es el nivel inicial o desfase entre los datos y el origen;  $\beta_s$  es la pendiente que cambia en cada estación  $s = 1, \dots, m$ ; finalmente,  $\mu_t$  es el error de ajuste en la etapa  $t$  y se comporta como una fuente de variación con media cero y varianza constante y es independiente del resto de los componentes del modelo. Por su naturaleza, el modelo de la ecuación (3.1) es un modelo de recta con pendiente variable al que nos referiremos simplemente como MLIN (léase modelo lineal). En el presente trabajo, se hacen varias abreviaturas como la anterior y por ello, invitamos al lector a consultar el índice de abreviaciones.

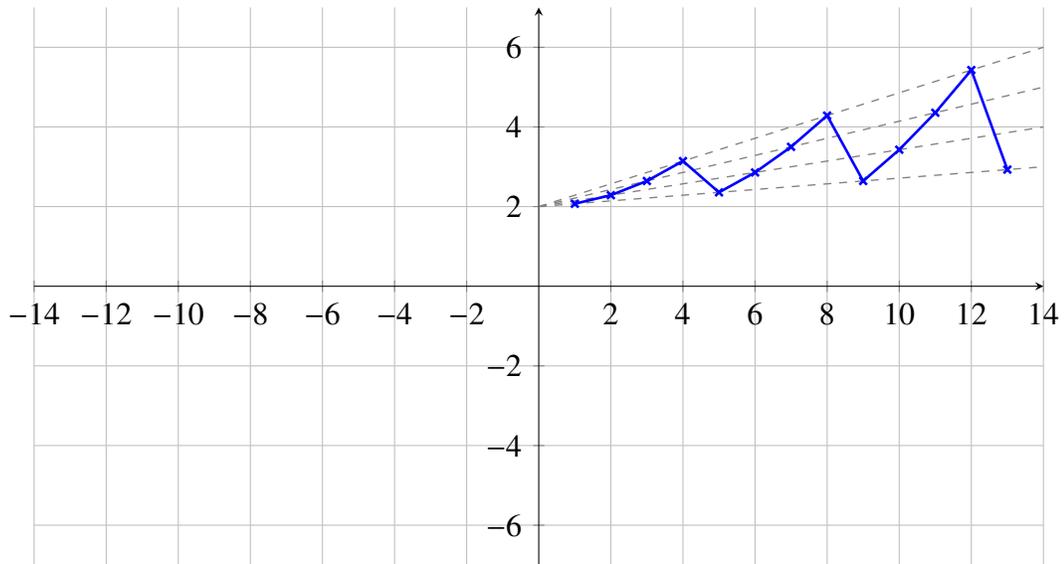
### 3.1.1. Interpretación geométrica

Como ya había sido mencionado, la idea detrás del MLIN es ajustar una línea cuya pendiente varíe con cada estación. Otra manera de ver esta idea se muestra en las figuras 3.1 y 3.2. En dichas gráficas se aprecia que el MLIN es equivalente a ajustar una línea recta a cada conjunto de datos que representa una estación. Es decir, una recta para las observaciones realizadas en el primer trimestre del año, otra línea para las observaciones hechas en el segundo trimestre y así sucesivamente hasta el último trimestre del año. Las rectas se interceptan en el punto en el que el eje de las ordenadas vale cero ( $x_0 = 0$ ). En dicho punto, el valor del eje de las ordenadas por el cual pasan las líneas es siempre igual al desfase que hay entre la serie y el origen ( $\beta_0$ ).

En la gráfica 3.1 se observa que la coordenada de intersección de las rectas es  $(0, 0)$ , mientras que en la gráfica 3.2 es  $(0, 2)$ . Esto muestra que el punto de intersección pasará siempre por la recta  $x_t = 0$ . La explicación se encuentra en la ecuación (3.1) de la cual se puede ver que, siempre que  $x_t$  sea igual a cero, el valor correspondiente de  $y_t$  será  $\beta_0$  para toda  $s$ , es decir, para todas las rectas. En este punto, el modelo supone implícitamente que  $\mu_t = 0$  ya que cero es el valor esperado de la fuente de variación. Estas son las razones por las cuales todas las rectas se interceptan en la coordenada  $(0, \hat{\beta}_0)$ .



**Figura 3.1** – Serie simulada con el MLIN y con intersección en (0,0)



**Figura 3.2** – Serie simulada con el MLIN y con intersección en (0,2)

### 3.1.2. Tres representaciones matemáticas y sus ventajas

En la literatura podemos encontrar tres maneras de representar un modelo de pronóstico. La primera es utilizando *variables binarias*, la segunda es la *notación vectorial* y la tercera es la *notación estacional*. Estas notaciones así como sus ventajas se discuten a continuación.

### Notación de variables binarias

Utilizando variables binarias, podemos representar al MLIN de la siguiente manera

$$y_t = \beta_0 + \left( \sum_{s=1}^m D_{s,t} \beta_s \right) x_t + \mu_t, \quad (3.2)$$

donde  $D_{s,t}$  es la variable binaria que toma el valor de uno si la estación  $s$  coincide con el tiempo  $t$  o cero de otra forma. Esta notación es muy utilizada en el área de econometría y su valor pedagógico es incuestionable.

### Notación vectorial

La notación con variables binarias implica que podemos representar cada una de las observaciones de la siguiente manera

$$\begin{aligned} y_1 &= \beta_0 + \beta_1(1) + \mu_1 \\ y_2 &= \beta_0 + \beta_2(2) + \mu_2 \\ &\vdots \\ y_m &= \beta_0 + \beta_m(m) + \mu_m \\ y_{m+1} &= \beta_0 + \beta_1(m+1) + \mu_{m+1} \\ &\vdots \\ y_{2m} &= \beta_0 + \beta_m(2m) + \mu_{2m} \\ &\vdots \\ &\vdots \\ y_n &= \beta_0 + \beta_m(n) + \mu_n. \end{aligned}$$

Si utilizamos vectores y matrices tenemos

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu} \quad (3.3)$$

donde  $\mathbf{y} = [y_1, y_2, \dots, y_n]'$ ,  $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_m]'$ ,  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_n]'$ ,  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{\lfloor t/m \rfloor}, \mathbf{X}_{\lceil t/m \rceil}]'$  donde cada submatriz  $\mathbf{X}_i$  tiene la forma

$$\mathbf{X}_i = \begin{bmatrix} 1 & (i-1)m+1 & 0 & \dots & 0 \\ 1 & 0 & (i-1)m+2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & im \end{bmatrix}.$$

Lo anterior es válido para  $i = 1, \dots, \lfloor t/m \rfloor$  mientras que  $\mathbf{X}_{\lceil t/m \rceil}$  se define como

$$\mathbf{X}_{\lceil t/m \rceil} = \begin{bmatrix} 1 & (\lceil t/m \rceil - 1)m + 1 & 0 & \dots & 0 & \dots & 0 \\ 1 & 0 & (\lceil t/m \rceil - 1)m + 2 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & n & \dots & 0 \end{bmatrix}.$$

Es decir,  $\mathbf{X}_{\lceil t/m \rceil}$  es como cualquier  $\mathbf{X}_i$  salvo por el detalle de que las filas de  $\mathbf{X}_{\lceil t/m \rceil}$  llegan hasta donde haya datos; es decir, el número de filas de  $\mathbf{X}_{\lceil t/m \rceil}$  es  $\text{mod}(t, m)$  mientras que el de cualquier otra submatriz  $\mathbf{X}_i$  es siempre  $m$ . Lo anterior se hace con el fin de abarcar los casos en los que el número de datos no es un múltiplo de  $m$ . Si el número de datos si es divisible entre  $m$  entonces, simplemente se saca  $\mathbf{X}_{\lceil t/m \rceil}$  de  $\mathbf{X}$  y se redefine ésta como  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{\lfloor t/m \rfloor}]'$ . La notación vectorial tiene la ventaja de ser más compacta que la notación de variables binarias.

### Notación estacional

Una manera poco convencional pero muy útil de representar un modelo de pronóstico es ver a su índice temporal  $t$  como una función de períodos y estaciones. En el cuadro 3.1 aparecen tabulados los datos de una serie con estacionalidad.

Las columnas del cuadro 3.1 están indexadas por  $s$  (con  $s = 1, \dots, m$  siendo  $m$  el número de estaciones) y representan las estaciones (meses, cuatrimestres, etc). Las filas están indexadas por  $i$  con  $i = 1, \dots, k_s$  (donde  $k_s$  es el último año para el cual hay

**Cuadro 3.1** – Datos tabulados en función de períodos y estaciones.

Períodos $i$	Estaciones $s$			
	1	2	...	$m$
1	$y_{1,1}$	$y_{1,2}$	...	$y_{1,m}$
2	$y_{2,1}$	$y_{2,2}$	...	$y_{2,m}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$k_s$	$y_{k_s,1}$	$y_{k_s,2}$	...	$y_{k_s,m}$

observaciones de la estación  $s$ ) y representan los períodos (normalmente expresados en términos de años). Del cuadro 3.1 se deduce que la representación del tiempo del MLIN como una función de los períodos y las estaciones es  $t = (i - 1)m + s$ ; las observaciones  $y_t$  se escriben como  $y_{(i-1)m+s}$  sólo que por brevedad emplearemos el termino  $y_{i,s}$  (lo mismo se aplica para  $x_{i,s}$ ); los parámetros  $\beta_s$  se representan de esta forma ya que éstos dependen únicamente de las estaciones. Después de los cambios, la ecuación (3.1) queda así

$$y_{i,s} = \beta_0 + \beta_s x_{i,s} + \mu_{i,s} \quad (3.4)$$

La reescritura del modelo, no sólo ayuda a conocer el período y la estación donde se ubica cada observación, sino que también ayuda a facilitar la estimación de parámetros y por esta razón será utilizada por el resto de este trabajo.

### 3.1.3. Pronóstico con el primer modelo: caso de ejemplo

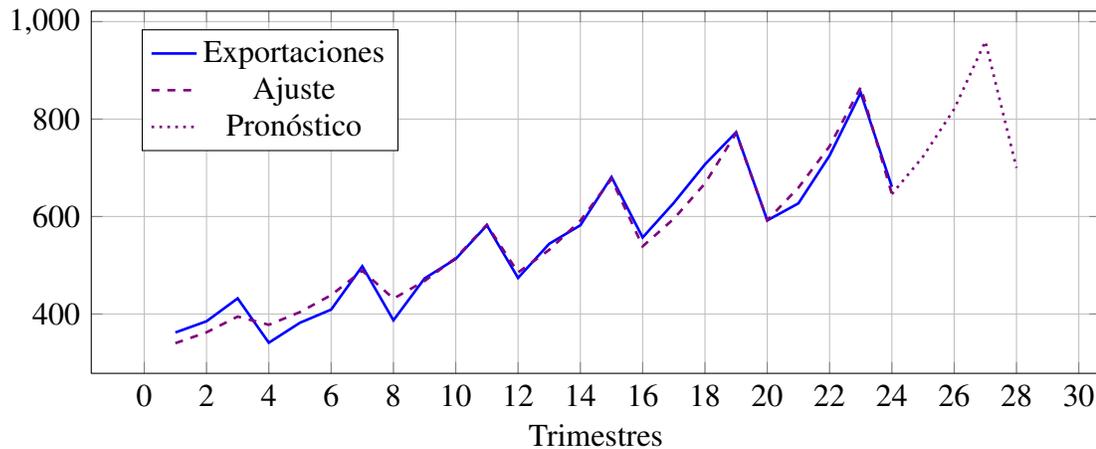
Los datos del cuadro 3.2 representan las exportaciones trimestrales de una compañía francesa. Estos datos fueron tomados del texto clásico de [Makridakis et al. \(1998, pág. 162\)](#). Las estaciones están dadas por los trimestres del año y por ello, hay cuatro estaciones en cada período (año).

En este ejemplo, las estaciones están dadas por trimestres. En otros casos, pueden estar dadas por semanas, meses, bimestres, etc. De igual manera, los períodos pueden ser meses, semestres, sexenios, etc. Después de estimar los valores de los parámetros del MLIN

**Cuadro 3.2** – Exportaciones trimestrales de una compañía francesa

Años <i>i</i>	Trimestres <i>j</i>			
	1	2	3	4
1	362	385	432	341
2	382	409	498	387
3	473	513	582	474
4	544	582	681	557
5	628	707	773	592
6	627	725	854	661

(la estimación se presenta en el siguiente capítulo), se obtiene el pronóstico que se muestra en la figura 3.3.



**Figura 3.3** – Ajuste y pronóstico con el MLIN

Según podemos observar en la figura 3.3, el MLIN no sólo sigue muy bien a la serie de datos durante el período de ajuste, sino que también arroja un pronóstico intuitivamente esperado, por que la tendencia del pronóstico sigue creciendo y las amplitudes de los ciclos se amplían, justo como se comportan los datos de la serie. Un indicio más numérico que intuitivo es la *suma de desviaciones cuadradas* que fue de 511 para el MLIN en el período de ajuste y es 17 veces menor al que resulta de seguir la metodología del libro de texto (el método multiplicativo de Holt-Winters) que fue de 8743,8. Un error tan pequeño podría implicar que el MLIN *sobreparametriza* la serie, pero esto no es probable ya que el número de parámetros que este método utiliza es menor que los del método de Holt-Winters.

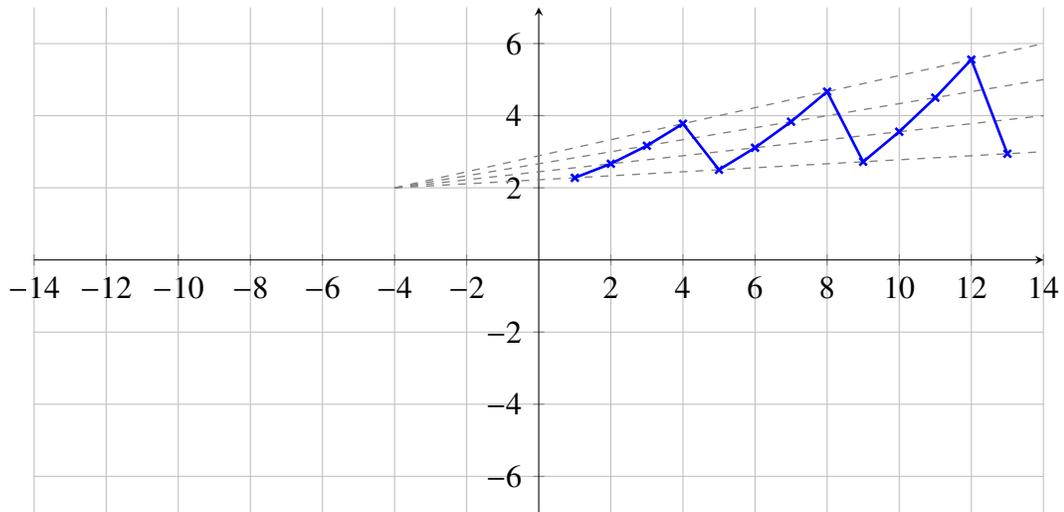
## 3.2. El segundo modelo propuesto

El  $MLIN$  es un modelo simple para el pronóstico de STTLEC. La idea que lo subyace es sencilla y el modelo da buenos resultados frente a metodologías ya establecidas como el método multiplicativo de Holt-Winters. Según lo mencionado en la interpretación geométrica, sabemos que el  $MLIN$  ajusta una recta a cada conjunto de datos y que las rectas se interceptan en el eje de las ordenadas ( $x_t = 0$ ). Lo anterior resulta restrictivo y podría ocasionar que el  $MLIN$  modele los primeros períodos de observaciones con una amplitud excesivamente pequeña ya que el punto de intersección estará siempre muy cerca de esos períodos. Por esta razón, se hace necesario proponer un modelo cuyo punto de intersección no se limite a los valores sobre el eje de las ordenadas, sino que pueda moverse a lo largo del plano. Dicho modelo es el tema de esta sección. En las próximas subsecciones discutiremos su interpretación geométrica, notaciones matemáticas y presentaremos una comparación de su pronóstico contra el del  $MLIN$ .

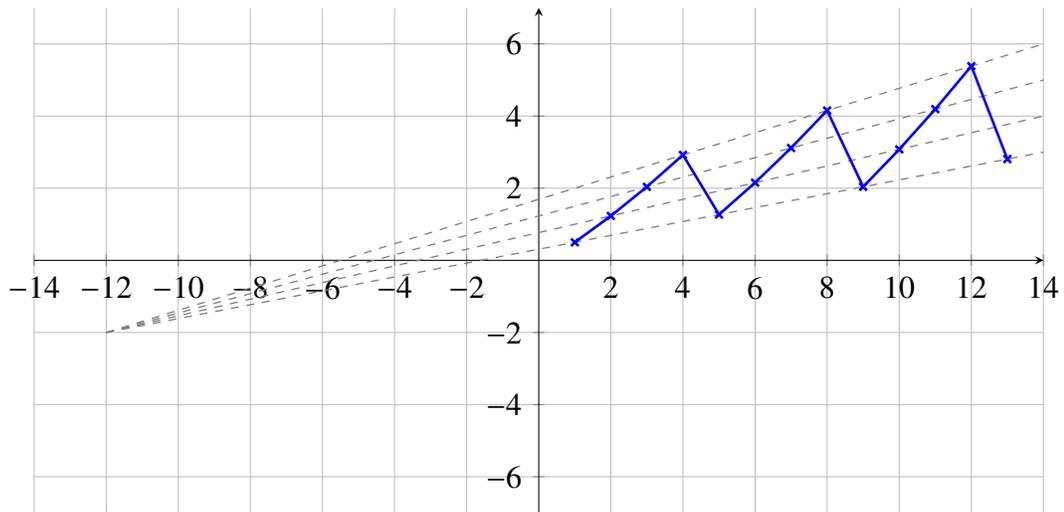
### 3.2.1. Interpretación geométrica

Para prevenir la problemática causada por el hecho de que las rectas se interceptan siempre en el eje de las ordenadas, nosotros proponemos un modelo similar al  $MLIN$  en cuanto a que ajusta una recta a cada estación pero con la diferencia de que las líneas se interceptan en cualquier punto del plano y no sólo en el eje de las ordenadas. En pocas palabras, lo que proponemos es un modelo que busca un punto en el plano y a partir de él se trazan rectas a cada uno de los conjuntos de datos que representan una estación. En las figuras 3.4 y 3.5 se muestra esta idea.

Siguiendo esta idea, supongamos que se quiere ajustar una línea recta a cada conjunto de datos que representan una estación. Si tenemos (como lo hemos supuesto) un conjunto de  $m$  líneas rectas  $y_{s,t} = \alpha_s + \beta_s x_{s,t}$  entonces, estas rectas tendrán un punto de intersección común si todas pasan por el punto  $(\tau, \beta_0)$ . En dicho punto tenemos que  $y_{1,\tau} = y_{2,\tau} = \dots =$



**Figura 3.4** – Serie generada con el segundo modelo y punto de intersección (-4,2)



**Figura 3.5** – Serie generada con el segundo modelo y punto de intersección (-12,-2)

$y_{m,\tau} = \beta_0$ . Lo anterior conduce a que  $\alpha_s = \beta_0 - \beta_s\tau$ . Es decir, los desfases de cada línea son ahora una función de cada una de sus pendientes. Al sustituir en las  $m$  líneas rectas tenemos  $y_{s,t} = \beta_0 + \beta_s(x_{s,t} - \tau)$ . Por cada  $t$  existen  $m$  posibles valores para  $y_{s,t}$  y debemos seleccionar aquel que corresponda a la recta cuya  $s$  sea igual a  $\text{mod}(t - 1, m) + 1$  ya que esta línea (más específicamente, la pendiente de esta línea) es la que corresponde a la estación que se observa en la etapa  $t$ . Lo anterior es simplemente tomar la línea que desde un principio se quería ajustar a la estación  $s$ . Para que el modelo esté terminado, falta añadir una fuente de

variación. Al hacerlo nos queda

$$y_{s,t} = \beta_0 + \beta_s(x_{s,t} - \tau) + \mu_{s,t}, \quad (3.5)$$

donde  $\mu_{s,t}$  es un proceso con media cero y varianza constante. Debido a su naturaleza, la ecuación (3.5) es un modelo de rectas con intersección en el plano al que haremos referencia simplemente como MNoLIN (léase modelo no lineal). A continuación, veremos tres notaciones matemáticas de uso común en la literatura para representar de manera más adecuada al MNoLIN.

### 3.2.2. Tres representaciones matemáticas y sus ventajas

#### Notación estacional

Utilizando la notación estacional, podemos representar al MNoLIN así

$$y_{s,i} = \beta_0 + \beta_s(x_{s,i} - \tau) + \mu_{s,i}. \quad (3.6)$$

La ecuación (3.6) muestra que el MNoLIN es equivalente al MLIN salvo por el detalle de que el tiempo es desplazado  $\tau$  unidades ( $x_{s,i} - \tau$ ). Cuando  $\tau$  es igual a cero, el MNoLIN es equivalente al MLIN.

#### Notación con variables binarias

También podemos reescribir el MNoLIN con variables binarias de la siguiente manera

$$y_t = \beta_0 + \left( \sum_{s=1}^m \beta_s D_{s,t} \right) (x_t - \tau) + \mu_t, \quad (3.7)$$

### Notación vectorial

Finalmente, la notación vectorial para el MNoLIN quedaría de la siguiente manera

$$\mathbf{y} = (\mathbf{X} - \tau \mathbf{D})\boldsymbol{\beta} + \boldsymbol{\mu}, \quad (3.8)$$

donde  $\mathbf{y}$ ,  $\mathbf{X}$ ,  $\boldsymbol{\beta}$  y  $\boldsymbol{\mu}$  se definen de manera similar a los de la ecuación 3.3 mientras que  $\mathbf{D}$  se define como  $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n]'$  donde  $\mathbf{d}_t$  se define a su vez como  $\mathbf{d}_t = [0|0, \dots, 1, \dots, 0]$  es decir, la primera partición del renglón  $\mathbf{d}_t$  consistirá de un elemento igual a cero mientras que en la segunda partición, todos los  $m$  elementos tomarán el valor de cero salvo el  $(\text{mod}(t - 1, m) + 1)$ -ésimo que tomará el valor de uno.

### 3.2.3. Pronóstico con el segundo modelo: comparación

En la figura 3.6 se muestra el ajuste y pronóstico del MNoLIN a las exportaciones de la compañía francesa. El MNoLIN se ajustó muy bien a las amplitudes de las estaciones en todos los años. Esto es especialmente notorio en el primer y segundo año donde prácticamente no hubo diferencia de ajuste entre estos períodos y el resto, a diferencia de lo que ocurrió con el MLIN (ver figura 3.3).

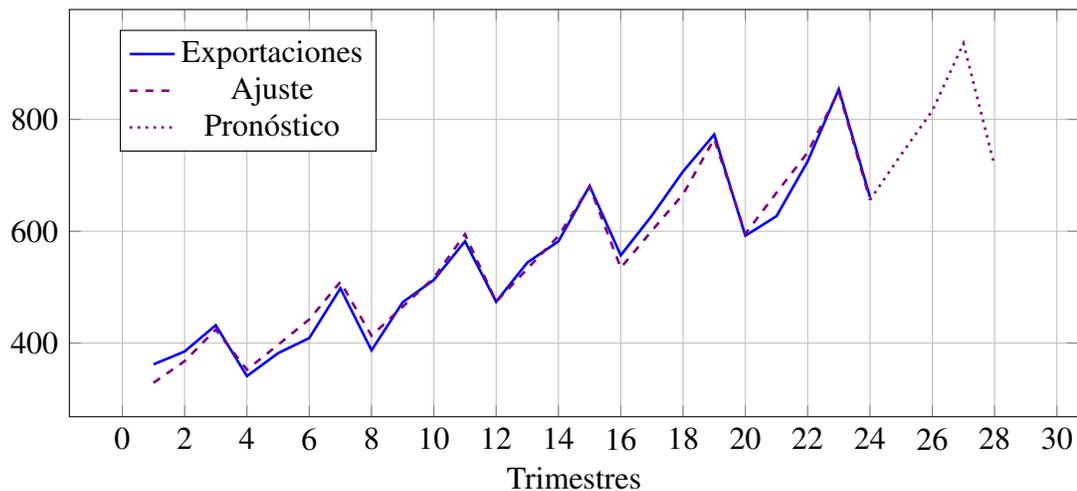


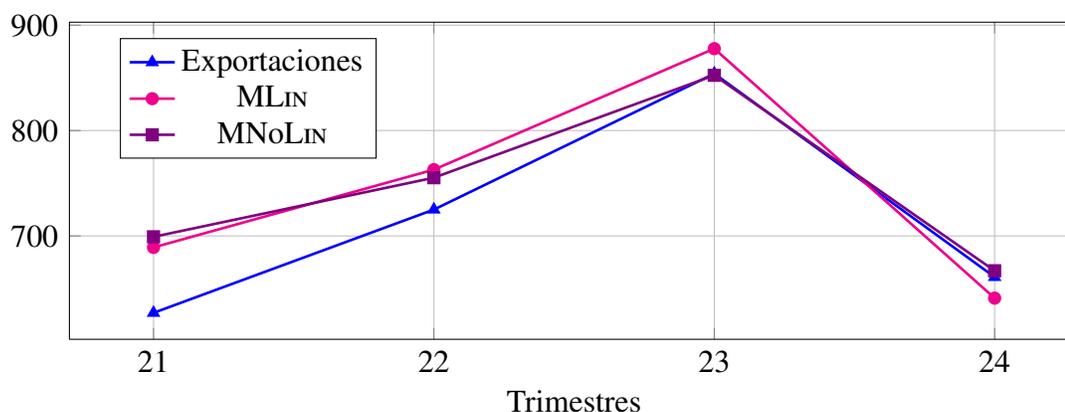
Figura 3.6 – Ajuste y pronóstico con el MNoLIN.

Para hacer una comparación entre el MLIN y el MNoLIN, decidimos pronosticar el último período de la serie y comparar el error de pronóstico entre ambos modelos. En el cuadro 3.2.3 se muestran los resultados y en la figura 3.7 se gráfica esta información.

**Cuadro 3.3** – Pronósticos con los modelos propuestos.

Trimestres	Observaciones	Modelos		Mejor modelo
		MLIN	MNoLIN	
21	627	689.1	699.1	MLIN
22	725	762.9	755.2	MNoLIN
23	854	877.7	852.4	MNoLIN
24	661	641.0	666.9	MNoLIN

En el cuadro 3.2.3 se aprecia que el MLIN pronosticó mejor el trimestre número 21. La observación correspondiente a este trimestre se comporta como un evento raro ya que ninguno de los dos pronósticos estuvo cerca de dicha observación. En la figura 3.7 se observa que la distancia entre los pronósticos y la observación correspondiente al trimestre número 21 es más grande que el resto. Este fenómeno es lo inverso a lo que intuitivamente se esperaría; entre más lejos está el dato a pronosticar, mayor será el error de pronóstico. Por estas razones, creemos que la observación número 21 no es muy representativa de la serie. Para el resto de las observaciones, no se observaron anomalías de este tipo y el mejor método de pronóstico fue el MNoLIN.



**Figura 3.7** – Pronósticos del MLIN y del MNoLIN en un período de prueba.

### 3.3. Conclusiones

En este capítulo, se han presentado dos modelos para el pronóstico de STTLEC. El primer modelo, consiste en una recta cuya pendiente varía de acuerdo a la estación correspondiente a cada  $t$ . Las ventajas de este primer modelo son su simplicidad pero sobre todo su linealidad. Su principal desventaja es lo restrictivo que resulta el hecho de que cada recta que representa una estación se intercepte con las otras rectas en algún punto del eje de las ordenadas.

El segundo modelo es una generalización del primero y consiste en ajustar una recta a cada estación del año y suponer que todas las rectas se interceptan en algún lugar del plano y no sólo en algún lugar del eje de las ordenadas. La principal ventaja de este modelo es su capacidad de abarcar un mayor número de STTLEC. Su principal desventaja radica en su no linealidad. Esto complica la estimación de parámetros al igual que la caracterización estadística para este modelo.

Según el estudio comparativo, el modelo no lineal pronosticó mejor en tres de cuatro ocasiones al compararlo contra el modelo lineal. Sin embargo, será necesario pronosticar más series para obtener resultados más confiables respecto a cual modelo de pronóstico es mejor, pero antes, en la siguiente sección se discutirá algo que quedó pendiente en esta; la estimación de parámetros.

---

---

# CAPÍTULO 4

---

## Estimación de parámetros

En este capítulo, se mostrará la manera de estimar los parámetros del MLIN y del MNoLIN. Debido a que el MLIN es un modelo de regresión lineal, la estimación de sus parámetros bajo criterios tradicionales se puede realizar con facilidad. Sin embargo, este no es el caso del MNoLIN ya que es un modelo de regresión no lineal pues sus parámetros se multiplican entre sí. Para este último modelo, se proponen dos maneras de estimar sus parámetros. Al final, se presentarán resultados de experimentos que indican cual de las dos estimaciones es la mejor.

### 4.1. Estimación de parámetros para el primer modelo

Los estimadores *mínimos cuadrados ordinarios* (MCO) son aquellos que minimizan la *suma de las desviaciones cuadradas* (SDC). Dicha suma es una función que resulta de mirar los parámetros del modelo como variables a estimar. Los estimadores MCO se obtienen al resolver el siguiente problema de optimización

$$\min_{\beta_0, \beta_1, \dots, \beta_m} \text{SDC} = \sum_{s=1}^m \sum_{i=1}^{k_s} \mu_{i,s}^2 = \sum_{s=1}^m \sum_{i=1}^{k_s} (y_{i,s} - (\beta_0 + \beta_s x_{i,s}))^2 \quad (4.1)$$

A continuación se deducen los valores críticos de  $\beta_0$  y  $\beta_s \forall s$ . En el Apéndice A se demuestra que los valores críticos minimizan la SDC.

Para obtener el valor crítico de  $\beta_0$  es necesario obtener la derivada parcial de la SDC

con respecto a este parámetro e igualar a cero como se muestra a continuación

$$\frac{\partial \text{SDC}}{\partial \beta_0} = -2 \sum_{s=1}^m \sum_{i=1}^{k_s} (y_{i,s} - (\beta_0 + \beta_s x_{i,s})) = 0, \quad (4.2)$$

al despejar tenemos que el valor crítico de  $\beta_0$  está dado por

$$\hat{\beta}_0 = \frac{1}{n} \sum_{s=1}^m \sum_{i=1}^{k_s} y_{i,s} - \frac{1}{n} \sum_{s=1}^m \sum_{i=1}^{k_s} \beta_s x_{i,s}. \quad (4.3)$$

repetiendo lo mismo para  $\beta_s$  tenemos

$$\frac{\partial \text{SDC}}{\partial \beta_s} = -2 \sum_{i=1}^{k_s} (y_{i,s} - (\beta_0 + \beta_s x_{i,s})) x_{i,s} = 0, \quad \forall s, \quad (4.4)$$

$$\hat{\beta}_s = \frac{\sum_{i=1}^{k_s} (y_{i,s} - \beta_0) x_{i,s}}{\sum_{i=1}^{k_s} x_{i,s}^2}, \quad \forall s. \quad (4.5)$$

Al evaluar  $\hat{\beta}_0$  en este punto (sustituyendo  $\beta_s$  de la ecuación (4.3) por  $\hat{\beta}_s$  de la ecuación (4.5)) y despejando  $\hat{\beta}_0$  tenemos

$$\hat{\beta}_0 = \frac{\sum_{s=1}^m \sum_{i=1}^{k_s} y_{i,s} w_{i,s}}{\sum_{s=1}^m \sum_{i=1}^{k_s} w_{i,s}}, \quad (4.6)$$

donde

$$w_{i,s} = 1 - x_{i,s} \frac{\sum_{j=1}^{k_s} x_{j,s}}{\sum_{j=1}^{k_s} x_{j,s}^2}, \quad \forall i, \quad \forall s. \quad (4.7)$$

El estimador  $\hat{\beta}_0$  es una función de los datos conocidos  $x_{i,s}$  y  $y_{i,s} \forall i, s$ . Al sustituir (4.6) en (4.5) tenemos

$$\hat{\beta}_s = \frac{\sum_{i=1}^{k_s} y_{i,s} x_{i,s}}{\sum_{i=1}^{k_s} x_{i,s}^2} - \frac{\sum_{i=1}^{k_s} x_{i,s}}{\sum_{i=1}^{k_s} x_{i,s}^2} \cdot \frac{\sum_{s=1}^m \sum_{i=1}^{k_s} y_{i,s} w_{i,s}}{\sum_{s=1}^m \sum_{i=1}^{k_s} w_{i,s}}, \quad \forall s. \quad (4.8)$$

Si utilizamos la «mucho más compacta» notación vectorial tenemos que los estimadores MCO de  $\beta$  estarán dados por

$$\hat{\beta} = [\mathbf{X}' \mathbf{X}]^{-1} \mathbf{X}' \mathbf{y} \quad (4.9)$$

## 4.2. Estimación de parámetros para el segundo modelo

El MNoLIN pertenece a la clase de los llamados *modelos de regresión no lineal* debido a que sus parámetros se multiplican entre sí. En este contexto, los estimadores que minimizan la SDC reciben el nombre de *mínimos cuadrados no lineales* (MCNL). Generalmente, no hay una manera fácil ni desarrollos analíticos para obtener los estimadores MCNL y se vuelve necesario utilizar métodos numéricos para encontrarlos.

En la primera parte de esta sección, revisaremos el algoritmo estándar para encontrar los estimadores MCNL, el algoritmo de Levenberge-Marquardt; en la segunda parte, proponemos un algoritmo simple y eficiente para la estimación de los parámetros del MNoLIN que explota la estructura del modelo; al final, se realiza un experimento para comparar la eficacia de ambos procedimientos de optimización.

### 4.2.1. El algoritmo de Levenberg-Marquardt

El algoritmo de Levenberg-Marquardt (al que haremos referencia en esta sección como LEV\_MARQ) es una técnica iterativa para localizar el valor mínimo de funciones de SDC no lineales y de múltiples variables. Esta técnica es considerada la manera estándar de encontrar los estimadores MCNL y por ello se utiliza no sólo en pronósticos sino en varias disciplinas.

El algoritmo de LEV\_MARQ puede ser visto como una combinación del algoritmo de descenso acelerado y del método de Gauss-Newton. Si la solución inicial se encuentra lejos del valor óptimo, el procedimiento de LEV\_MARQ se comporta como el algoritmo de descenso acelerado, el cual trabaja eficientemente si el valor óptimo está lejos. Si la solución inicial está cerca del óptimo, el algoritmo de LEV\_MARQ se comporta como el de Gauss-Newton; más eficiente entre más cerca del óptimo se encuentre. Los pasos del algoritmo de LEV\_MARQ son los siguientes:

**Algoritmo de Levenberg-Marquardt:**  $\mathbf{P} = \text{LEV\_MARQ}(\mathbf{y})$

*Paso 1:* Inicializar:

$l$ , parámetro de Marquardt.

$\alpha$ , magnitud del paso de búsqueda.

$\varepsilon$ , precisión.

$\mathbf{P}$ , vector de parámetros dado por  $\mathbf{P} = [\tau, \beta_0, \beta_1, \dots, \beta_m]'$  ó  $\mathbf{P} = [\tau, \beta']'$ .

$\mathbf{J}$ , matriz jacobiana definida como  $\mathbf{J} = [\mathbf{J}_1, \dots, \mathbf{J}_{\lfloor n/m \rfloor}, \mathbf{J}_{\lceil n/m \rceil}]'$  donde cada  $\mathbf{J}_r$  para  $r = 1, \dots, \lfloor n/m \rfloor$  tiene la forma

$$\mathbf{J}_r = \begin{bmatrix} -\beta_1 & 1 & g_\tau(1) & 0 & \dots & 0 \\ -\beta_2 & 1 & 0 & g_\tau(2) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ -\beta_m & 1 & 0 & 0 & \dots & g_\tau(m) \end{bmatrix},$$

con  $g_\tau(s) = (r-1)m + s - \tau$ . La matriz  $\mathbf{J}_{\lceil n/m \rceil}$  está dada por

$$\mathbf{J}_{\lceil n/m \rceil} = \begin{bmatrix} -\beta_1 & 1 & g_{\lceil n/m \rceil}(1) & 0 & \dots & 0 & \dots & 0 \\ -\beta_2 & 1 & 0 & g_{\lceil n/m \rceil}(2) & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 & \ddots & \vdots \\ -\beta_{\lceil n/m \rceil} & 1 & 0 & 0 & \dots & n - \tau & \dots & 0 \end{bmatrix}.$$

*Paso 2:* Resolver  $(\mathbf{J}'\mathbf{J} + \mathbf{I})h = -\text{SDC}$  para encontrar  $h$ , la dirección de búsqueda;

*Paso 3:*  $\mathbf{P} := \mathbf{P} + \alpha h$ ;

*Paso 4:* Actualizar  $l$ ;

*Paso 5:* **si**  $\|\alpha h\| > \varepsilon \|P\|$ ,

**entonces ir a Paso 2;**

**si no ir a Paso 6;**

*Paso 6:* **devolver P**

Muchos detalles de los pasos anteriores pueden ser refinados y en especial aquellos relacionados con la inicialización de  $\alpha$  y  $l$  y la actualización de este último parámetro. Nosotros utilizaremos la versión del algoritmo de LEV\_MARQ codificada por Richard Shrager, Arthur Jutan y Ray Muzic. Ellos programaron el algoritmo de LEV\_MARQ en Octave, un language de programación altamente compatible con Matlab. El código se encuentra en el sitio <http://octave.sourceforge.net/optim/function/leasqr.html>.

El vector de parámetros fue iniciado de la forma « $\mathbf{P}_0 = [1, 1, \bar{\beta}_1, \dots, \bar{\beta}_m]'$ » donde  $\bar{\beta}_s \forall s$  son los estimadores de las pendientes resultantes de regresar una recta sobre cada conjunto de datos que representa una estación, la justificación de este procedimiento se encuentra en la sección 5.4 del capítulo 5; el tamaño mínimo del paso de búsqueda fue « $\text{minstep} = 0,001\mathbf{e}$ » donde  $\mathbf{e}$  es un vector de unos de  $(m + 2) \times 1$ ; el máximo tamaño del paso de búsqueda fue « $\text{minstep} = \mathbf{e}$ »; el máximo número de iteraciones fue de « $\text{niter} = 50$ »; finalmente, la tolerancia de mejora fraccional en la suma escalar de cuadrados fue « $\text{stol} = 0,001$ ».

#### 4.2.2. El algoritmo de estimación diseñado para este modelo

Para encontrar los estimadores MCNL mediante el algoritmo de LEV\_MARQ, es necesario minimizar una función no lineal de  $m + 2$  variables. Para realizar esta labor de manera más eficiente, nosotros proponemos un algoritmo de optimización que saca ventaja de la estructura de la función. El procedimiento propuesto consiste de dos fases. En la primera se realiza una *búsqueda lineal* y en la segunda se aplica el algoritmo de la *sección dorada*. Antes de ver detalladamente los procedimientos de cada fase, estableceremos primero que encontrar los estimadores MCNL equivale a solucionar el siguiente problema de optimización no lineal

$$\min_{\tilde{\tau}, \tilde{\beta}} \text{SDC} = \tilde{\mu}'\tilde{\mu} = (\mathbf{y} - (\mathbf{X} - \tilde{\tau}\mathbf{D})\tilde{\beta})'(\mathbf{y} - (\mathbf{X} - \tilde{\tau}\mathbf{D})\tilde{\beta}) \quad (4.10)$$

Sacando las derivadas parciales e igualando a cero tenemos

$$\tilde{\beta}_{\min} = [(\mathbf{X} - \tilde{\tau}\mathbf{D})'(\mathbf{X} - \tilde{\tau}\mathbf{D})]^{-1}(\mathbf{X} - \tilde{\tau}\mathbf{D})'\mathbf{y}, \quad (4.11)$$

$$\tilde{\tau}_{\min} = [(\mathbf{D}\tilde{\beta})'(\mathbf{D}\tilde{\beta})]^{-1}(\mathbf{D}\tilde{\beta})'(\mathbf{X}\tilde{\beta} - \mathbf{y}). \quad (4.12)$$

Las variables del vector  $\tilde{\beta}_{\min}$  son a su vez una función de la variable  $\tilde{\tau}$  y la variable  $\tilde{\tau}_{\min}$  es una función del vector  $\tilde{\beta}$ . Resulta conveniente ver a la SDC como una función de  $\tilde{\tau}$  ya que de esta manera, el problema de optimización original dado por la ecuación (4.10) y en el que era necesario estimar  $m + 2$  variables, se reduce a tener que estimar sólo una. El problema (4.10) equivale a

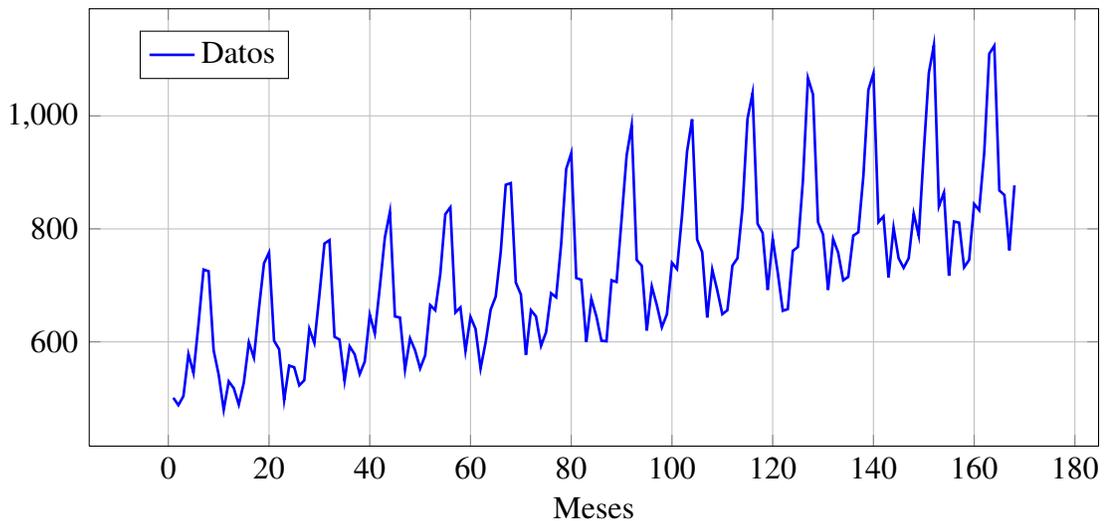
$$\min_{\tilde{\tau}} \text{SDC}_{\min} = \tilde{\mu}_{\min}' \tilde{\mu}_{\min} \quad (4.13)$$

donde

$$\tilde{\mu}_{\min} = (\mathbf{y} - (\mathbf{X} - \tilde{\tau} \mathbf{D}) \tilde{\beta}_{\min}), \quad (4.14)$$

$$= (\mathbf{y} - (\mathbf{X} - \tilde{\tau} \mathbf{D}) [(\mathbf{X} - \tilde{\tau} \mathbf{D})' (\mathbf{X} - \tilde{\tau} \mathbf{D})]^{-1} (\mathbf{X} - \tilde{\tau} \mathbf{D})' \mathbf{y}). \quad (4.15)$$

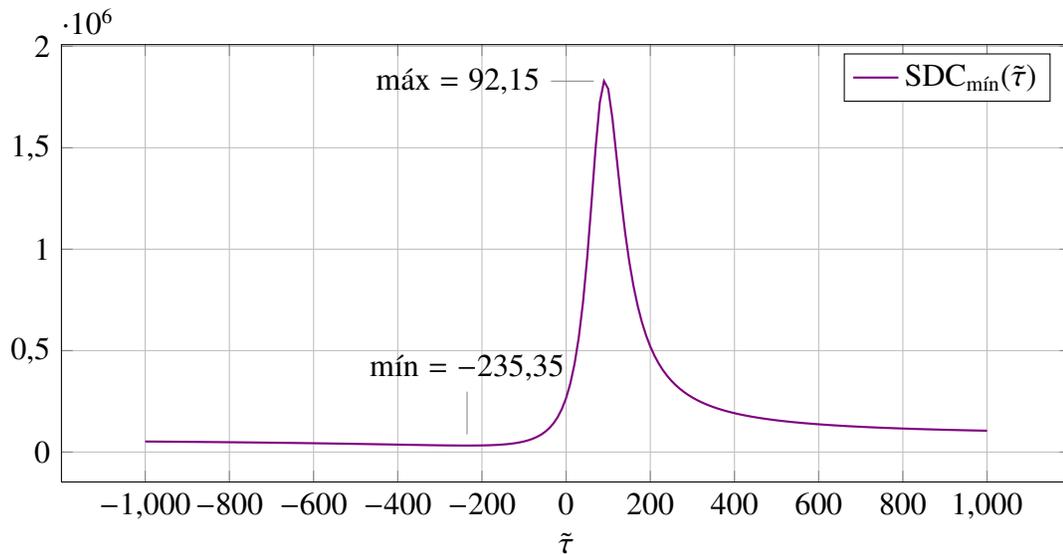
Para ilustrar las vicisitudes de la estimación de parámetros, utilizaremos como serie de ejemplo los datos de habitaciones de un hotel tomados de [Bowerman y O'Connell \(1993\)](#) y que se muestran en la figura 4.1. Se observa que la serie posee tendencia lineal y estacionalidad creciente.



**Figura 4.1** – Datos de habitaciones de un hotel.

En la figura 4.2 se muestra la función  $\text{SDC}_{\min}$  (en cientos de miles) como una función de  $\hat{\tau}$  para los datos de habitación de hotel. En la figura 4.2 se puede apreciar que  $\text{SDC}_{\min}$

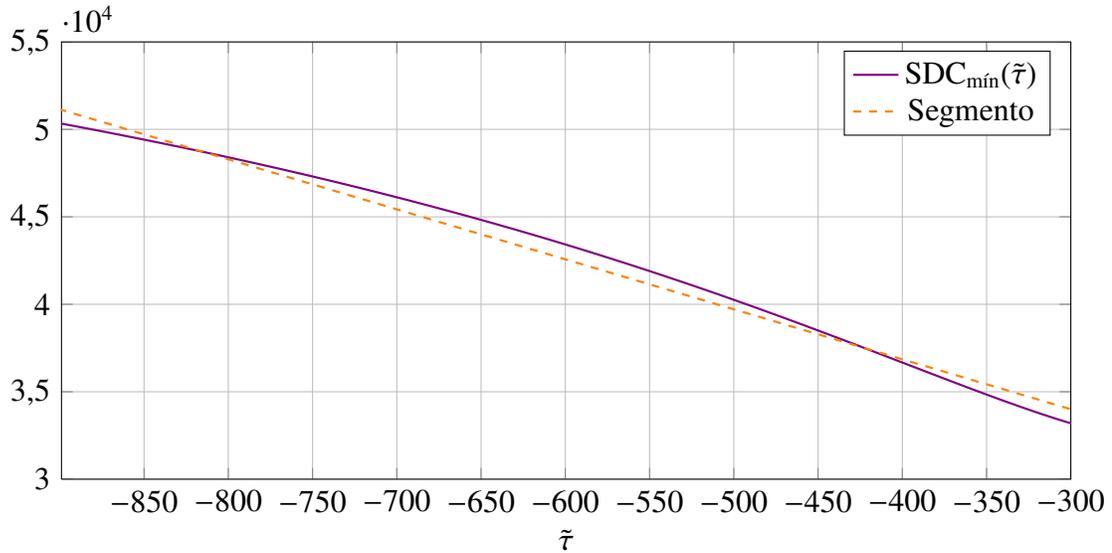
no es una función *convexa*, tiene un valor máximo en  $\tilde{\tau} = 92,15$  y un valor mínimo en  $\tilde{\tau} = -235,35$ . Al valor de  $\tilde{\tau}$  que minimiza la  $SDC_{\min}$  lo representaremos como  $\hat{\tau}$  ya que esta es la notación común para hacer referencia a estimadores que minimizan la SDC. También haremos referencia al vector  $\tilde{\beta}_{\min}$  evaluado en  $\tilde{\tau} = \hat{\tau}$  simplemente como  $\hat{\beta}$ .



**Figura 4.2** – La SDC como función de  $\tilde{\tau}$  ( $SDC_{\min}(\tilde{\tau})$ ).

Como ya ha sido mencionado anteriormente, la función  $SDC_{\min}$  no es convexa en el intervalo  $\tilde{\tau} \in [-1000, 1000]$ . Esto resulta obvio al mirar la figura 4.2. Sin embargo, al mirar la figura 4.2 podría parecer que la función si es convexa en el intervalo  $\tilde{\tau} \in [-1000, \text{máx} = 92,15]$  ya que en este intervalo no se observan irregularidades. Lo mismo podríamos decir sobre el intervalo  $\tilde{\tau} \in [\text{máx} = 92,15, 1000]$ . Si los intervalos  $\tilde{\tau} \in [-1000, \text{máx} = 92,15]$  y  $\tilde{\tau} \in [\text{máx} = 92,15, 1000]$  fueran convexos, podríamos aplicar un algoritmo en ellos para la búsqueda del óptimo en regiones convexas. Sin embargo, la función  $SDC_{\min}(\tilde{\tau})$  no es convexa en los intervalos  $\tilde{\tau} \in [-1000, \text{máx} = 92,15]$  y  $\tilde{\tau} \in [\text{máx} = 92,15, 1000]$ . En la figura 4.3 se muestra este hecho para un intervalo dentro de  $[-1000, 92,15]$ ;  $\tilde{\tau} \in [-900, -300]$ . Lo que se observa es que el segmento de recta con inicio y fin mayores a la  $SDC_{\min}$  tiene algunos puntos menores a la  $SDC_{\min}$ . Esto contradice la definición de convexidad y por tanto, la función no es convexa en dicho intervalo. Por esta razón, no es indicado

resolver este tipo de problemas con algoritmos clásicos como el de Newton ya que éstos exigen convexidad. El problema de que la región de búsqueda no sea convexa, no es



**Figura 4.3** – La no convexidad de  $SDC_{\min}(\tilde{\tau})$  para  $\tilde{\tau} \in [-900, -300]$ .

impedimento para la aplicación del método de Newton, sin embargo, no se puede garantizar la convergencia del algoritmo si el espacio de búsqueda no es convexo. Resulta más indicado aplicar algoritmos que no supongan la convexidad del espacio de búsqueda. El algoritmo de la *sección dorada* no supone convexidad sino *unimodalidad*.

Se entiende por región unimodal un intervalo en el que hay un sólo valor  $x_{\min}$  que minimiza una función  $f$  que es monótonamente decreciente  $\forall x \leq x_{\min}$  y monótonamente creciente  $\forall x \geq x_{\min}$  dentro del intervalo.

La unimodalidad es una suposición más débil que la convexidad. Una función convexa es unimodal pero lo inverso no es necesariamente cierto. Por esta razón, en este trabajo suponemos que el valor óptimo se encuentra en una región unimodal y por lo tanto, utilizaremos el algoritmo de la sección dorada para encontrar el óptimo en dicho intervalo. El pseudocódigo de la búsqueda de la sección dorada (SEC\_DOR) se presenta en el siguiente listado.

**Algoritmo de la sección dorada:** SEC\_DOR( $f, a, b$ )

*Paso 1:* **igualar:**

$$\phi = (1 + \sqrt{5})/2, c = (b + a\phi)/(1 + \phi), d = (a + b\phi)/(1 + \phi);$$

**Paso 2: repetir:**

**si**  $SDC(c) < SDC(d)$ ,

**entonces**  $b = d, d = c, c = (b + a\phi)/(1 + \phi)$ ;

**si no**  $a = c, c = d, d = (a + b\phi)/(1 + \phi)$ ;

**hasta que**  $|a - b| < \epsilon$ ;

**Paso 3: devolver**  $(c + d)/(2)$ ;

El algoritmo SEC\_DOR toma una función  $f$  y un intervalo  $[a, b]$  en el que ésta es unimodal y regresa el valor en el intervalo que minimiza la función dada. Esto se logra al ir reduciendo el intervalo en porciones proporcionales a la *fracción áurea* hasta encerrar el mínimo en un intervalo tan pequeño como se desee. Comúnmente, se designa al valor a mitad del intervalo final como el óptimo.

El algoritmo de la sección dorada supone que tenemos una función y un intervalo en el que ésta es unimodal. Sin embargo, dicho intervalo también debe ser encontrado con un algoritmo de búsqueda. Nosotros proponemos una *búsqueda lineal* a lo largo de  $\tilde{\tau}$ .

A grandes rasgos, la búsqueda lineal propuesta consiste de dos fases. En la primera, el algoritmo buscará una región unimodal hacia la derecha. El paso de búsqueda aumentará en potencias de 2 y el ciclo terminará cuando el algoritmo detecte un cambio de pendiente en la función. El intervalo unimodal estará definido por los valores de la última y antepenúltima iteraciones del ciclo. Ahí se aplicará el algoritmo de la sección dorada y se obtendrá el valor mínimo en ese intervalo. En la segunda fase, se hará lo mismo pero hacia la derecha. Al final, se elegirá el valor de la primera y segunda iteración que minimicen la  $SDC_{\min}$ . El procedimiento resultante se muestra en el siguiente listado.

**Algoritmo de la búsqueda lineal con sección dorada:**  $\hat{\tau} = \text{BLSD}(y)$

**Paso 1: igualar**  $i = 1$ ;

**Paso 2: si**  $i \leq 2$ ,

**entonces**  $\tilde{\tau}_0 = (i - 1)(n + 1), k = 0$ ;

**si no ir a Paso 6;**

**Paso 3: repetir:**

$$\tilde{\tau}_{k+1} = \tilde{\tau}_k + (-1)^i 2^k, k = k + 1;$$

**hasta que**  $SDC_{\min}(\tilde{\tau}_{k+1}) \geq SDC_{\min}(\tilde{\tau}_k)$ ;

**Paso 4:**  $\hat{\tau}_i = \text{SEC\_DOR}(f = SDC_{\min}, a = \tilde{\tau}_{k+(-1)^{i-1}}, b = \tilde{\tau}_{k+(-1)^i})$ ;

**Paso 5: igualar**  $i = i + 1$ ; **ir a Paso 2**;

**Paso 6: si**  $SDC_{\min}(\hat{\tau}_1) \leq SDC_{\min}(\hat{\tau}_2)$  **entonces**  $\hat{\tau} = \hat{\tau}_1$  **si no**  $\hat{\tau} = \hat{\tau}_2$

**Paso 7: devolver**  $\hat{\tau}$

En términos más precisos, el algoritmo para minimizar la  $SDC_{\min}$  consiste en inicializar la variable  $\tilde{\tau}$  en 0 para el primer ciclo (búsqueda hacia la izquierda) y en  $n + 1$  para el segundo (búsqueda hacia la derecha). Estos puntos de arranque suponen que el punto de intersección de las rectas se encuentra fuera de los datos observados. Si existe una buena razón para suponer lo contrario, el valor del punto de arranque del segundo ciclo debe cambiarse a 0. En el primer ciclo, el valor de  $\tilde{\tau}$  se decrementa en potencias de 2 hasta que el nuevo valor de la  $SDC_{\min}$  sea peor o igual al actual ( $SDC_{\min}(\tilde{\tau}_{k+1}) \geq SDC_{\min}(\tilde{\tau}_k)$ ). Una vez que esto ha ocurrido, podemos deducir que la curvatura de la  $SDC_{\min}$  ha cambiado y que un valor mínimo se encuentra entre el nuevo valor  $\tilde{\tau}_{k+1}$  y el valor anterior al actual  $\tilde{\tau}_{k-1}$ . Debido a que la búsqueda es en este caso hacia la izquierda, sabemos que  $\tilde{\tau}_{k+1} < \tilde{\tau}_{k-1}$ . Por esta razón, le daremos al procedimiento SEC\_DOR los siguientes datos  $f=SDC_{\min}$ ,  $a = \tilde{\tau}_{k+1}$  y  $b = \tilde{\tau}_{k-1}$ . En el segundo ciclo, todo es similar al primero salvo que esta vez la búsqueda será hacia la derecha y al final del ciclo tendremos  $\tilde{\tau}_{k+1} > \tilde{\tau}_{k-1}$  y por esta razón debemos dar al procedimiento SEC\_DOR los datos de la siguiente manera  $f=SDC_{\min}$ ,  $a = \tilde{\tau}_{k-1}$  y  $b = \tilde{\tau}_{k+1}$ . Al final, el algoritmo compara el óptimo local encontrado hacia la izquierda con aquel encontrado hacia la derecha y devuelve el que minimiza la  $SDC_{\min}$ . El procedimiento resultante es una búsqueda lineal con sección dorada y haremos referencia a él por sus siglas (BLSA).

### 4.2.3. Experimentación.

En este capítulo, se han propuesto dos maneras de estimar los parámetros del MNO<sub>LIN</sub>. Sin embargo, no hay una manera directa de saber cual de las técnicas de estimación propuestas es la mejor. Por ejemplo, a pesar de que los algoritmos de LEV<sub>MARQ</sub> y el de la BLS<sub>D</sub> tienen como objetivo minimizar la SDC, en la práctica, ambos métodos no llegan a la misma solución.

Con el objetivo de saber cual de las técnicas es más eficaz al minimizar la SDC, hemos decidido comparar el algoritmo de LEV<sub>MARQ</sub> contra el de BLS<sub>D</sub>. En esta sección, veremos los lineamientos seguidos y los resultados de dicho experimento.

#### Lineamientos observados en el experimento

Para saber cual de estos algoritmos es mejor basta saber cual de ellos minimiza el valor de SDC ya que éste es el objetivo de ambos. Para obtener datos que puedan clarificar esta interrogante, decidimos diseñar series con las siguientes características:

- El punto de origen se simulo con  $\tau \sim U(-500, 0)$  y  $\beta_0 \sim U(0, 1000)$ , donde  $U(a, b)$  es una distribución uniforme en el intervalo  $[a, b]$ .
- Simulamos un cono dentro del cual generaríamos las rectas correspondientes a cada estación. El ángulo inferior ( $A_{\text{inf}}$ ) del cono fue simulado así,  $A_{\text{inf}} \sim U(0, \frac{\pi}{2})$  mientras que el ángulo superior ( $A_{\text{sup}}$ ) se simuló a partir del ángulo inferior de la manera  $A_{\text{sup}} \sim U(A_{\text{inf}}, \frac{\pi}{2})$ . El resto de las pendientes de cada línea se simuló así  $\beta_s \sim U(A_{\text{inf}}, A_{\text{sup}}) * U(0, 1000)$   $s = 1, \dots, 12$ .
- De la serie generada tomamos la media  $\bar{y}$  y con ella generamos un ruido autorregresivo  $\mu_t$  con  $E(\mu) = 0$  y desviación constante de la forma  $\sqrt{\text{Var}(\mu_t)} \sim U(0,01, 0,15) * \bar{y}$ .
- Para generar el ruido autorregresivo de orden  $p = 18$  (AR(18)) primero había que saber si sus parámetros tomarían valores diferentes de cero. Una vez hecho esto, se

le daba al parámetro un valor en el intervalo  $[-0,5, 0,5]$ . Para los primeros cinco parámetros hicimos que la probabilidad de ser diferente de cero fuera  $P(\phi_i \neq 0) = 0,5$   $i = 1, \dots, 5$ , mientras que para el resto la probabilidad fue  $P(\phi_i \neq 0) = 0,1$   $i = 6, \dots, 18$ . Una vez generados los parámetros, revisamos que cumplieran las condiciones de invertibilidad y estabilidad necesarias para los modelos autorregresivos. Si no lo hacían, simplemente generábamos otro conjunto de parámetros. Para que la serie tuviera media cero hicimos  $\mu_t = \mu_t - \bar{\mu}$  y para que su desviación fuese la requerida hicimos  $\mu_t = \mu_t * (\text{desviación de } \mu_t) / (\text{desviación deseada})$ .

- Las series fueron generadas para 175 casos con observaciones desde 25 hasta 200. Para cada caso, generamos 10 series para hacer un total de 1750 series.
- Una vez que se generaron las 1750 series, decidimos aplicar el método de LEV\_MARQ y el de la BLSD a cada caso.

## Resultados

Debido a que el objetivo que se quería minimizar era la SDC, obtuvimos mucha variabilidad como para poder graficar uno a uno los resultados de cada algoritmo. En vez de esto, graficamos los coeficientes resultantes de dividir la SDC obtenida con el algoritmo de la BLSD ( $SDC_{BLSD}$ ) entre la SDC obtenida con el algoritmo de LEV\_MARQ ( $SDC_{LEV\_MARQ}$ ) para cada serie. El resultado se muestra en la figura 4.4. En la figura 4.5 se muestra el histograma de frecuencias asociado a la razón  $SDC_{BLSD}/SDC_{LEV\_MARQ}$ .

En el histograma de frecuencias de la figura 4.5, el valor «1<sup>+</sup>» representa todos los valores mayores a 1. De las gráficas 4.4 y 4.5 se observa que el algoritmo propuesto mejoró al estándar en 1738 de 1750 pruebas. El algoritmo propuesto fue más de diez veces mejor en 1311 ocasiones. La mejora del algoritmo de la sección dorada con búsqueda lineal se debe al hecho de que restringe la búsqueda a una sola dimensión a diferencia del algoritmo de LEV\_MARQ que realiza una búsqueda en 14 variables.

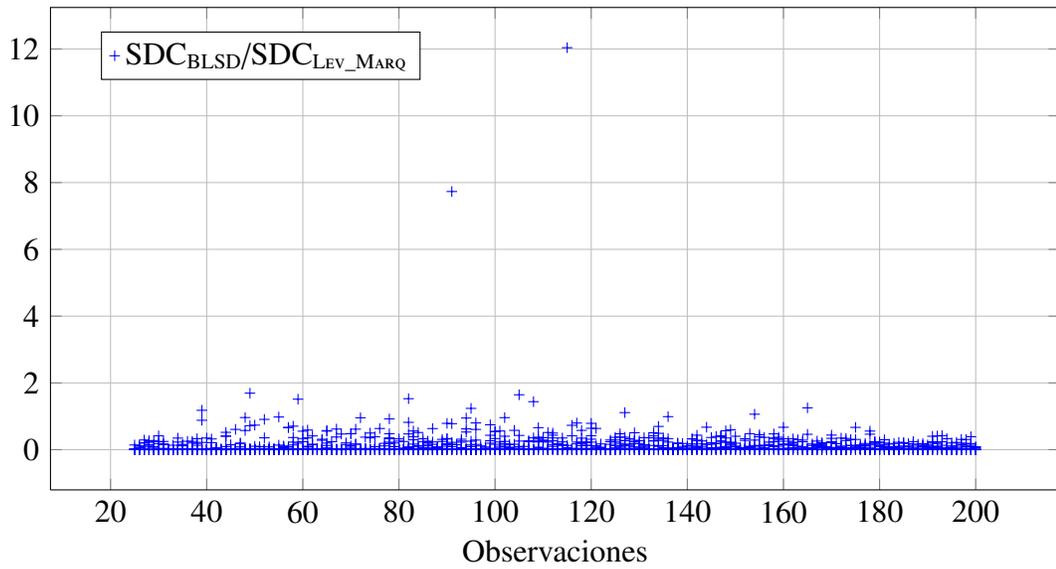


Figura 4.4 – Coeficiente de la SDC con cada algoritmo.

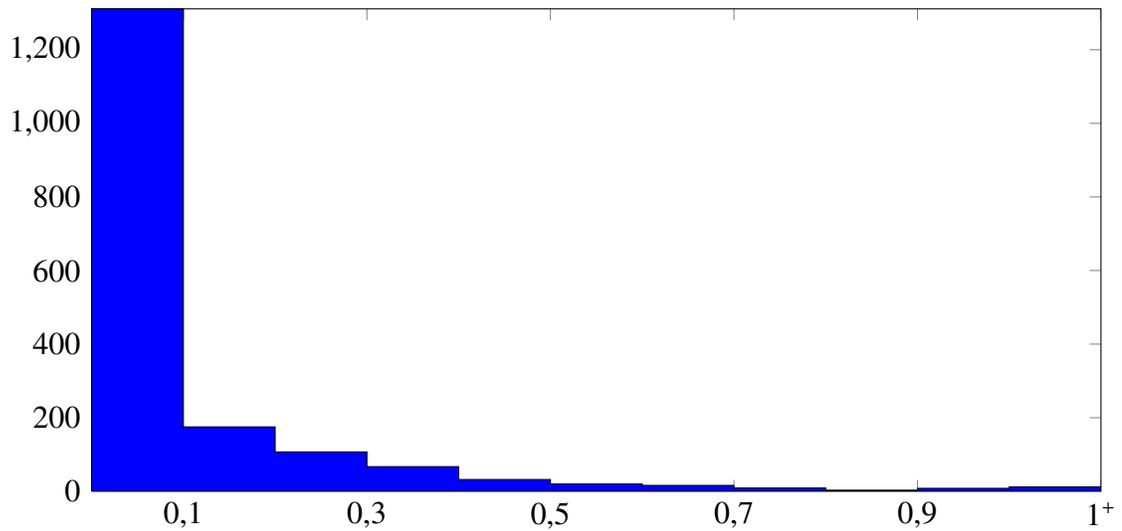


Figura 4.5 – Histograma de frecuencias de la razón  $SDC_{BLS}/SDC_{LEV\_MARQ}$ .

### 4.3. Conclusiones

En este capítulo, hemos encontrado estimadores que minimizan la SDC tanto del  $MLIN$  como del  $MNoLIN$ .

Para el  $MLIN$ , propusimos los estimadores MCO. Dichos estimadores se pudieron obtener de forma analítica gracias a que el  $MLIN$  es un modelo lineal en sus parámetros.

Para el  $MNoLIN$ , propusimos los estimadores MCNL. Para encontrar dichos estimado-

res, es necesario utilizar un algoritmo de optimización adecuado y eficiente. Un algoritmo que toma ventaja de la estructura del MNoLIN fue propuesto y se comparó contra la técnica estándar; el algoritmo de LEV\_MARQ. Los resultados fueron rotundamente favorables para el algoritmo que propusimos y por esta razón recomendamos utilizar el algoritmo de búsqueda lineal con sección dorada para estimar los parámetros del MNoLIN.

En el siguiente capítulo veremos la justificación de utilizar estimadores tales que minimicen la SDC en cada modelo además de su caracterización estadística. También veremos los pros y contras de una técnica alternativa para la estimación de parámetros.

---

---

## CAPÍTULO 5

---

### Caracterización estadística y estimaciones alternativas

**S**e estudiarán las propiedades estadísticas de los estimadores que minimizan la SDC (MCO para el MLIN y MCNL para el MNoLIN) bajo dos suposiciones sobre la fuente de variación, a saber: los elementos de  $\mu$  siguen un proceso NID( $0, \sigma^2$ ) o un proceso de media cero,  $E(\mu) = \mathbf{0}$ . A pesar de que ninguna de las suposiciones anteriores afectan el proceso de encontrar los estimadores que minimizan la SDC, si cambian tanto las propiedades estadísticas de los estimadores como los intervalos predictivos.

Para el caso del MNoLIN, se ofrece una estimación de parámetros alternativa, más simple que el algoritmo de la sección dorada con búsqueda lineal y con propiedades estadísticas más deseables. Al final, realizaremos un experimento para conocer que tan buenos son estos estimadores comparados con los de MCNL.

Antes, daremos un repaso a dos importantes teoremas de regresión lineal. Estas pruebas serán útiles para demostrar las propiedades de los parámetros bajo cualquiera de los dos supuestos sobre su fuente de variación.

## 5.1. Dos teoremas importantes sobre regresión lineal

El análisis de regresión lineal es un tema extenso y estudiado desde la época de Johann Carl Friedrich Gauss. Este gran matemático propuso un teorema que sería complementado por otro gran matemático, Andrei Andreyevich Markov. A la postre, a éste se le conocería como el teorema de Gauss-Markov. Primero veremos las suposiciones de las que depende la demostración y luego veremos el teorema en sí.

### 5.1.1. Suposiciones de Gauss-Markov

Tan famosas como el teorema son las suposiciones necesarias para su demostración y a las cuales se les denomina «las suposiciones de Gauss-Markov». Sea  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$  un modelo de regresión lineal. Entonces, las suposiciones de Gauss-Markov son:

1. El modelo subyacente a los datos es lineal en los parámetros.
2. Multicolinealidad imperfecta, las columnas de  $\mathbf{X}$  son independientes.
3. Media condicional igual a cero,  $E(\boldsymbol{\mu} | \mathbf{X}) = \mathbf{0}$ .
4. Varianza finita y constante,  $\text{Var}(\boldsymbol{\mu} | \mathbf{X}) = \sigma^2 < \infty$ .
5. Covarianza igual a cero,  $\text{Cov}(\boldsymbol{\mu}(t), \boldsymbol{\mu}(t') | \mathbf{X}) = 0, \forall t \neq t'$ .

Si las suposiciones anteriores fueran verdaderas entonces, es posible demostrar que los estimadores MCO de  $\boldsymbol{\beta}$  ( $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ ) son MELI<sup>1</sup> (Mejor Estimador Lineal Insesgado). Lo que esto significa es que no hay un estimador lineal e insesgado de  $\boldsymbol{\beta}$  cuyos elementos posean una varianza menor a la varianza de los elementos de  $\hat{\boldsymbol{\beta}}$ . A continuación se presenta formalmente este teorema.

**Teorema 1.** *Si se cumplen las suposiciones de Gauss-Markov, los estimadores MCO son los mejores estimadores lineales insesgados de los parámetros del modelo lineal.*

<sup>1</sup>Equivalente en español de BLUE (*Best Linear Unbiased Estimator*)

*Demostración.* El estimador de MCO es un vector y en consecuencia su varianza es una matriz. Por ello, lo que se demostrará será que  $\hat{\beta}$  minimiza la varianza para cualquier combinación lineal de los coeficientes estimados,  $l' \hat{\beta}$ . Se observa que

$$\begin{aligned}\text{Var}(l' \hat{\beta}) &= l' \text{Var}(\hat{\beta}) l \\ &= l' [\sigma^2 (\mathbf{X}' \mathbf{X})^{-1}] l \\ &= \sigma^2 l' (\mathbf{X}' \mathbf{X})^{-1} l\end{aligned}$$

Nótese que  $\text{Var}(l' \hat{\beta})$  es un escalar. Sea  $\tilde{\beta}$  otro estimador lineal insesgado de  $\beta$  que sea una combinación lineal de los datos. De lo que se trata es de demostrar que  $\text{Var}(l' \tilde{\beta}) \geq \sigma^2 l' (\mathbf{X}' \mathbf{X})^{-1} l$ .

Primero, se observa que se puede escribir cualquier otro estimador de  $\beta$  que sea una combinación lineal de los datos en la forma

$$\tilde{\beta} = [(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' + \mathbf{B}] \mathbf{y} + \mathbf{b}_0,$$

siendo  $\mathbf{B}$  una matriz de  $p \times n$ , y  $\mathbf{b}_0$  es un vector de  $p \times 1$  constantes, que ajusta en forma adecuada al estimador MCO para formar el estimador alternativo. A continuación se observa que si el modelo es correcto entonces

$$\begin{aligned}\text{E}(\tilde{\beta}) &= \text{E}([(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' + \mathbf{B}] \mathbf{y} + \mathbf{b}_0) \\ &= [(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' + \mathbf{B}] \text{E}(\mathbf{y}) + \mathbf{b}_0 \\ &= [(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' + \mathbf{B}] \mathbf{X} \beta + \mathbf{b}_0 \\ &= (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X} \beta + \mathbf{B} \mathbf{X} \beta + \mathbf{b}_0 \\ &= \beta + \mathbf{B} \mathbf{X} \beta + \mathbf{b}_0.\end{aligned}$$

Por consiguiente,  $\tilde{\beta}$  es insesgado si, y sólo si, tanto  $\mathbf{b}_0 = \mathbf{0}$  como  $\mathbf{B} \mathbf{X} = \mathbf{0}$ . La varianza del

estimador  $\tilde{\beta}$  es

$$\begin{aligned}
 \text{Var}(\tilde{\beta}) &= \text{Var}\left(\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right]\mathbf{y}\right), \\
 &= \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right] \text{Var}(\mathbf{y}) \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right]', \\
 &= \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right] \sigma^2 \mathbf{I} \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right]', \\
 &= \sigma^2 \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right] \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{B}\right]', \\
 &= \sigma^2 \left[(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{B}\mathbf{B}'\right],
 \end{aligned}$$

porque  $\mathbf{B}\mathbf{X} = \mathbf{0}$ , lo cual a su vez implica que  $(\mathbf{B}\mathbf{X})' = \mathbf{X}'\mathbf{B}' = \mathbf{0}$ . Como resultado,

$$\begin{aligned}
 \text{Var}(l' \tilde{\beta}) &= l' \text{Var}(\tilde{\beta}) l, \\
 &= l' \left(\sigma^2 \left[(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{B}\mathbf{B}'\right]\right) l, \\
 &= \sigma^2 l' (\mathbf{X}'\mathbf{X})^{-1} l + \sigma^2 l' \mathbf{B}\mathbf{B}' l, \\
 &= \text{Var}(l' \hat{\beta}) + \sigma^2 l' \mathbf{B}\mathbf{B}' l.
 \end{aligned}$$

Primero se observa que el resultado del producto  $\mathbf{B}\mathbf{B}'$  es una matriz cuadrada con todos sus elementos mayores o iguales a cero. Por consiguiente  $\sigma^2 l' \mathbf{B}\mathbf{B}' l \geq \mathbf{0}$ . A continuación, se ve que se puede definir a  $l^* = \mathbf{B}' l$  y por tanto,

$$l' \mathbf{B}\mathbf{B}' l = l^* l^* = \sum_{i=1}^p l_i^{*2},$$

que debe ser estrictamente mayor que 0 para cierta  $l \neq \mathbf{0}$ , a menos que  $\mathbf{B} = \mathbf{0}$ . Por lo anterior, el estimador MCO de  $\beta$  es el mejor estimador lineal insesgado. Esto completa la demostración. □

### 5.1.2. Suposiciones débiles de Gauss-Markov

En la práctica es raro que las suposiciones de Gauss-Markov se cumplan. Es común, por ejemplo, que la fuente de variación no sea una sucesión de variables independientes (incumplimiento del quinto supuesto) o que la varianza no sea constante (incumplimiento del cuarto supuesto). En este contexto, las suposiciones más realistas serían las siguientes

1. El modelo subyacente a los datos es lineal en los parámetros.
2. Multicolinealidad imperfecta, las columnas de  $\mathbf{X}$  son independientes.
3. Media condicional igual a cero,  $E(\boldsymbol{\mu} | \mathbf{X}) = \mathbf{0}$ .

A los anteriores nos referiremos como los supuestos *débiles* de Gauss-Markov. El calificativo «débiles» evoca que los nuevos supuestos son mejores por ser menos en cantidad. Sin embargo, recordemos que «entre menos suponemos, menos podemos concluir». Por ejemplo, bajo las suposiciones débiles, los estimadores MCO ya no son los mejores estimadores lineales insesgados. Pero no todo está perdido; las propiedades de los estimadores MCO al aceptar los supuestos débiles de Gauss-Markov se enuncian en el siguiente teorema

**Teorema 2.** *Si las suposiciones débiles de Gauss-Markov se cumplen entonces, los estimadores MCO son insesgados y consistentes.*

Aquí nos limitaremos a demostrar que los estimadores MCO son insesgados. Para la demostración de su consistencia, el lector puede consultar (Wooldridge (2001), pág. 353).

*Demostración.* Los estimadores MCO bajo las suposiciones débiles de Gauss-Markov son insesgados ya que la fuente de variación es un proceso de media cero y por tanto

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\right), \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{y}), \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}), \end{aligned}$$

$$\begin{aligned}
&= (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' [\mathbf{X} \boldsymbol{\beta} + \mathbf{E}(\boldsymbol{\mu})], \\
&= (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{X} \boldsymbol{\beta}, \\
&= \boldsymbol{\beta}.
\end{aligned}$$

El supuesto de ruido con media cero se expresa matemáticamente como  $\mathbf{E}(\boldsymbol{\mu}) = \mathbf{0}$  y por esta razón, los estimadores MCO son insesgados. Esto completa la demostración.  $\square$

## 5.2. Caracterización estadística del MLIN

Como ya ha sido mencionado, las propiedades estadísticas de los estimadores MCO pueden variar dependiendo de las suposiciones acerca de la distribución del proceso  $\boldsymbol{\mu}$ . Primero veremos lo que le ocurre tanto a los parámetros como a los intervalos predictivos del MLIN cuando se supone que los elementos de  $\boldsymbol{\mu}$  siguen un proceso NID( $0, \sigma^2$ ) y después, analizaremos el caso en el que esos elementos siguen un proceso ARMA( $p, q$ ).

### 5.2.1. Las desviaciones siguen un proceso NID( $0, \sigma^2$ )

#### Propiedades estadísticas de $\hat{\boldsymbol{\beta}}$

Si los elementos de  $\boldsymbol{\mu}$  siguen un proceso NID( $0, \sigma^2$ ) entonces, los estimadores MCO serán los mejores estimadores lineales insesgados como se establece en la siguiente proposición.

**Proposición 1.** *Si las desviaciones de una serie, cuyo proceso subyacente es el MLIN, siguen un proceso NID( $0, \sigma^2$ ) entonces, la serie cumplirá con los cinco supuestos de Gauss-Markov y el estimador  $\hat{\boldsymbol{\beta}}$  será el mejor estimador lineal insesgado de  $\boldsymbol{\beta}$ .*

*Demostración.* A continuación se muestra la validez de cada una de las suposiciones:

1. El modelo subyacente a la serie es lineal. Esto resulta evidente ya que ninguno de los parámetros se multiplican entre sí y este hecho no cambia cualquiera que sea la distribución de la fuente de variación ya que esta se suma al modelo y no tiene ninguna otra interacción con los parámetros. Se podría pensar que los parámetros no son lineales ya que son variables binarias que se «prenden» y «apagan» dependiendo de la estación. Sin embargo, esto no conduce a la no linealidad de los parámetros pues quien se «prende» y «apaga» dependiendo de la estación son las variables binarias  $D_{s,t}$ . Las variables binarias están incluidas en la matriz  $\mathbf{X}$  y forman parte de las variables independientes y dichas variables pueden tomar cualquier valor sin afectar la linealidad del modelo ya que esta linealidad compete únicamente a los parámetros. Los estimadores MCO mantienen su linealidad pues los parámetros son lineales (como lo muestra la ecuación (4.9)) además de ser insesgados como se muestra a continuación

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{y}), \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}, \\ &= \boldsymbol{\beta}. \end{aligned}$$

2. Multicolinealidad imperfecta, las columnas de  $\mathbf{X}$  son linealmente independientes. Para demostrar este hecho, debemos recordar primero que las columnas de una matriz son linealmente *dependientes* si y sólo si existen constantes  $a_1, a_2, \dots, a_{m+1}$  no todas cero tales que

$$a_1 \mathbf{X}^1 + a_2 \mathbf{X}^2 + \dots + a_{m+1} \mathbf{X}^{m+1} = \mathbf{0}. \quad (5.1)$$

siendo  $\mathbf{X}^j$  la  $j$ -ésima columna de  $\mathbf{X}$ . Si la única solución de (5.1) es aquella en la que todas las constantes son cero entonces, se concluye que las columnas de  $\mathbf{X}$  son independientes y el supuesto de multicolineridad imperfecta mantiene su validez.

Para demostrar que la matriz  $\mathbf{X}$  del MLIN cumple con este supuesto, demostraremos que la única solución factible de (5.1) es la trivial. Recordemos que la matriz  $\mathbf{X}$  esta dada por  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{\lfloor t/m \rfloor}, \mathbf{X}_{\lceil t/m \rceil}]'$  donde cada submatriz  $\mathbf{X}_i$  tiene la forma

$$\mathbf{X}_i = \begin{bmatrix} 1 & (i-1)m+1 & 0 & \dots & 0 \\ 1 & 0 & (i-1)m+2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & im \end{bmatrix}.$$

siendo lo anterior válido para  $i = 1, \dots, \lfloor t/m \rfloor$  mientras que  $\mathbf{X}_{\lceil t/m \rceil}$  se define como

$$\mathbf{X}_{\lceil t/m \rceil} = \begin{bmatrix} 1 & (\lceil t/m \rceil - 1)m + 1 & 0 & \dots & 0 & \dots & 0 \\ 1 & 0 & (\lceil t/m \rceil - 1)m + 2 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & n & \dots & 0 \end{bmatrix}.$$

Según la definición de independencia lineal, para que las columnas de  $\mathbf{X}$  sean linealmente independientes, es necesario y suficiente que la única solución del sistema  $\mathbf{X}\mathbf{a} = \mathbf{0}$  (donde  $\mathbf{a} = [a_1, a_2, \dots, a_{m+1}]'$ ) sea la trivial,  $\mathbf{a} = \mathbf{0}$ . Empezaremos notando que si tenemos una serie con las primeras  $m$  observaciones entonces, tendremos que resolver el siguiente sistema de ecuaciones

$$\begin{aligned} a_1 + a_2 + 0 + \dots + 0 &= 0 \\ a_1 + 0 + 2a_3 + \dots + 0 &= 0 \\ \vdots & \quad \quad \quad \vdots \quad \ddots \quad \quad \quad \vdots = 0 \\ a_1 + 0 + 0 + \dots + ma_{m+1} &= 0. \end{aligned} \tag{5.2}$$

La solución del sistema (5.2) es  $[a_2 = a_1, a_3 = \frac{1}{2}a_1, \dots, \frac{1}{m}a_{m+1} = a_1]$ . Para conocer el valor de las constantes  $a_2, \dots, a_{m+1}$  es necesario conocer el valor de  $a_1$ . Claramente,  $a_1$  puede tomar cualquier valor diferente de cero, lo que conduciría a una solución no

trivial. En este caso, las columnas de  $\mathbf{X}$  no serían independientes. Sin embargo, basta añadir una observación más al sistema para tener  $m + 1$  ecuaciones de la forma

$$\begin{aligned}
 a_1 + a_2 + 0 + \dots + 0 &= 0 \\
 a_1 + 0 + 2a_3 + \dots + 0 &= 0 \\
 \vdots & \\
 a_1 + 0 + 0 + \dots + ma_{m+1} &= 0 \\
 a_1 + (m+1)a_2 + 0 + \dots + 0 &= 0.
 \end{aligned} \tag{5.3}$$

La solución del sistema (5.3) para  $a_3, \dots, a_{m+1}$  es la misma que la del sistema (5.2). La diferencia es que en el sistema (5.3)  $a_1$  no puede tomar cualquier valor, sino que esta definida de la siguiente manera

$$\begin{aligned}
 a_1 + a_2 &= 0 \\
 a_1 + (m+1)a_2 &= 0.
 \end{aligned} \tag{5.4}$$

El sistema (5.4) tiene solución única. Esto se debe a que su determinante es diferente de cero como se muestra a continuación

$$m + 1 - 1 = m \neq 0. \tag{5.5}$$

La solución única del sistema (5.4) es la trivial ( $a_1 = 0$  y  $a_2 = 0$ ). Lo anterior ocurre con cualquier observación que se añada a las  $m$  observaciones del sistema (5.2). El resto de las  $a$ 's tomaran valores iguales a cero debido a que  $a_1 = 0$ . Esto demuestra que la solución única del sistema  $\mathbf{Xa} = \mathbf{0}$  cuando hay más de  $m$  observaciones es la trivial y por tanto, las columnas de  $\mathbf{X}$  son independientes y el supuesto de multicolinealidad imperfecta mantiene su validez.

3. Media condicional igual a cero,  $E(\mu | \mathbf{X}) = 0$ . Este supuesto se cumple debido a

que los elementos de  $\boldsymbol{\mu}$  siguen un proceso NID( $0, \sigma^2$ ) el cual tiene media cero por definición y por tanto, el vector  $\boldsymbol{\mu}$  tiene media cero.

4. Varianza finita y constante,  $\text{Var}(\boldsymbol{\mu}(t)|\mathbf{X}) = \sigma^2 < \infty, \forall t$ . Esto se cumple debido a que la varianza de cada elemento de  $\boldsymbol{\mu}$  es constante y finita debido a que siguen una distribución NID( $0, \sigma^2$ ).
5. Covarianza igual a cero,  $\text{Cov}(\boldsymbol{\mu}(t), \boldsymbol{\mu}(t')|\mathbf{X}) = 0, \forall t \neq t'$ . Según la definición de NID( $0, \sigma^2$ ), las variables aleatoria que siguen este proceso son independientemente distribuidas y por lo tanto, su covarianza es cero.

Lo anterior demuestra que si las desviaciones del MLIN siguen una distribución normal entonces, este modelo cumplirá con las suposiciones de Gauss-Markov y por lo tanto, los estimadores MCO del MLIN serán los mejores estimadores lineales insesgados de  $\boldsymbol{\beta}$ . Esto completa la demostración.  $\square$

### Intervalos predictivos

Para encontrar intervalos predictivos de la forma  $y_{n+h} \pm t_{\alpha/2, v} \sqrt{\text{Var}(\hat{\mu}_{n+h}|\mathbf{X})}$  donde  $t_{\alpha/2, v}$  es un valor de la distribución  $t$  de Student con  $v$  grados de libertad y  $\hat{\mu}_{n+h} = y_{n+h} - \hat{y}_{n+h}$ , debemos conocer  $\text{Var}(\hat{\mu}_{n+h}|\mathbf{X})$  que representa la varianza del error de pronóstico en el que se incurre al pronosticar  $h$  pasos hacia adelante en la etapa  $n$ . Lo primero por hacer es encontrar una expresión para el error de pronóstico como se muestra a continuación

$$\begin{aligned} \hat{\mu}_{n+h} = y_{n+h} - \hat{y}_{n+h} &= \beta_0 + \left( \sum_{s=1}^m D_{s, n+h} \beta_s \right) x_{n+h} + \mu_{n+h} - \left( \hat{\beta}_0 + \left( \sum_{s=1}^m D_{s, n+h} \hat{\beta}_s \right) x_{n+h} \right) \\ &= (\beta_0 - \hat{\beta}_0) + \left( \sum_{s=1}^m D_{s, n+h} (\beta_s - \hat{\beta}_s) \right) x_{n+h} + \mu_{n+h}. \end{aligned}$$

La expresión anterior representa el error de pronóstico en el que se incurre al pronosticar  $h$  pasos hacia adelante. Para obtener la varianza del error de pronóstico debemos obtener el

segundo momento del error de pronóstico. Esto lo haremos elevando al cuadrado el error y luego obteniendo su valor esperado. El cuadrado del error de pronóstico es

$$\begin{aligned}
\hat{\mu}_{n+h}^2 &= \left( (\beta_0 - \hat{\beta}_0) + \left( \sum_{s=1}^m D_{s,n+h}(\beta_s - \hat{\beta}_s) \right) x_{n+h} \right)^2 + \\
&\quad \left( (\beta_0 - \hat{\beta}_0) + \left( \sum_{s=1}^m D_{s,n+h}(\beta_s - \hat{\beta}_s) \right) x_{n+h} \right) \mu_{n+h} + \mu_{n+h}^2, \\
&= (\beta_0 - \hat{\beta}_0)^2 + 2(\beta_0 - \hat{\beta}_0) \left( \sum_{s=1}^m D_{s,n+h}(\beta_s - \hat{\beta}_s) \right) x_{n+h} + \\
&\quad \left( \sum_{s=1}^m D_{s,n+h}(\beta_s - \hat{\beta}_s) \right)^2 x_{n+h}^2 + \\
&\quad \left( (\beta_0 - \hat{\beta}_0) + \left( \sum_{s=1}^m D_{s,n+h}(\beta_s - \hat{\beta}_s) \right) x_{n+h} \right) \mu_{n+h} + \mu_{n+h}^2,
\end{aligned}$$

y el segundo momento del error de pronóstico (que es la varianza) esta dado por

$$\begin{aligned}
\text{Var}(\hat{\mu}_{n+h} | \mathbf{X}) &= \text{E}(\hat{\mu}_{n+h}^2 | \mathbf{X}) - \left( \text{E}(\hat{\mu}_{n+h} | \mathbf{X}) \right)^2 \\
&= \text{E}((\beta_0 - \hat{\beta}_0)^2 | \mathbf{X}) + \\
&\quad 2 \text{E} \left( (\beta_0 - \hat{\beta}_0) \left( \sum_{s=1}^m D_{s,n+h}(\beta_s - \hat{\beta}_s) \right) x_{n+h} \middle| \mathbf{X} \right) + \\
&\quad \text{E} \left( \left( \sum_{s=1}^m D_{s,n+h}(\beta_s - \hat{\beta}_s) \right)^2 x_{n+h}^2 \middle| \mathbf{X} \right) + \\
&\quad \text{E} \left( (\beta_0 - \hat{\beta}_0) + \left( \sum_{s=1}^m D_{s,n+h}(\beta_s - \hat{\beta}_s) \right) x_{n+h} \middle| \mathbf{X} \right) \text{E}(\mu_{n+h} | \mathbf{X}) + \\
&\quad \text{E}(\mu_{n+h}^2 | \mathbf{X}).
\end{aligned}$$

Debido a que  $\mu_{n+h}$  es un proceso NID(0,  $\sigma^2$ ) independiente de  $\mathbf{X}$ , tenemos que  $\text{E}(\mu_{n+h} | \mathbf{X}) = 0$  y  $\text{E}(\mu_{n+h}^2 | \mathbf{X}) = \sigma^2$ . Con esta información, podemos simplificar la varianza de  $\hat{\mu}_{n+h}$  de la siguiente manera

$$\text{Var}(\hat{\mu}_{n+h} | \mathbf{X}) = \text{Var}(\hat{\beta}_0 | \mathbf{X}) + 2 \left( \sum_{s=1}^m \text{Cov}(\hat{\beta}_0, \hat{\beta}_s | \mathbf{X}) D_{s,n+h} \right) x_{n+h} +$$

$$\left( \sum_{s=1}^m \text{Var}(\hat{\beta}_s | \mathbf{X}) D_{s,n+h} \right) x_{n+h}^2 + \sigma^2.$$

Los elementos  $\text{Var}(\hat{\beta}_0 | \mathbf{X})$ ,  $\text{Var}(\hat{\beta}_s | \mathbf{X})$  y  $\text{Cov}(\hat{\beta}_0, \hat{\beta}_s | \mathbf{X}) \forall s$  corresponden respectivamente a los elementos  $\mathbf{VC}(1, 1)$ ,  $\mathbf{VC}(1 + s, 1 + s)$  y  $\mathbf{VC}(1, 1 + s) \forall s$  de  $\mathbf{VC}$  que es la matriz de varianzas y covarianzas de  $\hat{\beta}$  y que esta dada por

$$\mathbf{VC}(\hat{\beta} | \mathbf{X}) = \sigma^2 [\mathbf{X}' \mathbf{X}]^{-1}. \quad (5.6)$$

## 5.2.2. Las desviaciones siguen un proceso ARMA( $p, q$ )

### Propiedades estadísticas de $\hat{\beta}$

Si la fuente de variación exhibe correlaciones entonces, los estimadores MCO dejarán de cumplir con la quinta de las suposiciones de Gauss-Markov y perderán su condición de mejores estimadores lineales insesgados. Sin embargo, los supuestos débiles de Gauss-Markov se siguen cumpliendo como se demuestra en la siguiente proposición

**Proposición 2.** *Si la fuente de variación de una serie que tiene como modelo subyacente al MLIN sigue una distribución ARMA( $p, q$ ) entonces, se cumplirán las suposiciones débiles de Gauss-Markov y los estimadores MCO serán insesgados y consistentes.*

*Demostración.* A continuación se verá la demostración para cada uno de los supuestos:

1. El modelo subyacente a la serie es lineal. Como en la proposición anterior, la distribución de las desviaciones no afecta la linealidad del modelo debido a que ésta se suma a la serie y no tiene ninguna otra interacción con los parámetros.
2. Multicolinealidad imperfecta, las columnas de  $\mathbf{X}$  son independientes. La demostración de este punto es idéntica que aquella hecha en el caso de ruido NID( $0, \sigma^2$ ) pues la distribución de la fuente de variación no interviene en este punto.

3. Media condicional igual a cero,  $E(\boldsymbol{\mu} | \mathbf{X}) = \mathbf{0}$ . Esta suposición se cumple debido a que por definición, las series que siguen un proceso ARMA( $p, q$ ) tienen valor esperado igual a cero.

Lo anterior demuestra que si las desviaciones del MLIN siguen un proceso ARMA( $p, q$ ) entonces, se cumplirán los supuestos débiles de Gauss-Markov y por lo tanto, los estimadores MCO serán insesgados y consistentes. Esto completa la demostración.  $\square$

### Intervalos predictivos

Según [Box et al. \(2008, pág. 402\)](#), para obtener los intervalos predictivos de la forma  $y_{n+h} \pm t_{\alpha/2, v} \sqrt{\text{Var}(\hat{\mu}_{n+h} | \mathbf{X})}$  bajo el supuesto de ruido ARMA( $p, q$ ), es necesario suponer *información perfecta*. Esto implica suponer que tanto los parámetros del modelo como los del proceso ARMA( $p, q$ ) son conocidos. De esta manera el error de pronóstico estará dado por

$$\begin{aligned} \hat{\mu}_{n+h} &= y_{n+h} - \hat{y}_{n+h} = \beta_0 + \left( \sum_{s=1}^m D_{s, n+h} \beta_s \right) x_{n+h} + \mu_{n+h} - \left( \hat{\beta}_0 + \left( \sum_{s=1}^m D_{s, n+h} \hat{\beta}_s \right) x_{n+h} \right), \\ &= \beta_0 + \left( \sum_{s=1}^m D_{s, n+h} \beta_s \right) x_{n+h} + \mu_{n+h} - \left( \beta_0 + \left( \sum_{s=1}^m D_{s, n+h} \beta_s \right) x_{n+h} \right), \\ &= \mu_{n+h}, \end{aligned}$$

y la varianza del error de pronóstico será

$$\text{Var}(\hat{\mu}_{n+h} | \mathbf{X}, \tau, \boldsymbol{\beta}, \phi(B), \theta(B)) = \sigma_\varepsilon^2 \sum_{j=0}^{h-1} \psi_j^2, \quad (5.7)$$

donde  $\sigma_\varepsilon^2$  es la varianza de la fuente de variación del proceso ARMA( $p, q$ ) y  $\psi(B) = \phi(B)^{-1}\theta(B)$ , siendo  $\phi(B)$  y  $\theta(B)$  polinomios de operadores regresivos<sup>2</sup> tales que

$$\phi(B)y_t = \theta(B)\varepsilon_t,$$

$$y_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}.$$

### 5.3. Caracterización estadística del MNoLIN

Debido a su naturaleza no lineal, no hay una manera analítica de encontrar los intervalos predictivos del MNoLIN ni tampoco hay forma de deducir las propiedades estadísticas de los estimadores MCNL. Cuando la naturaleza no lineal de un modelo impide deducir sus propiedades estadísticas, lo que se hace en la literatura es suponer que algunos parámetros son conocidos. Este es el caso de la metodología ARIMA( $p, d, q$ ) y de los métodos de suavizado exponencial. Por esta razón, supondremos que algunos de los parámetros del MNoLIN son conocidos. En este caso, lo mejor es suponer que  $\tau$  es conocido ya que esto simplifica enormemente la deducción de las propiedades estadísticas. Si suponemos que  $\tau$  es conocido entonces, el MNoLIN se convierte en un modelo lineal del que podremos deducir tanto propiedades estadísticas como intervalos predictivos con facilidad. En la práctica, habrá que reemplazar  $\tau$  por  $\hat{\tau}$ . Lo anterior aplica a los dos posibles supuestos sobre la fuente de variación como veremos a continuación.

#### 5.3.1. Las desviaciones siguen un proceso NID( $0, \sigma^2$ ) y $\tau$ es conocido

##### Propiedades estadísticas de $\hat{\beta}$

En la siguiente proposición se establecen las propiedades estadísticas de los estimadores  $\hat{\beta}$ .

<sup>2</sup>Un operador regresivo  $B$  tiene la propiedad de «atrasar» o «rezagar» una variable de la forma  $B^p y_t = y_{t-p}$ .

**Proposición 3.** *Si las desviaciones de una serie, cuyo modelo subyacente es el MNoLIN, se comportan como un proceso  $NID(0, \sigma^2)$  y  $\tau$  es conocido entonces, la serie cumple con los supuestos de Gauss-Markov y los estimadores  $\hat{\beta}$  equivaldrán a los estimadores MCO y por lo tanto,  $\hat{\beta}$  será el mejor estimador lineal insesgado de  $\beta$ .*

*Demostración.* A continuación se demuestra cómo se cumple cada una de las suposiciones de Gauss-Markov:

- Si  $\tau$  es conocido entonces, el MNoLIN se transforma en un modelo lineal, debido a que los parámetros a estimar que causaban la no linealidad del modelo al multiplicarse entre sí ( $\tau$  y  $\beta$ ) ya no causan una no linealidad pues  $\tau$  es ya un dato conocido como cualquier componente de  $\mathbf{X}$  y no necesita ser estimado. Sólo queda estimar  $\beta$ , que es un vector de parámetros lineales y por tanto, el modelo es lineal.
- Multicolinealidad imperfecta, las columnas de  $(\mathbf{X} - \tau \mathbf{D})$  son linealmente independientes. Para demostrar que la matriz  $(\mathbf{X} - \tau \mathbf{D})$  del MNoLIN cumple con este supuesto, demostraremos que la única solución factible del sistema  $(\mathbf{X} - \tau \mathbf{D})\mathbf{a}$  es  $\mathbf{a} = \mathbf{0}$ . La matriz  $(\mathbf{X} - \tau \mathbf{D})$  está dada por  $(\mathbf{X} - \tau \mathbf{D}) = [(\mathbf{X} - \tau \mathbf{D})_1, (\mathbf{X} - \tau \mathbf{D})_2, \dots, (\mathbf{X} - \tau \mathbf{D})_{\lfloor t/m \rfloor}, (\mathbf{X} - \tau \mathbf{D})_{\lfloor t/m \rfloor}']$  donde cada submatriz  $(\mathbf{X} - \tau \mathbf{D})_i$  tiene la forma

$$(\mathbf{X} - \tau \mathbf{D})_i = \begin{bmatrix} 1 & (i-1)m + 1 - \tau & 0 & \dots & 0 \\ 1 & 0 & (i-1)m + 2 - \tau & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & im - \tau \end{bmatrix}.$$

siendo lo anterior válido para  $i = 1, \dots, \lfloor t/m \rfloor$  mientras que  $(\mathbf{X} - \tau \mathbf{D})_{\lfloor t/m \rfloor}$  se define como

$$(\mathbf{X} - \tau \mathbf{D})_{\lfloor t/m \rfloor} =$$



La solución del sistema (5.9) para  $a_3, \dots, a_{m+1}$  es la misma que en el sistema (5.8). La diferencia es que en el sistema (5.9),  $a_1$  no puede tomar cualquier valor, sino que está definida de la siguiente manera

$$\begin{aligned} a_1 + (1 - \tau)a_2 &= 0 \\ a_1 + (m + 1 - \tau)a_2 &= 0 \end{aligned} \quad (5.10)$$

Para que el sistema (5.10) tenga solución, es necesario que el determinante sea diferente de cero. El determinante está dado por

$$m + 1 - \tau - (1 - \tau) = m \neq 0. \quad (5.11)$$

Debido a que el determinante del sistema anterior es diferente de cero, se concluye que el sistema tiene solución única; la trivial ( $a_1 = 0, a_2 = 0$ ). Lo anterior se cumple para cualquier observación que se añada a las primeras  $m$  observaciones. Debido a que  $a_1$  se vuelve cero, las constantes  $a_2, \dots, a_m$  toman también el valor de cero y por lo tanto, las columnas de  $(\mathbf{X} - \tau \mathbf{D})$  son independientes siempre que haya más de  $m + 1$  observaciones y  $\tau$  sea conocido, lo cual implica que haya  $m + 2$  datos.

- Media condicional igual a cero,  $E(\boldsymbol{\mu} | \mathbf{X}, \tau, \mathbf{D}) = 0$ . Este supuesto se cumple debido a que la fuente de variación sigue un proceso NID( $0, \sigma^2$ ). Los procesos NID( $0, \sigma^2$ ) tienen media cero por definición y por consiguiente,  $E(\boldsymbol{\mu} | \mathbf{X}, \tau, \mathbf{D}) = 0$ .
- Varianza finita y constante,  $\text{Var}(\boldsymbol{\mu} | \mathbf{X}) = \sigma^2 < \infty$ . Este supuesto se cumple también por la definición de los procesos NID( $0, \sigma^2$ ).
- Covarianza igual a cero,  $\text{Cov}(\boldsymbol{\mu}(t), \boldsymbol{\mu}(t') | \mathbf{X}) = 0, \forall t \neq t'$ . Este supuesto se cumple también por la definición de los procesos NID( $0, \sigma^2$ ).

Las últimas tres propiedades se deducen por el simple hecho de que la fuente de

variación es un proceso NID(0,  $\sigma^2$ ). Falta demostrar que los estimadores MCNL denotados por  $\hat{\beta}$  equivalen a los MCO.

Como se demostró en el primer punto, si  $\tau$  es conocido entonces, el MNoLIN es un modelo lineal y el estimador MCO de sus parámetros será  $[(\mathbf{X} - \tau \mathbf{D})(\mathbf{X} - \tau \mathbf{D})]^{-1}(\mathbf{X} - \tau \mathbf{D})' \mathbf{y}$ . Dicho estimador es equivalente al estimador MCNL dado por la ecuación (4.11). Esto demuestra que si  $\tau$  es conocido entonces, el estimador MCNL es equivalente al estimador MCO y como se cumplen los supuestos de Gauss-Markov,  $\hat{\beta}$  es el mejor estimador de  $\beta$ . Esto completa la demostración.  $\square$

### Intervalos predictivos

Para el MNoLIN, la varianza del error de pronóstico bajo la suposición de que  $\tau$  es conocido y de que  $\mu$  sigue un proceso NID(0,  $\sigma^2$ ) se deduce de manera similar a la varianza del error de pronóstico del MLIN salvo por el detalle de que hay que sustituir  $x_t$  por  $x_t - \tau$  pues está es la única diferencia entre ambos modelos. De esta manera, para el MNoLIN, la varianza de  $\hat{\mu}_{n+h}$  está dada por

$$\begin{aligned} \text{Var}(\hat{\mu}_{n+h} | \mathbf{X}, \tau) &= \text{Var}(\hat{\beta}_0 | \mathbf{X}, \tau) + 2 \left( \sum_{s=1}^m \text{Cov}(\hat{\beta}_0, \hat{\beta}_s | \mathbf{X}, \tau) D_{s,n+h} \right) (x_{n+h} - \tau) \\ &\quad + \left( \sum_{s=1}^m \text{Var}(\hat{\beta}_s | \mathbf{X}, \tau) D_{s,n+h} \right) (x_{n+h} - \tau)^2 + \sigma^2. \end{aligned}$$

donde  $\text{Var}(\hat{\beta}_0 | \mathbf{X}, \tau)$ ,  $\text{Var}(\hat{\beta}_s | \mathbf{X}, \tau)$  y  $\text{Cov}(\hat{\beta}_0, \hat{\beta}_s | \mathbf{X}, \tau) \forall s$  se obtienen de la siguiente matriz de varianzas y covarianzas

$$\mathbf{VC}(\hat{\beta} | \mathbf{X}, \tau, \mathbf{D}) = \sigma^2 [(\mathbf{X} - \tau \mathbf{D})' (\mathbf{X} - \tau \mathbf{D})]^{-1}. \quad (5.12)$$

### 5.3.2. Las desviaciones siguen un proceso $ARMA(p, q)$ y $\tau$ es conocido

#### Propiedades estadísticas de $\hat{\beta}$

Las propiedades estadísticas de los estimadores obtenidos bajo estos supuestos se enuncian en la siguiente proposición

**Proposición 4.** *Si las desviaciones de una serie, cuyo modelo subyacente es el MNoLIN, siguen un proceso  $ARMA(p, q)$  y  $\tau$  es conocido entonces, la serie en cuestión cumplirá con los supuestos débiles de Gauss-Markov, los estimadores  $\hat{\beta}$  equivaldrán a los estimadores MCO y por tanto,  $\hat{\beta}$  será un estimador insesgado y consistente de  $\beta$ .*

*Demostración.* A continuación se muestra que la serie cumple con cada uno de los supuestos débiles de Gauss-Markov:

1. El modelo subyacente a la serie es lineal. Como en la proposición anterior, la distribución de las desviaciones del modelo no afecta la linealidad del modelo debido a que ésta se suma a la serie y no tiene ninguna otra interacción con los parámetros. El modelo sería no lineal si el parámetro  $\tau$  fuera desconocido pero suponemos que no lo es y por tanto, es como cualquier otro componente de  $\mathbf{X}$ .
2. Multicolinealidad imperfecta, las columnas de  $\mathbf{X}$  son independientes. La demostración de este punto es idéntica que aquella hecha en el caso de ruido  $NID(0, \sigma^2)$  pues la distribución de la fuente de variación no interviene en este punto.
3. Media condicional igual a cero,  $E(\mu | \mathbf{X}) = \mathbf{0}$ . Esto se debe a que por definición, las series que siguen un proceso  $ARMA(p, q)$  tienen valor esperado igual a cero.

Lo anterior demuestra que si las desviaciones del MNoLIN siguen un proceso  $ARMA(p, q)$  entonces, se cumplirán los supuestos débiles de Gauss-Markov. Falta demostrar que los estimadores MCNL equivalen a los MCO. Esta prueba es similar a la hecha en la

proposición anterior pues la distribución de la fuente de variación no se relaciona con dicha demostración. Por lo tanto,  $\hat{\beta}$  es un estimador lineal e insesgado de  $\beta$ . Esto completa la demostración.  $\square$

### Intervalos predictivos

De manera similar a lo hecho para el MLIN, para obtener los intervalos predictivos del MNoLIN de la forma  $y_{n+h} \pm t_{\alpha/2, v} \sqrt{\text{Var}(\hat{\mu}_{n+h} | \mathbf{X})}$  bajo el supuesto de ruido ARMA( $p, q$ ) es necesario suponer que tanto los parámetros del modelo como los del proceso ARMA( $p, q$ ) son conocidos. De esta manera, la varianza del error de pronóstico estará dada por

$$\text{var}(\hat{\mu}_{n+h} | \mathbf{X}, \psi) = \sigma_{\varepsilon}^2 \sum_{j=0}^{h-1} \psi^2, \quad (5.13)$$

donde  $\psi(B) = \theta(B)\phi(B)^{-1}$ .

## 5.4. Estimadores alternativos para el MNoLIN

En la sección anterior vimos como deducir las propiedades estadísticas de una serie cuyo modelo subyacente es el MNoLIN y su fuente de variación se comporta ya sea como un proceso NID( $0, \sigma^2$ ) ó ARMA( $p, q$ ). Dicha deducción supone que  $\tau$  es conocido, sin embargo, en la práctica no es posible conocer con certeza el valor de  $\tau$ . Por esta razón, sería deseable no tener que suponer que  $\tau$  es conocido para deducir las propiedades estadísticas de los estimadores. En esta sección, se propone un procedimiento de estimación que no necesita suponer que  $\tau$  es conocido para deducir las propiedades estadísticas de los estimadores obtenidos con dicho procedimiento. Primero, veremos como obtener dichos estimadores y luego, veremos los resultados de un experimento en el que se compara este método de estimación con el procedimiento de la búsqueda lineal con sección dorada.

### 5.4.1. Estimadores analíticos

La deducción de los estimadores alternativos se basa en el hecho de que el MNoLIN ajusta una línea recta a cada conjunto de datos que representa una estación y supone que todas las rectas tienen un origen común en algún lugar del plano.

Sean  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  un conjunto de coordenadas sobre las cuales se regresa una línea recta. Entonces, el estimador MCO de la pendiente de la recta (representado por  $\hat{\beta}_{\text{MCO}}$ ) será

$$\hat{\beta}_{\text{MCO}} = \frac{n \sum_{t=1}^n y_t x_t - \sum_{t=1}^n y_t \sum_{t=1}^n x_t}{n \sum_{t=1}^n x_t^2 - \left( \sum_{t=1}^n x_t \right)^2}. \quad (5.14)$$

Haciendo lo anterior para cada estación, es decir, ajustando una línea recta a cada conjunto de coordenadas que representan una estación  $((x_{1,s}, y_{1,s}), (x_{2,s}, y_{2,s}), \dots, (x_{k_s,s}, y_{k_s,s}), (s = 1, \dots, m))$  tenemos

$$\bar{\beta}_s = \frac{k_s \sum_{i=1}^{k_s} y_{i,s} x_{i,s} - \sum_{i=1}^{k_s} y_{i,s} \sum_{i=1}^{k_s} x_{i,s}}{k_s \sum_{i=1}^{k_s} x_{i,s}^2 - \left( \sum_{i=1}^{k_s} x_{i,s} \right)^2}, \quad (5.15)$$

donde  $\bar{\beta}_s$  es el estimador MCO de la pendiente de cada recta regresada sobre los datos correspondientes a la estación  $s$ . Los estimadores  $\bar{\beta}_s$  tienen las siguientes ventajas:

- No necesitan de un algoritmo para ser estimados pues están dados por fórmulas analíticas.
- Si la fuente de variación se comporta como un proceso NID(0,  $\sigma^2$ ) entonces,  $\bar{\beta}_s$  son los mejores estimadores lineales insesgados.
- Si la fuente de variación se comporta como un proceso ARMA( $p, q$ ) entonces, los estimadores  $\bar{\beta}_s$  son insesgados y consistentes.
- Los resultados anteriores no suponen que  $\tau$  sea conocido.

Lo anterior se deduce del hecho de que los estimadores  $\bar{\beta}_s$  son los estimadores MCO de la pendiente de cada una de las rectas que subyacen a cada estación generada con el MNoLIN. El hecho de que el modelo subyacente a cada estación sea una recta podría no ser claro a simple vista pues el MNoLIN es un modelo de regresión no lineal. Sin embargo, este hecho se puede demostrar como se verá a continuación.

**Proposición 5.** *El modelo subyacente a cada una de las estaciones de una serie, cuyo modelo subyacente es el MNoLIN, es una línea recta con pendiente  $\beta_s$ .*

*Demostración.* Basta mirar al modelo subyacente a cada estación de manera detenida y con algunas simplificaciones tenemos

$$\begin{aligned}
 y_{i,s} &= \beta_0 + \beta_s(x_{i,s} - \tau) + \mu_{i,s}, \\
 &= \beta_0 + \beta_s x_{i,s} - \beta_s \tau + \mu_{i,s}, \\
 &= \beta_0 - \beta_s \tau + \beta_s x_{i,s} + \mu_{i,s}, \\
 &= \alpha_s + \beta_s x_{i,s} + \mu_{i,s}.
 \end{aligned} \tag{5.16}$$

Evidentemente, la ecuación (5.16) describe una línea recta con desface  $\alpha_s = \beta_0 - \beta_s \tau$  y pendiente  $\beta_s$ . Esto completa la demostración.  $\square$

Las desviaciones de las observaciones contenidas en cada estación se comportan de la misma manera que las desviaciones de la serie pues son una muestra aleatoria de éstas. Así por ejemplo, si las desviaciones de la serie siguen un proceso NID( $0, \sigma^2$ ) entonces, las desviaciones de todas las observaciones correspondientes a una estación seguirán un proceso NID( $0, \sigma^2$ ) y por tanto,  $\bar{\beta}_s \forall s$  serán los mejores estimadores lineales insesgados.

Alternativamente, si las desviaciones de la serie se comportan como un proceso ARMA( $p, q$ ) entonces, las desviaciones correspondientes a las observaciones de cada estación seguirán también un proceso ARMA( $p, q$ ) aunque con valores diferentes de  $p$  y  $q$ . Sin embargo, los estimadores  $\bar{\beta}_s \forall s$  seguirán siendo lineales y consistentes.

A priori, se podría pensar que los estimadores alternativos, a los cuales nos referiremos como analíticos, dominan a los estimadores MCNL. Sin embargo, debemos notar que los estimadores analíticos sólo están disponibles para los parámetros  $\beta_s \forall s$ . No hay estimadores analíticos para  $\beta_0$  ni para  $\tau$ . En segundo lugar, los estimadores analíticos tienen muy pocos grados de libertad. Por ejemplo, si tenemos  $dm$  observaciones, cada uno de los estimadores  $\beta_s \forall s$  utilizará sólo  $d$  observaciones para ser estimado. Estas razones hacen necesaria la comparación contra los estimadores MCNL. A continuación se muestran los resultados de un experimento con este propósito.

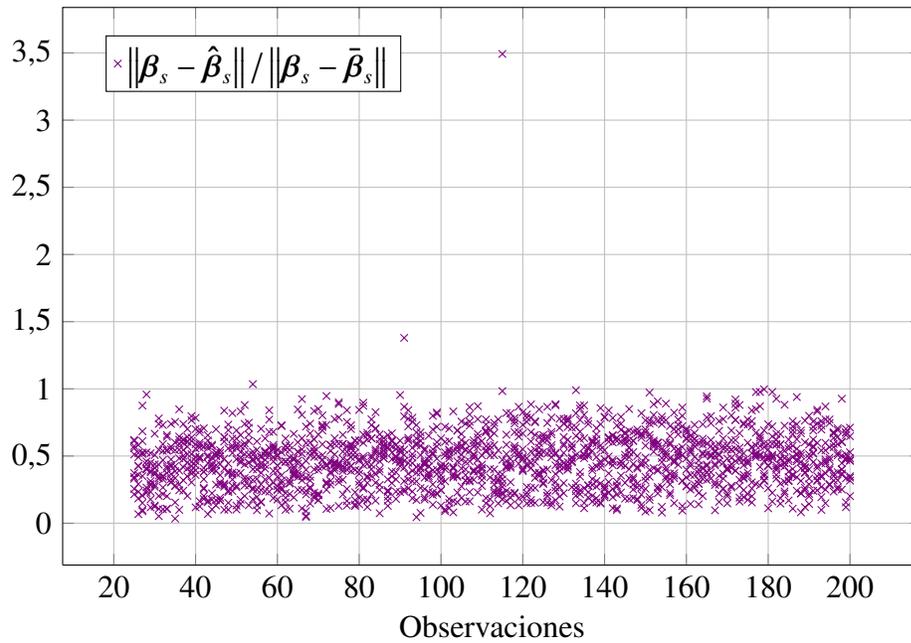
### 5.4.2. Experimentación

A continuación, describiremos los lineamientos observados y los resultados obtenidos de un experimento en el que se comparan los estimadores MCNL contra los analíticos. Debido a que hay estimadores analíticos sólo para  $\beta_s \forall s$ , la comparación se limita a la comparación de estos  $m$  parámetros.

#### Lineamientos observados durante el experimento

Para conocer cual de los estimadores del MNOLIN es mejor, utilizamos las series generadas en la sección anterior. Esta vez, comparamos las distancias euclidianas que hay entre los parámetros y los estimadores. Aquellos estimadores cuyas distancias euclidianas a los parámetros fueran las menores, serán los mejores. Esto se debe a que un estimador eficiente debe tender al parámetro que estima, sobre todo si la muestra es muy grande. Esto significa que la distancia euclidiana a los parámetros debe ser lo más pequeña posible. Por esta razón, debemos elegir los estimadores cuyas distancias euclidianas sean lo menor posible.

Sean  $\beta_s = [\beta_1, \dots, \beta_s]'$ ,  $\hat{\beta}_s = [\hat{\beta}_1, \dots, \hat{\beta}_s]'$  y  $\bar{\beta}_s = [\bar{\beta}_1, \dots, \bar{\beta}_s]'$ . Al igual que la suma de desviaciones cuadradas, las distancias euclidianas son magnitudes altamente variables y



**Figura 5.1** – Los cocientes  $\|\beta_s - \hat{\beta}_s\| / \|\beta_s - \bar{\beta}_s\|$ .

por ello, estudiaremos la razón  $\|\beta_s - \hat{\beta}_s\| / \|\beta_s - \bar{\beta}_s\|$ . Si la razón es menor a uno entonces, significa que para esa serie, los estimadores MCNL son mejores que los analíticos. Si la razón es mayor a uno, se concluirá lo contrario.

## Resultados

Las razones, para cada una de las series, se muestran en la figura 5.1. En la figura 5.2 se muestra el histograma de frecuencias. En la figura 5.1 se observa que la razón, para casi todas las series, está distribuida en el intervalo  $[0, 1]$ . Sólo hubo 3 casos en los que la razón fue mayor a 1. En el histograma de frecuencias, se aprecia que los radios se comportan de una manera aproximadamente normal en el intervalo  $[0, 1]$ , aunque la distribución está cargada más hacia la izquierda. La media es aproximadamente 0,5 lo que significa que en promedio podemos esperar que la distancia euclidiana  $\|\beta_s - \hat{\beta}_s\|$  será 50 % más pequeña que la distancia  $\|\beta_s - \bar{\beta}_s\|$ .

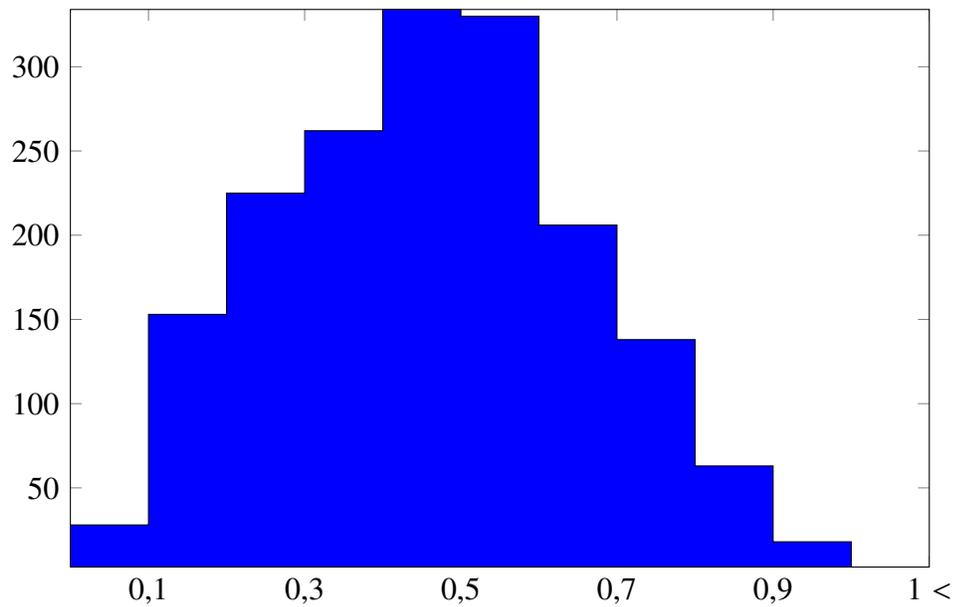


Figura 5.2 – Histograma de frecuencias de la razón  $\|\beta - \hat{\beta}_s\| / \|\beta - \bar{\beta}_s\|$ .

## 5.5. Conclusiones

En este capítulo, se dedujeron las propiedades estadísticas de los parámetros, tanto del MLIN como del MNoLIN, así como también los intervalos predictivos de ambos modelos.

Para el MLIN, la deducción se realizó con relativa facilidad, debido a que éste es un modelo de regresión lineal. Los estimadores MCO de este modelo fueron los mejores estimadores lineales insesgados para el caso de fuente de variación  $NID(0, \sigma^2)$  y fueron estimadores insesgados y consistentes para el caso de fuente de variación  $ARMA(p, q)$ . La naturaleza lineal del MLIN permitió que las deducciones anteriores, así como la de sus intervalos predictivos, se realizaran sin la necesidad de suponer que alguno de los parámetros a estimar era conocido.

En cambio, la deducción de las propiedades estadísticas de los estimadores del MNoLIN, así como la de sus intervalos predictivos, no se pudo realizar de manera directa debido a que el MNoLIN es un modelo de regresión no lineal. Cuando esto ocurre, lo que se hace en la literatura es suponer que algunos de los parámetros del modelo son conocidos. Esto con el objetivo de simplificar el análisis estadístico del modelo. Para el caso del

MNoLIN, la mejor opción fue suponer que el parámetro  $\tau$  era conocido. Así, el MNoLIN se transformó en un modelo de regresión lineal similar al MLIN. De esta manera, si la fuente de variación de la serie es un proceso NID( $0, \sigma^2$ ) y  $\tau$  es conocido, los estimadores MCNL son los mejores estimadores lineales insesgados. Por otra parte, si  $\tau$  es conocido pero la fuente de variación sigue un proceso ARMA( $p, q$ ) entonces, los estimadores MCNL son insesgados y consistentes. La deducción de los intervalos predictivos del MNoLIN descansa también sobre este supuesto.

Un conjunto de estimadores alternativos fueron propuestos para el MNoLIN; los estimadores analíticos. Dichos estimadores tienen varias ventajas sobre los estimadores MCNL. Por ejemplo, no necesitan de la implementación de complicados algoritmos de optimización no lineal para ser calculados pues están dados por formulas analíticas. Además, los estimadores analíticos no necesitan suponer que  $\tau$  es conocido para concluir que son los mejores estimadores lineales insesgados en el caso de ruido NID( $0, \sigma^2$ ) ó que son insesgados y consistentes en el caso de ruido ARMA( $p, q$ ). La desventaja principal de los estimadores analíticos radica en que sólo los hay disponibles para los parámetros  $\beta_s \forall s$ .

En un experimento en el que se compararon los estimadores MCNL contra los estimadores analíticos, se demostró que los primeros, tienen en promedio una distancia euclidiana a los parámetros estimados 50 % menor que la de los estimadores analíticos. Por esta razón se concluye que los MCNL son los mejores estimadores de los parámetros del MNoLIN, a pesar de que sus propiedades estadísticas como computacionales sean dominadas por aquéllas de los estimadores analíticos. Los estimadores MCNL serán utilizados por el resto de este trabajo.

En el siguiente capítulo, se estudiarán seis series temporales con tendencia lineal y estacionalidad creciente para comparar los dos modelos propuestos contra otra alternativa *apta* de la literatura. Las comparaciones se harán en capítulos posteriores.

---

---

# CAPÍTULO 6

---

## Series de ejemplo

Seis series reales, que presentan patrones de tendencia lineal y estacionalidad creciente, serán estudiadas en este capítulo. Las series fueron tomadas de diversos artículos y libros y representan actividades económicas del tipo venta y demanda. Las series serán empleadas para comparar el desempeño del MLIN, el MNoLIN y la alternativa *apta* de la literatura. Además, se describen tres *escenarios* bajo los cuales las series serán pronosticadas para así conocer el desempeño de cada modelo bajo diferentes circunstancias. Este capítulo será la base de la que partirán los tres capítulos siguientes, pues estos capítulos utilizarán las seis series para comparar el desempeño de los tres modelos de pronóstico bajo diferentes circunstancias o escenarios.

### 6.1. Descripción de las series

Las STTLEC pueden aparecer en actividades industriales muy diferentes. Así, podemos esperar su aparición en actividades tan diversas como lo son los datos de habitaciones de hotel o las ventas mexicanas de vehículos subcompactos. Por esta razón, en el cuadro 6.1 aparecen series que lo único que tienen en común es exhibir patrones de tendencia lineal y estacionalidad creciente. A las series del cuadro 6.1 se les ha asignado un número que será utilizado para identificar a cada serie en los cuadros en los que se muestren los resultados de los experimentos. Dicho número se muestra en la primera columna. En la segunda columna del cuadro 6.1, se describe el proceso que generó a cada serie. Así, la

**Cuadro 6.1** – Resumen de las series de ejemplo.

#	Serie	Procedencia	Meses
1	G11 (serie mensual, los períodos no están disponibles y su naturaleza es desconocida)	<a href="#">Makridakis et al. (1982)</a>	84
2	Ventas mexicanas de vehículos subcompactos de enero de 1995 a diciembre de 2004	<a href="#">INEGI (2005)</a>	108
3	Ventas de la «compañía X» de enero de 1965 a mayo de 1971	<a href="#">Chatfield y Prothero (1973)</a>	77
4	Datos de reservaciones hoteleras (el período es desconocido)	<a href="#">Bowerman y O'Connell (1993)</a>	168
5	Ventas totales estadounidenses de todas las tiendas detallistas (de enero de 1983 a diciembre de 1989)	<a href="#">Hanke y Reitsh (1996)</a>	84
6	Ventas industriales francesas (en francos) de papel para escribir e imprimir (el período es desconocido)	<a href="#">Makridakis et al. (1982)</a>	120

serie 2 corresponde a las ventas mexicanas de vehículos subcompactos. La serie 1, llamada «G11» recibe este nombre debido a que es la onceava serie en el compendio de series de Makridakis llamado «La competición M». La naturaleza de la serie es desconocida, lo cual es un caso común, como sucede con la serie 3, las ventas de la «compañía X». Cuando la naturaleza de una serie no es revelada, suele deberse a que el propietario de la información no desea que su competencia tenga información sobre sus ventas y menos sobre lo que proyecta vender. Salvo por el caso de las series 1 y 3, la naturaleza del resto de las series es conocida. En la tercera columna del cuadro 6.1 se presenta la fuente de donde se obtuvo la información sobre cada serie. En la cuarta columna, se muestra el número de meses durante el cual la serie exhibe patrones de tendencia lineal y estacionalidad creciente. Por ejemplo, aunque había más información sobre las ventas nacionales de vehículos subcompactos, esta información precede a un período de derrumbe en el cual la serie mostraba otros patrones. Tomamos la parte en la que esta serie exhibe patrones de tendencia lineal y estacionalidad creciente, pues suponemos que la serie se seguirá comportando de esa manera. Las gráficas de cada serie del cuadro 6.1 se muestran en las figuras 6.1, 6.2, 6.3, 6.4, 6.5 y 6.6.

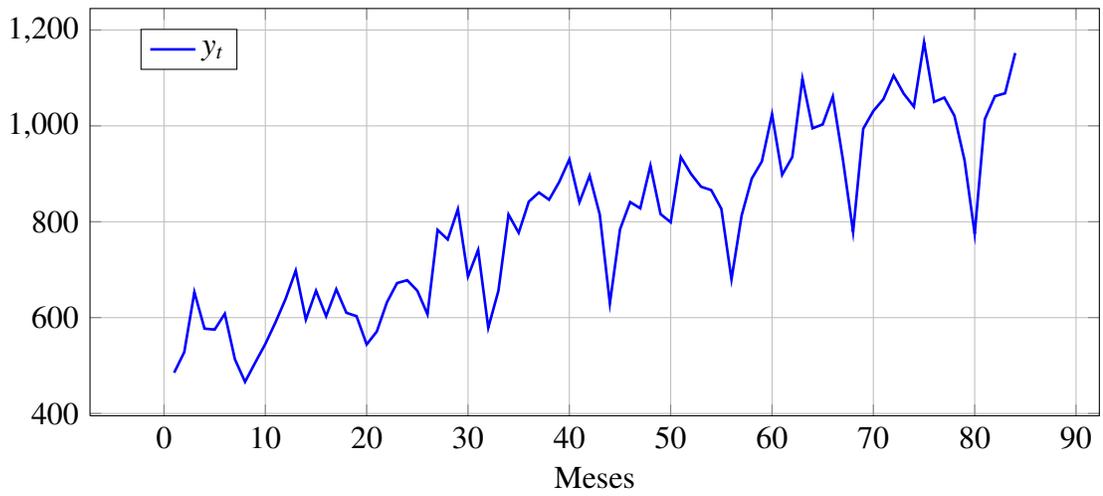


Figura 6.1 – Serie de ejemplo 1.

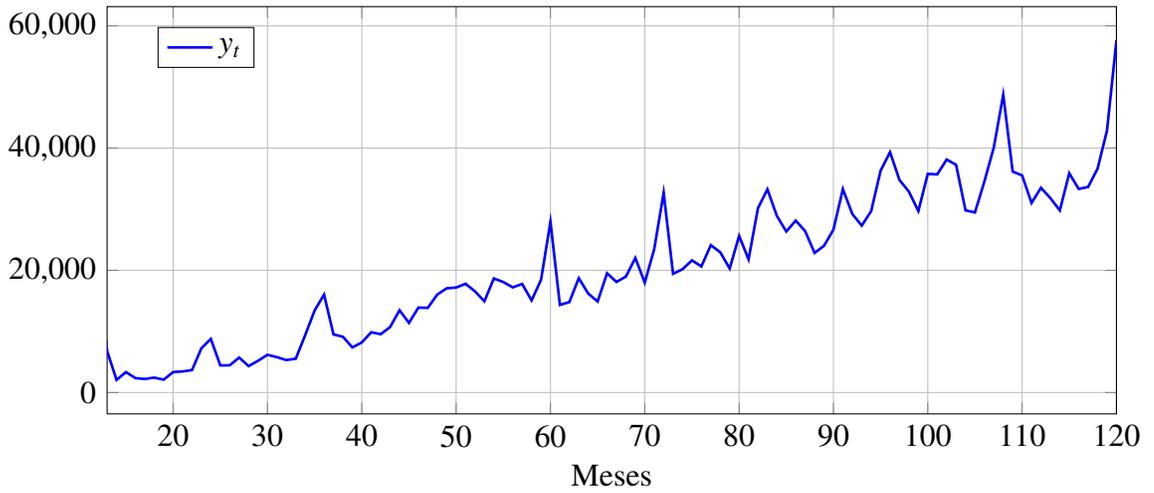


Figura 6.2 – Serie de ejemplo 2.

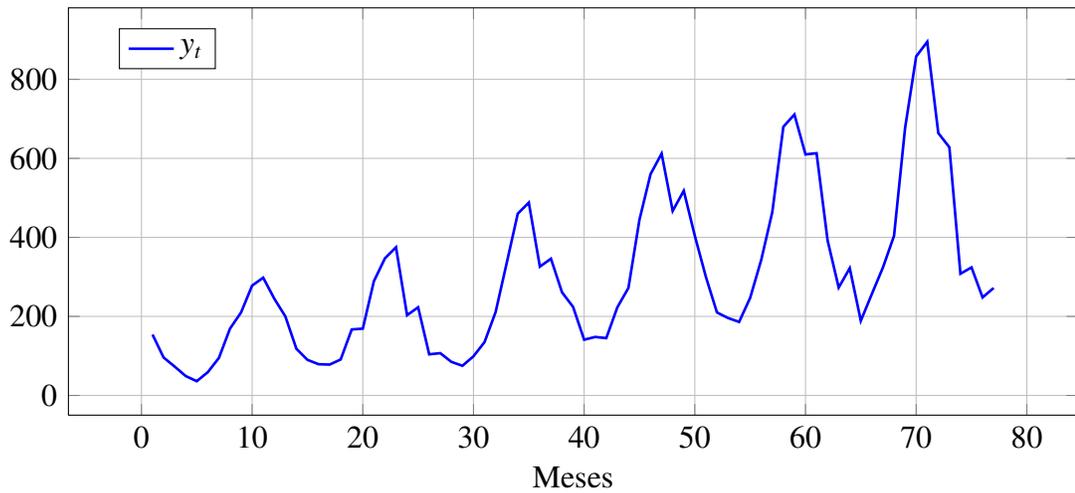


Figura 6.3 – Serie de ejemplo 3.

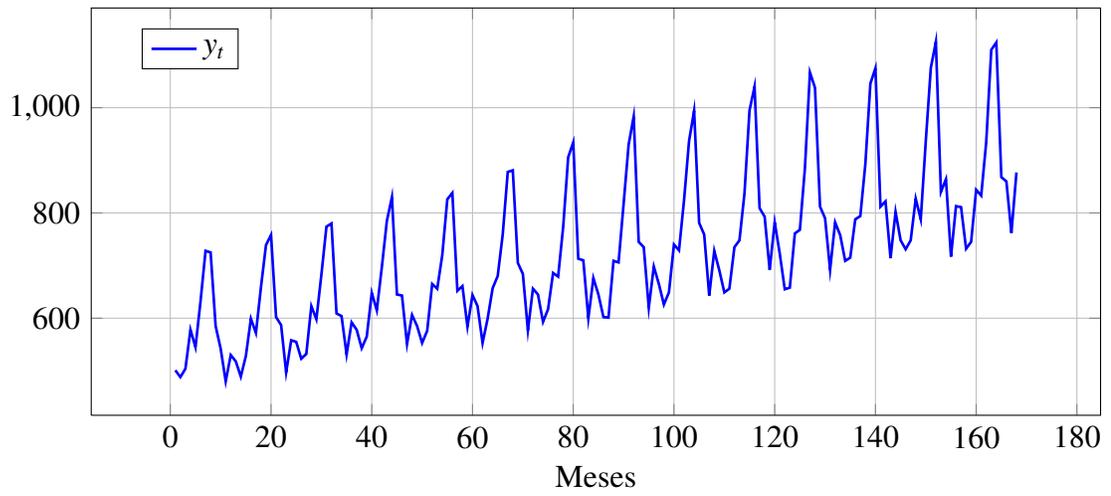


Figura 6.4 – Serie de ejemplo 4.

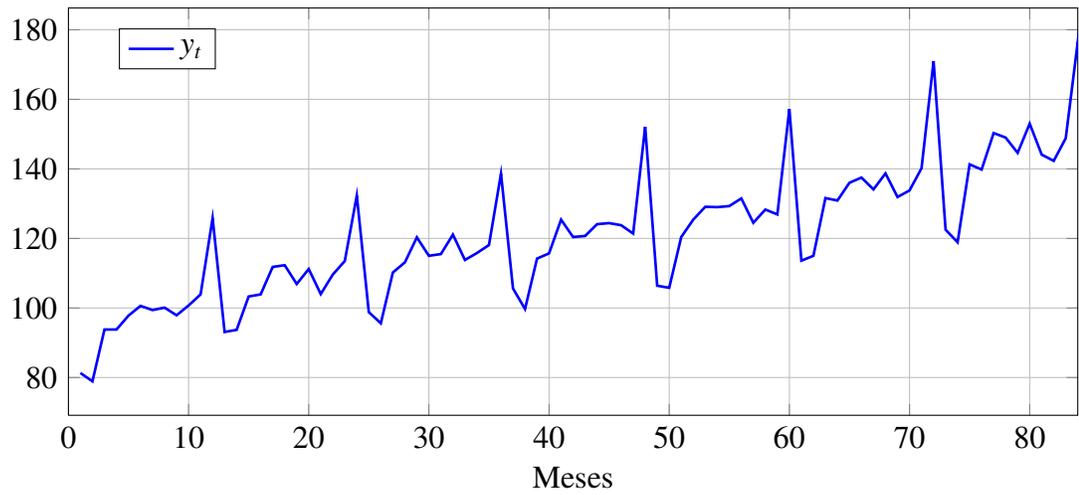


Figura 6.5 – Serie de ejemplo 5.

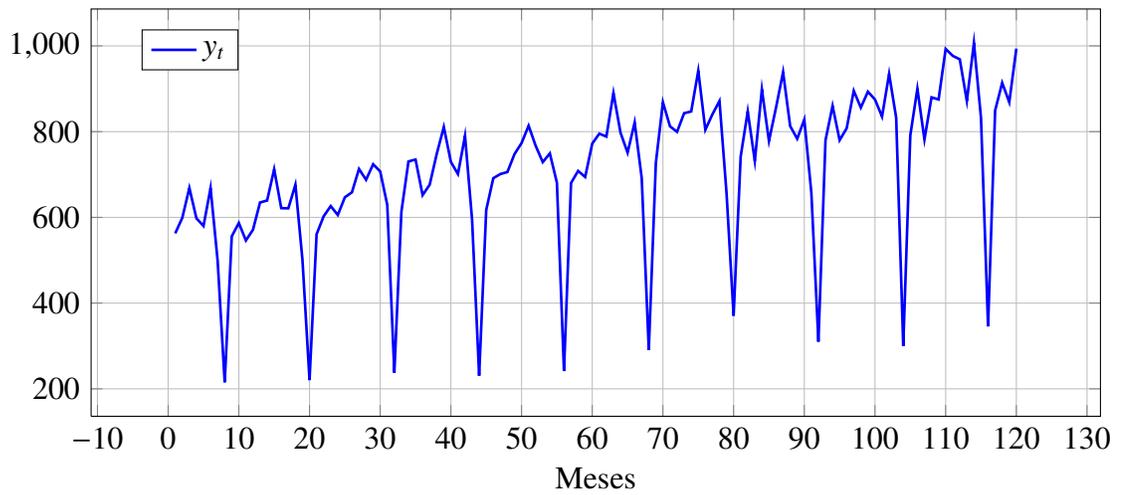


Figura 6.6 – Serie de ejemplo 6.

En las gráficas se observa que los períodos ocurren cada año y que las estaciones corresponden a cada mes. Por lo tanto, se trabajará con las series de cuadro 6.1 bajo la premisa de que  $m = 12$ . Sin embargo, los métodos utilizados pueden trabajar con cualquier valor de  $m$ .

## 6.2. El modelo *apto*

Para conocer la eficacia de los modelos propuestos, en los capítulos 7, 8 y 9 se comparará el desempeño del  $MLIN$  y el  $MNoLIN$  utilizando las series reales de este capítulo. También, se utilizará el sexto modelo con filtros diferenciadores y componentes deterministas, al que nos referiremos simplemente  $MALtDB$  (léase modelo altamente determinista de [Bowerman et al. \(1990\)](#)), para comparar el desempeño de los modelos propuestos contra alternativas de pronóstico establecidas en la literatura.

El  $MALtDB$  pertenece a la metodología de modelos  $ARMA(p, q)$  con filtros diferenciadores y componentes deterministas (véase la ecuación (2.19), página 19) y está especialmente diseñado para el pronóstico de  $STTLEC$ . Su estructura altamente determinista le permite pronosticar series temporales con observaciones faltantes. Según [Bowerman et al. \(1990\)](#), el  $MALtDB$  es mejor que la metodología  $ARIMA(p, d, q)$  con transformación logarítmica para el pronóstico de series que exhiben una marcada tendencia lineal. Según [Franses y Koehler \(1998\)](#), el  $MALtDB$  es el mejor modelo de pronóstico especialmente un paso adelante al compararlo con diversos modelos con filtros diferenciadores. Las razones anteriores hacen del  $MALtDB$  un modelo *apto* para el pronóstico de  $STTLEC$  y un buen modelo para comparar el desempeño tanto del  $MLIN$  como del  $MNoLIN$ .

### 6.3. Conjuntos de datos o escenarios

Nuestro interés consiste en comparar el desempeño del  $MLIN$ , el  $MNoLIN$  y el  $MALtDB$  al pronosticar  $STTLEC$  con tres diferentes conjuntos de datos o condiciones a los cuales haremos referencia como «escenarios» y que consisten en lo siguiente:

**Escenario 1** El número de observaciones es regular. Esto quiere decir que hay más de  $4m$  observaciones. Este escenario será el tema del capítulo 7;

**Escenario 2** El número de observaciones es escaso. Esto quiere decir que hay entre  $2m$  y  $4m$  observaciones. Este escenario será discutido en el capítulo 8;

**Escenario 3** El número de observaciones es regular pero hay datos faltantes aleatoriamente distribuidos en la serie. Este escenario será el tema del capítulo 9.

Para el primer caso, se pronosticarán las series modelando las autocorrelaciones de los residuales con modelos  $ARMA(p, q)$ . Para los otros dos casos, el único supuesto viable es el de ruido  $NID(0, \sigma^2)$ . Diversos estadísticos serán tomados en cada experimento para conocer la eficacia de los modelos en diversos ámbitos, como la magnitud promedio del error absoluto de pronóstico o la eficacia de los intervalos predictivos.

### 6.4. Conclusiones

Seis series, con patrones de tendencia lineal y estacionalidad creciente, han sido presentadas. También, han sido discutidas las razones por las cuales el  $MALtDB$  es considerado el modelo *apto* para comparar el desempeño del  $MLIN$  y del  $MNoLIN$ . En los siguientes tres capítulos, se compararán los tres modelos para conocer su desempeño bajo diversas circunstancias tales como: series con número regular de observaciones, series con escasa información y series con datos faltantes.

---

---

# CAPÍTULO 7

---

## Primer escenario: series con observaciones regulares

**A**nalizaremos los resultados obtenidos al utilizar las seis series del capítulo 6 para comparar el desempeño del MLIN, el MNoLIN y el MAltDB. Los resultados no se limitan a comparar el siguiente dato sino las siguientes doce observaciones. Adicionalmente, se presentan resultados relacionados con la eficacia de los intervalos predictivos como lo son la cobertura observada y la estrechez. Lo anterior se hace con base en estadísticos definidos en los lineamientos observados. Después, se discuten los resultados y con base en cada estadístico, se muestran resultados relacionados con el pronóstico, la cobertura observada y la estrechez de los intervalos predictivos. Al final, se presentarán las conclusiones de este comparativo.

### 7.1. Lineamientos observados

Para conocer la eficacia de los modelos al pronosticar varios pasos hacia adelante, tomamos las primeras  $n - 3m$  observaciones y pronosticamos las siguientes  $h = 1, \dots, (m = 12)$  observaciones. Esto lo hicimos así para las primeras  $n - 3m + j$  observaciones con  $j = 1, \dots, 2m$ . Es decir, pronosticamos un total de 24 veces las siguientes 12 observaciones. De los resultados se obtuvieron, para cada valor de  $h$ , los siguientes estadísticos:

- *Cobertura Observada del Intervalo Predictivo* (COB). Es la probabilidad observada para cada valor de  $h$  ( $h = 1, \dots, 12$ ) de que el dato pronosticado ocurra dentro del intervalo predictivo. Esto es,

$$\text{COB}_h = \frac{1}{2m} \sum_{j=1}^{2m} z_{h,n-3m+j},$$

donde  $z_{h,n-3m+j} \forall j$  son variables binarias que valen 1 si la observación pronosticada  $h$  pasos hacia adelante en el tiempo  $t = n - 3m + j$  ocurre dentro de los intervalos predictivos ó 0 de otro modo.

- *Estrechez Media del Intervalo Predictivo* (EST). Es la media de las amplitudes de los intervalos predictivos para cada  $h$ . Esto es,

$$\text{EST}_h = \frac{1}{2m} \sum_{j=1}^{2m} w_{h,n-3m+j},$$

donde  $w_{h,n-3m+j}$  es la amplitud de los intervalos predictivos para el valor pronosticado  $h$  pasos hacia adelante al tiempo  $t = n - 3m + j$ .

- *Desviación Absoluta Media* (DESV). Es la media de las desviaciones absolutas para cada  $h$ . Esto es,

$$\text{DESV}_h = \frac{1}{2m} \sum_{j=1}^{2m} |\hat{e}_{h,n-3m+j}|, \quad (7.1)$$

donde  $\hat{e}_{h,n-3m+j} = y_{n-3m+j+h} - \hat{y}_{h,n-3m+j}$  siendo  $\hat{y}_{h,n-3m+j}$  el pronóstico  $h$  pasos hacia adelante hecho en el tiempo  $t = n - 3m + j$  para la observación  $y_{n-3m+j+h}$ .

- *Desviación Absoluta Porcentual* (DESV %). Es la media  $\forall h$  de las desviaciones absolutas del pronóstico divididas entre el dato pronosticado. Esto es,

$$\text{DESV \%}_h = \frac{1}{2m} \sum_{j=1}^{2m} \frac{|\hat{e}_{n-3m+j+h}|}{\hat{y}_{n-3m+j+h}}$$

Todos los desarrollos estadísticos realizados en este trabajo suponen que la fuente de variación sigue un proceso NID( $0, \sigma^2$ ) o ARMA( $p, q$ ) en el caso de haber correlaciones en los residuales. Sin embargo, durante la realización del experimento, no fue necesario utilizar modelos ARMA( $p, q$ ) para eliminar las correlaciones de los datos. Para realizar esta labor, fue suficiente suponer que las desviaciones seguían un proceso AR( $p$ ). Los parámetros del proceso AR( $p$ ) fueron estimados maximizando su función de verosimilitud.

## 7.2. Resultados

Los resultados de este experimento se muestran en el Apéndice C. Los cuadros 7.1, 7.2 y 7.3 son tres ejemplos de estos cuadros de resultados.

**Cuadro 7.1** – Esc. 1; serie 1; MLIN; AR(1).

$h$	COB	EST	DESV	DESV %
1	0,750	163,235	52,468	5,898
2	0,833	171,842	52,764	5,821
3	0,791	172,797	52,038	5,673
4	0,833	172,910	50,133	5,441
5	0,875	172,924	48,551	5,206
6	0,916	172,925	45,559	4,819
7	0,875	172,925	46,098	4,817
8	0,875	172,926	46,412	4,810
9	0,875	172,926	45,304	4,643
10	0,875	172,926	45,980	4,708
11	0,875	172,926	45,041	4,576
12	0,875	172,926	46,865	4,745

**Cuadro 7.2** – Esc. 1; serie 1; MNoLIN; AR(12).

$h$	COB	EST	DESV	DESV %
1	0,791	141,480	43,282	4,773
2	0,833	141,480	41,313	4,502
3	0,833	141,480	41,667	4,497
4	0,875	141,480	40,509	4,352
5	0,875	141,480	42,319	4,486
6	0,916	141,480	39,516	4,113
7	0,875	141,480	41,089	4,265
8	0,875	141,480	42,822	4,417
9	0,875	141,480	43,290	4,444
10	0,875	141,480	45,103	4,621
11	0,875	141,480	43,323	4,414
12	0,875	141,480	42,669	4,325

**Cuadro 7.3** – Esc. 1; serie 1; MALtDB; AR(1, 3).

$h$	COB	EST	DESV	DESV %
1	0,583	123,962	51,040	5,652
2	0,583	123,962	48,557	5,280
3	0,625	123,962	47,715	5,131
4	0,625	123,962	49,561	5,292
5	0,625	123,962	47,993	5,069
6	0,666	123,962	44,104	4,614
7	0,625	123,962	45,572	4,731
8	0,625	123,962	45,253	4,636
9	0,625	123,962	46,240	4,757
10	0,625	123,962	45,382	4,635
11	0,666	123,962	43,193	4,352
12	0,666	123,962	45,016	4,521

En cada una de las columnas de los cuadros del Apéndice C (y por consiguiente, de los cuadros 7.1, 7.2 y 7.3), se muestran los estadísticos COB, EST, DESV y DESV % para los doce valores de  $h$  que representan los meses de un año. Además, en cada una de las descripciones de los cuadros aparece la siguiente información: el escenario (Esc. 1: el número de datos es regular), el número de la serie pronosticada (serie 1 para los cuadros 7.1, 7.2 y 7.3); el modelo con el que la serie se pronosticó (MLIN para el cuadro 7.1, MNoLIN para el cuadro 7.2 y MALtDB para el cuadro 7.3); fuente de variación con la que se modelaron los residuales (AR(1), AR(12) y AR(1, 3) para los cuadros 7.1 y 7.2 y 7.3 respectivamente).

Los resultados para la serie 1 indican que los modelos con mejor cobertura de intervalos predictivos son el MLIN y el MNoLIN. Esto se debe a que estos modelos arrojaron estadísticos COB más cercanos a la cobertura teórica, que es 95 %. El MALtDB tuvo un desempeño pobre en este aspecto de la comparación para la serie 1. Aunque los valores del estadístico COB del MLIN y del MNoLIN fueron buenos, distan de la cobertura teórica. Esto se debe a que el análisis realizado al deducir los intervalos predictivos modelando las desviaciones como ruido AR( $p$ ) no toma en cuenta la variabilidad de los parámetros. Esta es una desventaja de los modelos autorregresivos.

El segundo estadístico de interés es la amplitud media de los intervalos predictivos. Siempre es deseable que los intervalos predictivos sean tan reducidos como sea posible, siempre que cumplan o sean los más cercanos en alcanzar la cobertura teórica. En este aspecto, el modelo con intervalos predictivos más reducidos para la serie 1 fue el MAltDB. Esto se debe a que este modelo arrojó los valores más reducidos del estadístico Est para los doce valores de  $h$ . Le siguió el MNoLin y por último, el modelo menos eficaz en este aspecto fue el MLin.

Por último, el modelo que mejor pronosticó fue el MNoLin. Esto se debe a que este modelo minimizó los valores de los estadísticos Desv y por consiguiente del estadístico Desv %, a lo largo de los doce horizontes de pronóstico. El modelo que le siguió fue el MAltDB y por último el MLin.

Los resultados analizados en esta sección se desprenden del pronóstico de la serie 1 con el MLin, el MNoLin y el MAltDB. Sin embargo, este tipo de análisis no se puede realizar con todos los cuadros del Apéndice C pues la extensión del mismo sería muy grande. Así mismo, estamos interesados en realizar pruebas estadísticas que nos digan si los resultados de un modelo son significativamente diferentes a los obtenidos con otro. En particular, estamos interesados en realizar pruebas estadísticas que respondan a las siguientes interrogantes que surgen al comparar dos modelos:

1. ¿Cuál de los modelos arroja valores del estadístico Cob más cercanos a la cobertura teórica?
2. ¿Cuál de los modelos arroja los menores valores del estadístico Est?
3. ¿Cuál de los modelos minimiza los valores del estadístico Desv % y por consiguiente, del estadístico Desv?

En las secciones [7.2.3](#), [7.2.1](#) y [7.2.2](#) se da respuesta a cada una de las interrogantes anteriores.

### 7.2.1. Comparación de la cobertura de los intervalos predictivos basada en el estadístico $COB$

Existen dos criterios importantes que los intervalos predictivos deben cumplir. El primero es que se acerquen lo más posible a su valor teórico de cobertura y el segundo es que sean tan delgados como sea posible. Por ejemplo, si el valor teórico de cobertura, designado por el pronosticador, es de 95 %, se espera que en promedio, el 95 % de las observaciones ocurran dentro del intervalo predictivo. En esta sección, haremos pruebas de hipótesis sobre la cobertura observada de los intervalos predictivos mientras que en la siguiente, haremos las pruebas relacionadas con la estrechez.

Para comparar la proximidad de la cobertura observada de los intervalos predictivos de dos modelos a su cobertura teórica, primero, es necesario medir esta proximidad para cada modelo. Para medir la proximidad de la cobertura observada de los intervalos predictivos a su valor teórico, designado como 95 %, medimos la distancia de cada uno de los elementos de la columna  $COB$  al valor 0,95. Por ejemplo, para el caso de la serie 1 bajo el supuesto de ruido  $AR(1)$ , los resultados obtenidos con el  $MLIN$  (distancia entre la cobertura observada y la teórica) se obtiene de la diferencia absoluta entre la columna  $COB$  del cuadro 7.1 y el vector  $[0,95, \dots, 0,95]'$ . Esto es

$$COB_{dist} = \begin{bmatrix} 0,750 \\ 0,833 \\ 0,791 \\ 0,833 \\ 0,875 \\ 0,916 \\ 0,875 \\ 0,875 \\ 0,875 \\ 0,875 \\ 0,875 \\ 0,875 \end{bmatrix} - \begin{bmatrix} 0,950 \\ 0,950 \\ 0,950 \\ 0,950 \\ 0,950 \\ 0,950 \\ 0,950 \\ 0,950 \\ 0,950 \\ 0,950 \\ 0,950 \\ 0,950 \end{bmatrix} = \begin{bmatrix} 0,200 \\ 0,117 \\ 0,159 \\ 0,117 \\ 0,075 \\ 0,034 \\ 0,075 \\ 0,075 \\ 0,075 \\ 0,075 \\ 0,075 \\ 0,075 \end{bmatrix}. \quad (7.2)$$

El vector  $COB_{dist}$  representa la distancia para cada uno de los doce valores de  $h$  entre

la cobertura observada de los intervalos predictivos del  $MLIN$  y la cobertura teórica para la serie 1 bajo el supuesto de ruido  $AR(1)$ . Lo ideal sería que el valor de  $COB_{dist}$  fuera  $0$ , pues esto implicaría que la cobertura observada es igual a la teórica. Entonces, diremos que el modelo cuyo estadístico  $COB_{dist}$  sea significativamente menor al mismo estadístico de los otros dos modelos, será el mejor. Este resultado puede variar en cada serie. Para realizar estas pruebas, utilizamos la prueba de rangos con signo de Wilcoxon. Esta prueba es ideal para comparar dos variables cuyas realizaciones están correlacionadas. En este caso, las variables serán los modelos. Las observaciones de estos modelos estarán correlacionadas debido a que cada valor de  $h$  influye en los valores del estadístico  $COB_{dist}$ . Utilizaremos la prueba de Wilcoxon para comparar los estadísticos  $COB_{dist}$  de cada par de modelos para cada serie. Diremos que las diferencias son significativas, si los valor- $P$  son menores al nivel de significancia  $\alpha = 0,05$ . En el cuadro 7.4 se resumen los resultados obtenidos de este experimento

**Cuadro 7.4** – Comparación del estadístico COB ( $COB_{dist}$ ) para obs. regulares.

Series	MALtDB Vs.	MALtDB Vs.	MNoLIN Vs.
	MLIN	MNoLIN	MLIN
1	MLIN	MNoLIN	MNoLIN*
2	MLIN	MNoLIN*	MLIN
3	MLIN	MNoLIN	MLIN
4	MLIN	MNoLIN	MLIN*
5	MLIN	MNoLIN	MLIN*
6	MLIN*	MALtDB*	MLIN*

Los resultados del cuadro 7.4 indican que de la comparación MALtDB Vs. MLIN, el modelo con mejores resultados para las series 1, 2, 3, 4 y 5 fue el MLIN. En la serie 6, aparece el resultado MLIN\*, indicando que, aunque los resultados hayan sido favorables para el MLIN, no hubo suficiente evidencia para rechazar  $H_0$ . En cuanto a la comparación MALtDB Vs MNoLIN, los resultados 7.4 indican que para las series 1, 3, 4 y 5, el MNoLIN tuvo un mejor desempeño que el MALtDB mientras que en las series 2 y 6, los resultados no fueron concluyentes. La última de estas comparaciones, MNoLIN Vs. MLIN, indica que

el desempeño del MLIN fue superior que el del MNoLIN para las series 2 y 3, mientras que para las series 1, 4, 5 y 6, los resultados no fueron concluyentes.

En general, los resultados de esta prueba indican que los modelos propuestos, MLIN y MNoLIN, son mejores en el aspecto de la cobertura observada, dada por el estadístico  $COB_{dist}$ . Esto se debe a que el modelo apto, el MAltDB, lo hizo mejor que los modelos propuestos en tan sólo 1 de 12 comparaciones y la mejora ni siquiera fue significativa. En especial, el MLIN es el modelo que mejor desempeño tuvo en este aspecto del comparativo pues fue mejor en 11 de 12 comparaciones (aunque la diferencia fue significativa en sólo 7 de esas 11 ocasiones). El MNoLIN tuvo también un buen desempeño pues fue mejor en 6 ocasiones, aunque la mejora fue significativa en sólo 4 de estos 6 comparativos.

Estos fueron los resultados para el estadístico COB ( $COB_{dist}$ ). A continuación, veremos las pruebas y los resultados obtenidos con el estadístico Est.

### **7.2.2. Comparación de la estrechez de los intervalos predictivos con base en el estadístico Est**

Siempre es deseable que el intervalo predictivo sea tan reducido como sea posible mientras cumpla con la cobertura teórica especificada por el usuario. Por ejemplo, si el grosor del intervalo predictivo tiende a cero, esto significaría que nuestro modelo de pronóstico es muy preciso y que de hecho, está cada vez más cerca de pronosticar el dato deseado con una exactitud cada vez más grande. En esta sección, compararemos la estrechez de los intervalos predictivos de cada modelo para cada serie. Debido a que es deseable que los intervalos predictivos sean tan reducidos como se pueda, utilizaremos de nuevo la prueba de rangos con signo de Wilcoxon. Se supondrá que la media del estadístico del modelo que menores resultados arroje del estadístico Est, a lo largo de los doce valores de  $h$ , es menor que la media del modelo con el que se compara. De esta manera, se probará la hipótesis nula  $H_0: \mu_1 = \mu_2$  contra la alternativa  $H_1: \mu_1 < \mu_2$ . Las pruebas serán realizadas

al nivel de significancia  $\alpha = 0,05$ . Los resultados se muestran en el cuadro 7.5.

**Cuadro 7.5** – Comparación del estadístico Est para obs. regulares.

Serie	MALtDB Vs. MLIN	MALtDB Vs. MNoLIN	MNoLIN Vs. MLIN
1	MALtDB	MALtDB	MNoLIN
2	MALtDB	MALtDB	MNoLIN
3	MALtDB	MALtDB	MNoLIN
4	MALtDB	MALtDB	MNoLIN
5	MALtDB	MALtDB	MNoLIN
6	MALtDB	MNoLIN	MNoLIN

Los resultados del cuadro 7.5 indican que de la comparación MALtDB Vs. MLIN, los resultados fueron favorables para el MALtDB en cada una de las seis series utilizadas en el comparativo. Estos resultados fueron significativos en cada prueba. En cuanto a la comparación MALtDB Vs. MNoLIN, los resultados del cuadro 7.5 indican que el MALtDB mejora el desempeño del MNoLIN para las series 1, 2, 3, 4 y 5. La única serie cuyos intervalos predictivos resultaron más reducidos con el MNoLIN fue la serie 6. Por último, los resultados del comparativo MNoLIN Vs. MLIN indican que el MNoLIN es mejor que el MLIN para las seis series.

En general, los resultados del cuadro 7.5 indican que en cuanto a la estrechez de los intervalos predictivos, medida con el estadístico Est, el mejor modelo es el MALtDB. Este modelo fue mejor en 11 de las 12 veces que se comparó con los modelos MLIN y MNoLIN y las diferencias fueron significativas. El MNoLIN tuvo un buen desempeño ya que fue mejor en 7 de 12 comparativos y las diferencias también fueron significativas. Por último, el modelo menos exitoso en este aspecto del comparativo fue el MLIN ya que nunca superó el desempeño de los otros modelos.

Los resultados del estadístico Est contrastan bastante (salvo en el caso del MNoLIN) con los resultados obtenidos al comparar el desempeño de los modelos bajo el estadístico  $CoB_{dist}$ . Mientras el MLIN fue el mejor modelo en cuanto a los resultados del estadístico  $CoB_{dist}$  y el MALtDB fue el peor modelo para este estadístico, los resultados fueron

totalmente opuestos al comparar estos dos modelos utilizando el estadístico  $Est$ . En el caso del  $MNoLIN$ , este fue en ambos comparativos un buen modelo. Este fenómeno se debe a que el  $MALtDB$  utiliza muchos parámetros y tiende a sobreparametrizar las series. Las consecuencias de la sobreparametrización son una estimación mucho menor de la varianza del error de pronóstico. Debido a que el  $MALtDB$  subestima la varianza del error de pronóstico, produce intervalos predictivos muy estrechos. Por esta razón, los intervalos predictivos del  $MALtDB$  tienden a ser muy poco efectivos, de acuerdo con el estadístico  $CoB_{dist}$ . Sin embargo, al ser tan estrechos, son menores que los intervalos predictivos obtenidos con los modelos  $MLIN$  y  $MNoLIN$ . Lo opuesto a esta situación ocurre con el  $MLIN$ . Este modelo, sobrestima la varianza del error de pronóstico debido a que su capacidad de modelaje es muy reducida en comparación a su contraparte no lineal (el  $MNoLIN$ ). Debido a que el  $MLIN$  sobrestima el error de pronóstico, sus intervalos predictivos tenderán a ser muy grandes y las observaciones pronosticadas tenderán a ocurrir dentro de ellos (alta efectividad de acuerdo al estadístico  $CoB_{dist}$ ). Sin embargo, al ser los intervalos predictivos del  $MLIN$  tan grandes, serán ineficientes de acuerdo con el estadístico  $Est$  y quedarán superados en este aspecto por los intervalos del  $MALtDB$ . El único modelo que no comete sobreparametrización al mismo tiempo que modela bien es el  $MNoLIN$ .

En la siguiente sección, compararemos el desempeño de los tres modelos en cuanto a pronóstico se refiere utilizando el estadístico  $DESv \%$ .

### **7.2.3. Comparación del error de pronóstico basada en el estadístico**

#### **$DESv \%$**

Para comparar el desempeño de los modelos al pronosticar las seis series de ejemplo (cuadro 6.1), decidimos estudiar el comportamiento del estadístico  $DESv \%$  entre cada uno de los modelos para cada una de las series en los doce pasos a pronosticar. Esto se debe a que el estadístico  $DESv \%$  es una medida del error de pronóstico en la que no intervienen

las unidades de la serie pronosticada y por tanto, aquel modelo que lo minimice estará mejorando el pronóstico. Por esta razón, para comparar el desempeño de los modelos con las diferentes series, utilizamos, de nuevo, la prueba de rangos con signo de Wilcoxon. De manera similar a lo hecho para el estadístico Est, supusimos que el modelo que tenía los menores valores del estadístico DESV % para los doce valores de  $h$ , tendría una media menor que la del modelo con el que se compara. De esta forma, la hipótesis nula fue  $H_0: \mu_1 = \mu_2$ , mientras que la hipótesis alternativa fue  $H_1: \mu_1 < \mu_2$ . Las pruebas se realizaron al nivel de significancia  $\alpha = 0,05$ . Los resultados aparecen en el cuadro 7.6.

**Cuadro 7.6** – Comparación del estadístico DESV % para obs. regulares.

Series	MALtDB Vs.	MALtDB Vs.	MNoLIN Vs.
	MLIN	MNoLIN	MLIN
1	MALtDB	MNoLIN	MNoLIN
2	MLIN	MNoLIN	MLIN
3	MLIN	MNoLIN	MNoLIN
4	MALtDB	MNoLIN	MNoLIN
5	MALtDB	MNoLIN	MNoLIN
6	MALtDB	MNoLIN	MNoLIN

En el cuadro 7.6 se observa que de la comparación MALtDB Vs. MLIN, los resultados fueron favorables para el MALtDB para las series 1, 4, 5 y 6, mientras que el MLIN fue mejor para las series 2 y 3. En cuanto a los resultados del comparativo MALtDB Vs. MNoLIN, los resultados fueron favorables para el MNoLIN en cada una de las seis series comparadas. Por último, en el comparativo MNoLIN Vs. MLIN, tenemos que el MNoLIN lo hizo mejor para las series 1, 3, 4, 5 y 6. La única serie que fue mejor pronosticada en este comparativo por el MLIN fue la serie 2. Todos los resultados anteriores fueron significativos.

Al comparar el desempeño de los modelos propuestos, el MLIN y el MNoLIN, contra la alternativa apta de la literatura, el MALtDB, los resultados indican que el MNoLIN es el mejor modelo en cuanto a pronóstico se refiere (de acuerdo con el estadístico DESV %), ya que fue el mejor modelo en 11 de 12 comparaciones. El MALtDB fue mejor en 4 mientras que el MLIN fue el mejor modelo en 3 ocasiones. Estos resultados corroboran la idea de que

el MNoLIN modela mejor y sobreparametriza menos. Gracias a este equilibrio, el MNoLIN pronosticó mejor en este comparativo.

### **7.3. Conclusiones sobre el pronóstico de series con observaciones regulares**

Se pronosticaron las seis series del capítulo 6 con los modelos propuestos, el MLIN y el MNoLIN y también, con la alternativa apta de la literatura, el MALTDB. Se propusieron los estadísticos COB, EST, DESV y DESV % para conocer el comportamiento de los tres modelos en diferentes aspectos como lo son la cobertura observada del intervalo predictivo (COB), la estrechez de los intervalos predictivos (EST) y la magnitud media (DESV) y porcentual (DESV %) de los errores de pronóstico. Los estadísticos fueron utilizados para comparar el desempeño de los modelos en cada uno de estos aspectos.

Para saber si las diferencias entre los modelos son significativas, utilizamos la prueba de rangos con signo de Wilcoxon, una alternativa no paramétrica robusta que no supone normalidad en las observaciones. Las pruebas se realizaron al nivel de significancia  $\alpha = 0,05$ .

Para comparar la cobertura de los intervalos predictivos, utilizamos el estadístico COB. De acuerdo con los resultados, el modelo que mejor lo hizo con respecto a la cobertura observada de los intervalos predictivos fue el MLIN, seguido por el MNoLIN y por último, el MALTDB.

En cuanto a la estrechez de los intervalos predictivos, se utilizó el estadístico EST y los resultados indicaron que el mejor modelo fue el MALTDB, seguido por el MNoLIN y por último el MLIN.

En cuanto al pronóstico, se utilizó el estadístico DESV % para comparar el desempeño de los modelos en este aspecto ya que no toma en cuenta la unidades de las series observadas

sino sólo su error porcentual. Los resultados favorecieron rotundamente al MNoLIN muy por encima del MAltDB y del MLIN.

Los resultados anteriores son validos para las seis series con observaciones regulares. En el siguiente capítulo, veremos lo que ocurre cuando pronosticamos series con observaciones escasas.

---

---

## CAPÍTULO 8

---

### Segundo escenario: series con observaciones escasas

**E**n este segundo experimento, compararemos la eficacia de cada uno de los métodos al pronosticar series con pocas observaciones. Para saber si una serie es estacional, es necesario observar si su comportamiento en una estación se repite en el período siguiente o en el período anterior. Por ejemplo, si tenemos un período de observaciones y sospechamos que hay estacionalidad en ellos, es necesario otro período de observaciones para corroborar nuestra sospecha. Esto implica la necesidad de tener al menos  $2m$  observaciones para saber o confirmar si una serie es estacional. Por supuesto, si un experto en el fenómeno que se estudia opina que la serie es estacional aunque haya menos de  $2m$  observaciones, igual se puede asegurar que la serie es estacional.

De cualquier forma, los modelos propuestos en este trabajo no necesitan que haya  $2m$  observaciones para pronosticar una serie debido al bajo número de parámetros que manejan. Según vimos en el capítulo 5, el  $MLIN$  necesita de  $m + 1$  datos para dar un pronóstico y  $m + 2$  observaciones para estimar los intervalos predictivos mientras que el  $MNoLIN$  necesita de  $m + 2$  datos para dar un pronóstico y  $m + 3$  observaciones para estimar sus intervalos predictivos. Esto no ocurre con el  $MAI_tDB$  ya que este modelo necesita de  $2m$  observaciones para dar un pronóstico y  $2m + 1$  observaciones para obtener sus intervalos predictivos.

En este capítulo, se compararán cada uno de los modelos al pronosticar series con escasa información (de  $2m$  a  $4m$  observaciones). Primero, veremos los lineamientos observados durante el comparativo y luego discutiremos los resultados obtenidos al comparar los tres modelos en distintos aspectos de interés, como lo son el pronóstico, la cobertura y la estrechez de los intervalos predictivos.

## 8.1. Lineamientos observados

Para conocer la eficacia de los modelos en diferentes pasos hacia adelante de pronóstico y con poca información, tomamos las series del capítulo 6 (cuadro 6.1) y las redujimos en tamaño. Esto es, tomamos los primeros  $2m + j$  datos de cada serie con  $j = 1, \dots, 2m$  y pronosticamos las siguientes  $h = 1, \dots, (m = 12)$  observaciones. Es decir, pronosticamos un total de 24 veces las siguientes 12 observaciones añadiendo cada vez al conjunto de entrenamiento la siguiente observación. El haber tomado un número mínimo de observaciones a pronosticar de  $2m + 1$  se debe a que éste es el número mínimo de observaciones necesarias para que el MALTDB arroje intervalos predictivos. Para analizar los resultados, tomamos los mismos estadísticos que aquellos tomados cuando se pronosticaron las series con información regular de la sección 7.1 (página 87). Los estadísticos son: Cob (cobertura observada del intervalo predictivo), Est (estrechez media del intervalo predictivo), Desv (desviación absoluta media) y Desv % (desviación absoluta porcentual).

Uno de los problemas al pronosticar series con poca información radica en lo difícil que resulta caracterizar las autocorrelaciones de los residuales. En ningún caso, las autocorrelaciones de los residuales fueron significativas. Aunque esto no sea así, pues sabemos que los resultados cambian cuando hay más observaciones, simplemente supusimos que las desviaciones seguían un proceso NID( $0, \sigma^2$ ). En la práctica no habrá muchas opciones.

## 8.2. Resultados

Los resultados de este experimento se muestran en el apéndice D. En cada una de las columnas de los cuadros del apéndice D, se muestran los estadísticos de interés para los doce pasos pronosticados y que representan los meses de un año. En el cuadro 8.1 se muestra uno de los cuadros de resultados.

Los resultados del cuadro 8.1 se pueden contrastar con aquellos obtenidos al pronosticar la misma serie, con el mismo modelo, pero con más observaciones (Esc. 1), lo cual implica además, la remoción del ruido autorregresivo de las desviaciones. Este resultado se encuentra en el cuadro 7.1 (página 89) y se encuentra repetido en el cuadro 8.2 para facilitar la comparación.

**Cuadro 8.1** – Esc. 2; serie 1; MLIN; NID(0,  $\sigma^2$ ).

<i>h</i>	COB	EST	DESV	DESV %
1	0,958	358,907	68,787	8,679
2	0,958	353,025	70,940	8,900
3	0,958	347,603	75,506	9,412
4	0,958	342,605	76,991	9,507
5	0,958	337,993	79,518	9,791
6	0,958	333,730	83,285	10,228
7	0,958	329,778	81,723	9,973
8	0,958	326,104	80,685	9,737
9	0,958	322,680	77,233	9,204
10	0,958	319,477	73,561	8,728
11	0,958	316,475	71,896	8,516
12	0,958	313,651	69,251	8,202

**Cuadro 8.2** – Esc. 1; Series 1; MLIN; AR(1).

<i>h</i>	COB	EST	DESV	DESV %
1	0,750	163,235	52,468	5,898
2	0,833	171,842	52,764	5,821
3	0,791	172,797	52,038	5,673
4	0,833	172,910	50,133	5,441
5	0,875	172,924	48,551	5,206
6	0,916	172,925	45,559	4,819
7	0,875	172,925	46,098	4,817
8	0,875	172,926	46,412	4,810
9	0,875	172,926	45,304	4,643
10	0,875	172,926	45,980	4,708
11	0,875	172,926	45,041	4,576
12	0,875	172,926	46,865	4,745

En la descripción de los cuadros del apéndice D (y por lo tanto, en el cuadro 8.1) se presenta el escenario (Esc. 2: la información es escasa), la serie pronosticada (serie 1 para nuestro ejemplo), el modelo con el que se pronosticó la serie (el MLIN) y la suposición sobre la fuente de variación, la cual fue siempre NID(0,  $\sigma^2$ ), pues esta es la única suposición posible con tan poca información.

La primer diferencia que se puede observar es que hay un gran cambio en los resultados de la columna correspondiente al estadístico COB. Los valores del estadístico COB (la

cobertura observada) estuvieron más cerca de su valor teórico de cobertura (0,95) cuando hubo poca información. Este fenómeno pareció repetirse en el resto de las series. Sin embargo, el resto de los estadísticos se comportó mejor conforme había más observaciones (Esc. 1); la amplitud de los intervalos fue más reducida (Est), el promedio de las desviaciones fue menor (Desv) y por tanto, la desviación porcentual fue menor (Desv %).

La razón por la cual existe una notable mejoría en el estadístico Cob de las series con poca información es que para este tipo de series, se supuso que las desviaciones seguían un proceso  $NID(0, \sigma^2)$  en lugar de un proceso  $AR(p)$ , ya que cuando hay poca información, no se pueden caracterizar las desviaciones de una serie como ruido  $AR(p)$  y la alternativa es suponer normalidad. El análisis realizado para deducir los intervalos predictivos bajo la suposición de ruido  $NID(0, \sigma^2)$ , toma en cuenta la varianza de los estimadores ya que supone que éstos son variables aleatorias. Lo contrario ocurre al deducir los intervalos predictivos después de caracterizar las desviaciones como ruido  $AR(p)$ . La deducción de los intervalos predictivos bajo el supuesto de ruido  $AR(p)$  supone información perfecta (los estimadores son los parámetros). Esta suposición trae como consecuencia, intervalos predictivos muy reducidos y cuya cobertura observada dista bastantes de la teórica. Esto no ocurre cuando se supone que las desviaciones siguen un proceso  $NID(0, \sigma^2)$ .

A continuación, se comparara el desempeño del  $MLIN$ , el  $MNoLIN$  y el  $MAltDB$  al pronosticar series con escasa información. Los comparativos son similares a aquellos realizados cuando se compararon los modelos cuando había información regular (capítulo 7). Se presentarán resultados relacionados con el pronóstico, la cobertura y la estrechez de los intervalos predictivos.

### 8.2.1. Comparación de la cobertura de los intervalos predictivos basada en el estadístico $COB$

La comparación de los intervalos predictivos cuando la información es escasa, se realiza de manera similar a aquella de la sección 7.2.1 del capítulo 7 (página 92), solo que esta vez, los datos disponibles son muy pocos. Como en la subsección 7.2.1, utilizamos el estadístico  $COB_{dist}$ , que representa la distancia para cada uno de los doce valores de  $h$  entre la cobertura observada de los intervalos predictivos y su cobertura teórica. También, utilizamos la prueba de rangos con signo de Wilcoxon, para saber si la media de los valores del estadístico  $COB_{dist}$  de un modelo era menor que los de otro. Los resultados se presentan en el cuadro 8.3.

**Cuadro 8.3** – Comparación del estadístico  $COB$  ( $COB_{dist}$ ) para obs. escasas.

Series	MALtDB Vs.	MALtDB Vs.	MNoLIN Vs.
	MLIN	MNoLIN	MLIN
1	MLIN	MALtDB*	MLIN
2	MALtDB	MALtDB	MLIN
3	MLIN	MALtDB	MLIN
4	–	MNoLIN	MNoLIN
5	MLIN	MALtDB	MLIN
6	MLIN	MALtDB*	MLIN

Los resultados del cuadro 8.3 indican que de la comparación MALtDB Vs. MLIN, el MLIN lo hizo mejor para las series 1, 3, 5 y 6, mientras que el MALtDB lo hizo mejor sólo para la serie 2. El resultado de la serie 4, –, indica que la evidencia no fue suficiente para realizar la prueba. Esto se debió a que los doce valores del estadístico  $COB_{dist}$  obtenidos con los modelos MALtDB y MLIN para la serie 4 coincidieron, es decir, los doce valores fueron iguales. En cuanto a la comparación MALtDB Vs. MNoLIN, los resultados del cuadro 8.3 indican que el MALtDB lo hizo mejor para las series 1, 2, 3, 5 y 6, aunque los resultados fueron significativos sólo para las series 2, 3 y 5. En este caso, el MNoLIN lo hizo mejor para la serie 4. Por último, los resultados del cuadro 8.3 indican que de la comparación

MNoLIN Vs. MLIN, el MLIN lo hizo mejor para las series 1, 2, 3, 5 y 6, mientras que el MNoLIN lo hizo mejor sólo para la serie 4.

Los resultados anteriores favorecen al MLIN ya que este modelo lo hizo mejor en 9 de 12 comparaciones, seguido por el MALTDB, que lo hizo mejor en 6 de 12 comparaciones. Por último, el MNoLIN lo hizo mejor en tan sólo 2 de 12 casos. Estos resultados difieren un poco de aquellos obtenidos en la subsección 7.2.1. La diferencia más notable es que el MNoLIN paso de ser el segundo mejor modelo en este aspecto de la comparación, a ser el tercero. Esta diferencia se debe a que los intervalos predictivos fueron estimados bajo la suposición de desviaciones distribuidas normalmente. Debido a que esta suposición si permite tomar en cuenta la varianza de los estimadores, los intervalos del MALTDB crecieron mucho ya que este modelo utiliza 24 estimadores contra 14 del MNoLIN.

Esta fue la comparación de los modelos respecto al estadístico  $COB$  ( $COB_{dist}$ ). A continuación, veremos los resultados obtenidos al comparar los modelos utilizando el estadístico  $Est$ .

### **8.2.2. Comparación de la estrechez de los intervalos predictivos basada en el estadístico $Est$**

La comparación de la estrechez de los intervalos predictivos cuando hay poca información se realiza de manera similar a lo hecho en la subsección 7.2.2. Utilizamos el estadístico  $Est$  para esta comparación. En este caso, la prueba de rangos con signo de Wilcoxon nos dice si la media del estadístico  $Est$  de un modelo era significativamente menor a la del otro. Los resultados de este comparativo se muestran en el cuadro 8.4.

Los resultados del cuadro 8.4 indican que de la comparación MALTDB Vs. MLIN, el MALTDB lo hizo mejor para las series 3, 4, 5 y 6, mientras que el MLIN lo hizo mejor para las series 1 y 2. En cuanto a la comparación MALTDB Vs. MNoLIN, los resultados de la prueba indican que el MNoLIN lo hace mejor para las series 1, 2, 3, 4, 5 y 6, es decir, lo

**Cuadro 8.4** – Comparación del estadístico  $Est$  para obs. escasas.

Series	MALtDB Vs. MLIN	MALtDB Vs. MNoLIN	MNoLIN Vs. MLIN
1	MLIN	MNoLIN	MNoLIN
2	MLIN	MNoLIN	MNoLIN
3	MALtDB	MNoLIN	MNoLIN
4	MALtDB	MNoLIN	MNoLIN
5	MALtDB	MNoLIN	MNoLIN
6	MALtDB	MNoLIN	MNoLIN

hizo mejor siempre. Lo mismo ocurrió en la comparación MNoLIN Vs. MLIN.

El MNoLIN lo hizo mejor en 12 de 12 comparaciones, mientras que el MALtDB lo hizo mejor en 4 de las 12 comparaciones. El modelo menos favorecido en este aspecto de la prueba fue el MLIN pues lo hizo bien en sólo 2 ocasiones. Este resultado contrasta con aquel obtenido cuando el número de observaciones era regular (subsección 7.2.2). En aquella ocasión, el modelo que mejor lo hizo fue el MALtDB. La diferencia es que en esta ocasión, los intervalos predictivos fueron deducidos bajo la suposición de desviaciones que siguen un proceso  $NID(0, \sigma^2)$ , lo cual permitió incluir la varianza de los estimadores al deducir los intervalos predictivos. Debido a que el MALtDB utiliza 24 estimadores, sus intervalos predictivos terminaron siendo muy amplios y por esta razón, fueron superados en estrechez por los intervalos del MNoLIN, pues este último modelo utiliza solamente 14 estimadores.

Estos fueron los resultados en cuanto a estrechez de intervalos predictivos. A continuación, veremos los resultados obtenidos al comparar el desempeño en cuanto a pronóstico de los tres modelos utilizando el estadístico  $DESv \%$ .

### 8.2.3. Comparación del error de pronóstico basada en el estadístico

#### DESV %

Para comparar los resultados del pronóstico, procedimos de manera similar a la subsección 7.2.3 (página 96). Comparamos los estadísticos DESV % obtenidos con cada modelo. Utilizamos la prueba de rangos con signo de Wilcoxon para saber si la media del estadístico DESV % obtenida con un modelo era menor que la del otro. Los resultados de estas comparaciones se muestran en el cuadro 8.5.

**Cuadro 8.5** – Comparación del estadístico DESV % para obs. escasas.

Series	MALtDB Vs. MLIN	MALtDB Vs. MNoLIN	MNoLIN Vs. MLIN
1	MALtDB*	MNoLIN	MNoLIN
2	MALtDB*	MNoLIN	MNoLIN
3	MALtDB	MNoLIN	MNoLIN
4	MALtDB	MNoLIN	MNoLIN
5	MALtDB	MNoLIN	MNoLIN
6	MALtDB	MNoLIN	MNoLIN

Los resultados del cuadro 8.5 muestran que de la comparación MALtDB Vs. MLIN, el MALtDB lo hizo mejor para cada una de las seis series, aunque la diferencia no fue significativa para las series 1 y 2. En cuanto a la comparación MALtDB Vs. MNoLIN, los resultados fueron favorables para el MNoLIN en cada una de las seis series. Este resultado se repitió en la comparación MNoLIN Vs. MLIN, en la que el MNoLIN fue mejor en las seis comparaciones.

Estos resultados indican que el MNoLIN es el mejor modelo en cuanto a pronóstico se refiere pues lo hizo mejor en 12 de 12 comparaciones. Le siguió el MALtDB, que lo hizo mejor en 6 de 12 ocasiones. El MLIN nunca superó el desempeño de los otros dos modelos. Estos resultados son muy similares a los de la subsección 7.2.3, lo que indica que sin importar el número de datos, el MNoLIN supera el desempeño de los otros dos modelos.

### 8.3. Conclusiones sobre el pronóstico de series con observaciones escasas

Los resultados obtenidos al pronosticar series con información escasa se resumen a continuación:

- **La cobertura observada del MLIN fue la mejor.** Esto es similar a lo ocurrido cuando se pronosticaron las series con información regular. En la mayoría de los casos, la cobertura de los intervalos del MLIN fue significativamente mejor (más cercana a 95 %).
- **La estrechez de los intervalos predictivos del MNoLIN fue siempre menor.** La mejora en la estrechez de los intervalos predictivos del MNoLIN fue siempre significativa al compararla con la de los otros modelos.
- **El MNoLIN pronosticó mejor siempre.** De manera similar a lo ocurrido cuando hubo información regular, la media de los valores del error porcentual de pronóstico (DES<sub>V</sub> %) obtenida con el MNoLIN fue significativamente menor siempre. Esto confirma lo adecuado de este modelo para pronosticar series con cualquier cantidad de información.

De nuevo, los modelos propuestos tuvieron un mejor desempeño que el modelo apto de la literatura. A continuación, veremos como utilizar estos modelos para pronosticar series con información faltante, así como un experimento para comparar el desempeño de los modelos conforme el número de datos faltantes varía.

---

---

# CAPÍTULO 9

---

## Tercer escenario: series con observaciones faltantes

Comúnmente, se encuentran series temporales con observaciones faltantes. Las razones pueden ser muchas: el extravío de información, mediciones incorrectas que obligan a invalidar las observaciones tomadas, etc. Las metodologías de pronóstico habituales no se pueden aplicar en este tipo de series (al menos de manera directa) pues están hechas para pronosticar series temporales y las series con observaciones faltantes no cumplen con la definición de serie temporal:

Una serie temporal es una sucesión ordenada de variables igualmente espaciadas a lo largo del eje temporal.

La definición anterior fue tomada de [Makridakis et al. \(1998\)](#). Una serie con observaciones faltantes no es una sucesión de variables *igualmente espaciadas* y por tanto, no es una serie temporal en el sentido estricto de la definición.

En la práctica, sin embargo, las metodologías habituales para el pronóstico de series temporales pueden ser adaptadas para el pronóstico de series con observaciones faltantes. Por ejemplo, si tenemos una serie con 30 observaciones y a esta serie le falta la observación número 21, podemos pronosticar la observación número 21 a partir de las 20 observaciones anteriores y luego, podemos tomar el valor pronosticado de la observación número 21 como si fuera la observación misma. Este procedimiento puede tener serias desventajas.

Imaginemos por ejemplo, que la observación faltante es la número 2. Estimar el posible valor de la observación 2 a partir de la observación 1 podría muy probablemente conducir a un estimado muy malo. Además, no sería posible incluir efectos de estacionalidad en la observación 2 pues este tipo de efectos no se pueden deducir a partir de una única observación.

Sin embargo, lo anterior no sucede con los modelos propuestos (el MLIN y el MNoLIN) ni con el modelo apto de la literatura (MAltDB) pues son modelos capaces de pronosticar series con observaciones faltantes de manera directa, sin necesidad de estimar la observación faltante. Esto se debe al hecho de que los modelos mencionados son deterministas y tienen la capacidad de ser ajustados a cualquier conjunto de datos incluyendo las series con observaciones faltantes.

En este capítulo, veremos la forma de estimar los parámetros de los modelos MLIN, MNoLIN y MAltDB cuando hay datos faltantes. También, veremos un experimento en el que se estudia el comportamiento del pronóstico, conforme varía el número de observaciones faltantes, en cada una de las seis series de estudio.

## 9.1. Estimación cuando hay datos faltantes

Supongamos que tenemos una serie con observaciones faltantes de la forma  $\mathbf{y}_{falt} = [y_1, y_2, \dots, y_{j-1}, y_{j+1}, \dots, y_n]'$  donde falta la  $j$ -ésima observación. Notemos que  $\mathbf{y}_{falt}$  puede expresarse como una función de  $\mathbf{y}$  de la siguiente manera

$$\mathbf{y}_{falt} = \mathbf{I}_{falt} \mathbf{y} \quad (9.1)$$

donde  $\mathbf{I}_{falt}$  es una matriz unitaria *sin* el  $j$ -ésimo renglón. Así, suponiendo que el modelo subyacente a  $\mathbf{y}_{falt}$  es el MNoLIN, la serie puede ser representada de la siguiente manera

$$\mathbf{y}_{falt} = (\mathbf{X}_{falt} - \tau \mathbf{D}_{falt}) \boldsymbol{\beta} + \boldsymbol{\mu}_{falt} \quad (9.2)$$

donde  $\mathbf{X}_{falt} = \mathbf{I}_{falt} \mathbf{X}$ ,  $\mathbf{D}_{falt} = \mathbf{I}_{falt} \mathbf{D}$  y  $\boldsymbol{\mu}_{falt} = \mathbf{I}_{falt} \boldsymbol{\mu}$ . Para encontrar los estimadores MCNL, es necesario minimizar la cantidad  $\boldsymbol{\mu}'_{falt} \boldsymbol{\mu}_{falt}$ . Esto se puede hacer con el algoritmo de búsqueda lineal con sección dorada sin hacer mayores modificaciones en él. Si el modelo subyacente a la serie fuera el MLIN, los estimadores MCO de  $\boldsymbol{\beta}$  tendrían la forma  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'_{falt} \mathbf{X}_{falt})^{-1} \mathbf{X}'_{falt} \mathbf{y}_{falt}$ . Los estimadores del MALTDB pueden obtenerse de manera similar a los del MLIN pues ambos modelos son de naturaleza lineal.

Es de esperarse que los pronósticos empeoren si la serie temporal está incompleta y que el error de pronóstico empeore conforme aumenta el número de observaciones faltantes. Lo anterior podría tener algunas excepciones, por ejemplo, si retiramos de la serie observaciones aberrantes, el pronóstico en tal situación podría mejorar.

Nos interesa conocer el desempeño de los modelos conforme aumenta el número de datos faltantes en la serie. Esperamos que el modelo que mejor desempeño tuvo cuando no había datos faltantes, el MNoLIN, siga manteniendo esta superioridad aún cuando haya datos faltantes en la serie.

A continuación, se presentan los lineamientos y resultados de un experimento en el que se compara el desempeño de los diferentes modelos de pronóstico conforme varía el número de observaciones faltantes en las series. Los resultados incluyen el comparativo con los resultados obtenidos cuando no hubo observaciones faltantes (capítulo 7), que es equivalente a que hubieran 0 observaciones faltantes.

## 9.2. Lineamientos observados

Lo que se quiere es comparar el desempeño del  $MLIN$ , el  $MNoLIN$  y el  $MALtDB$  al variar el número de observaciones faltantes en cada serie, incluyendo el caso en que el número de observaciones faltantes es cero (la serie tiene todas las observaciones). Nuestra propuesta es un experimento en el que se toman las series del capítulo 6 que poseen información completa y obtener resultados del pronóstico cuando la información está completa (esto ya se hizo en el capítulo 7), cuando falta un dato, cuando faltan dos datos y así hasta que falten  $F$  observaciones en la serie. La forma de tomar observaciones de una serie debe ser tal que cada observación tenga la misma probabilidad de ser excluida de la serie. Además, la forma en que las observaciones faltantes se distribuyen a lo largo de la serie debe tomar en cuenta el modelo que más observaciones necesita para pronosticar; el  $MALtDB$ . Debido a que este modelo ajusta una línea recta a cada estación, es necesario que haya al menos dos observaciones de cada estación para que este modelo pueda pronosticar los siguientes  $m$  datos. Por esta razón, tomamos las seis series de ejemplo y tomamos un número variable de observaciones bajo la restricción de que las series con observaciones faltantes resultantes tuvieran al menos dos observaciones de cada estación. Los pasos que sigue el programa que toma las series completas y las devuelve con observaciones faltantes se muestra a continuación:

*Paso 1* Tomamos la serie y con  $n$  observaciones;

*Paso 2* A partir de la serie, se genera una matriz unitaria  $\mathbf{I}$  de  $n \times n$ ;

*Paso 3* Elegimos al azar un renglón de  $\mathbf{I}$ . El número del renglón se genera a partir de una distribución uniforme discreta de la forma  $U(1, n)$ ;

*Paso 4* Si el renglón elegido en el paso anterior ya había sido seleccionada o si equivale a dejar a una estación con menos de dos observaciones, volvemos al *Paso 3*;

*Paso 5* Eliminar de la matriz unitaria el renglón seleccionado;

*Paso 6* Ir al *Paso 3* hasta que el número de renglones faltantes en  $\mathbf{I}$  corresponda al número de observaciones faltantes deseado en la serie ( $F$ ). A la matriz resultante le llamamos  $\mathbf{I}_{falt}$ ;

*Paso 7* Devolver  $\mathbf{I}_{falt}$ .

Una vez que tenemos la matriz  $\mathbf{I}_{falt}$ , basta premultiplicarsela a las matrices de datos de los modelos y proceder de forma habitual al estimar los parámetros. Una vez hecho lo anterior, pronosticamos las siguientes 12 observaciones. Lo anterior se repitió 24 veces agregando cada vez la siguiente observación a las observaciones del conjunto de entrenamiento. Esto es similar al experimento en el que se pronosticaron series con información regular (página 87) salvo por las observaciones faltantes inducidas en este caso. Sea  $f$  la variable que representa el número de observaciones faltantes de una serie. Lo anterior se hizo para  $f = 0$  (capítulo 7) y luego para  $f = 1, \dots, F$  con  $F = 18$ .

La finalidad del experimento es comparar los resultados de pronóstico de cada modelo con cada serie al variar el número de observaciones faltantes. Lo que se debe comparar es el estadístico DESV % (la desviación absoluta porcentual). Esto fue lo que se hizo al comparar la eficacia del pronóstico cuando había información completa (subsección 7.2.3). Sin embargo, el número de estadísticos DESV % correspondientes a cada modelo con cada serie y con los 19 valores de  $f$  da un total de  $3 \times 6 \times 19 = 342$  columnas correspondientes al estadístico DESV %.

Para mantener los resultados manejables, decidimos obtener un valor a partir de cada columna DESV % que representara lo bien a lo mal que el modelo haya pronosticado cada serie en los 12 pasos pronosticados hacia adelante. Debido a que el valor ideal del estadístico DESV % es cero (pues esto implicaría un pronóstico perfecto sin desviaciones de ningún tipo), decidimos medir la distancia euclidiana del estadístico DESV % a cero. A esta

distancia de le conoce como norma euclidiana y está dada por la siguiente ecuación

$$\delta(\text{MDL}, \text{SR}, f) = (\text{DESV \%}'\text{DESV \%})^{1/2}, \quad (9.3)$$

donde la función  $\delta(\text{MDL}, \text{SR}, f)$  representa la norma euclidiana del vector  $\text{DESV \%}$  obtenido con el modelo  $\text{MDL}$  ( $\text{MDL} = \text{MLIN}, \text{MNoLIN}, \text{MAltDB}$ ), con la serie  $\text{SR}$  ( $\text{SR} = 1, \dots, 6$ ) y con un número de observaciones faltantes igual a  $f$  ( $f = 0, \dots, 18$ ).

### 9.2.1. Resultados

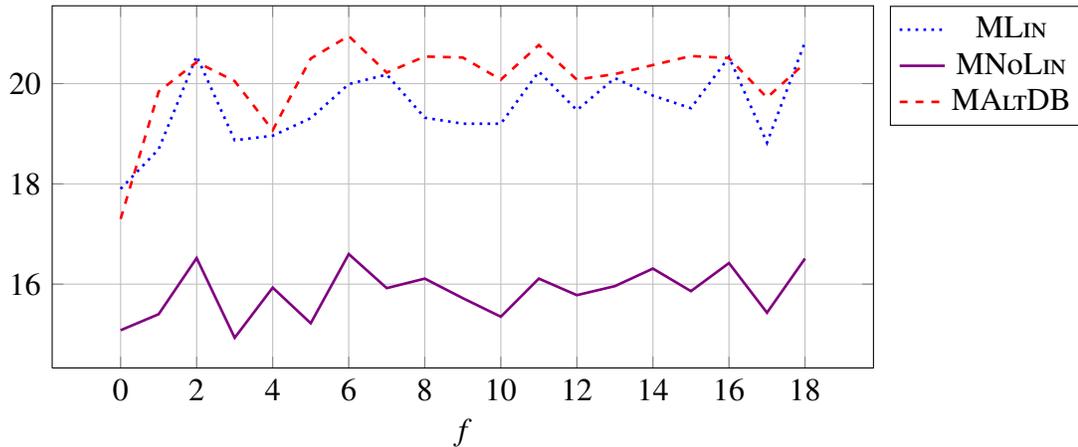
Los resultados de este experimento se muestran en el Apéndice E. En el cuadro 9.1 se muestra un ejemplo de estos cuadros.

**Cuadro 9.1** –  $\delta(\text{MDL}, \text{SR} = 1, f); \forall \text{MDL}; \forall f$ .

$f$	MLIN	MNoLIN	MAltDB
0	17,90	15,08	17,30
1	18,69	15,40	19,84
2	20,55	16,52	20,43
3	18,87	14,93	20,05
4	18,96	15,93	19,07
5	19,31	15,22	20,50
6	19,99	16,60	20,95
7	20,18	15,92	20,22
8	19,32	16,11	20,54
9	19,20	15,72	20,52
10	19,20	15,35	20,08
11	20,24	16,11	20,77
12	19,47	15,78	20,08
13	20,11	15,96	20,19
14	19,76	16,31	20,37
15	19,51	15,86	20,55
16	20,53	16,42	20,51
17	18,82	15,43	19,72
18	20,83	16,51	20,41

En cada cuadro del Apéndice E (y por tanto en el cuadro 9.1) aparece la norma euclidiana conforme se varía el número de observaciones faltantes y el modelo con el que se pronostican. Por ejemplo, en la descripción del cuadro 9.1 aparece  $\delta(\text{MDL}, \text{SR} = 1, f)$ ,

$\forall \text{MDL}, \forall f$ . Esto significa que se obtubieron los valores de la función  $\delta(\text{MDL}, \text{SR}, f)$  para los 18 valores de  $f$  con los modelos  $\text{MDL} = \text{MLIN}, \text{MNoLIN}, \text{MAltDB}$  utilizando la serie 1 ( $\text{SR} = 1$ ). Estos resultados se pueden graficar como se muestra en la figura 9.1.



**Figura 9.1** – Resultados de la función  $\delta(\text{MDL}, \text{SR} = 1, f), \forall \text{MDL}, \forall f$

En la figura 9.1 se pueden apreciar los distintos valores de la función  $\delta(\text{MDL}, \text{SR} = 1, f)$  al variar los valores de  $\text{MDL}$  y  $f$ . Se aprecia que a lo largo de los 19 valores de  $f$ , el modelo que minimizó el valor de la función fue  $\text{MDL} = \text{MNoLIN}$ . Por ejemplo, el  $\text{MNoLIN}$  fue el mejor modelo de pronóstico cuando no hubo observaciones faltantes ( $f = 0$ ). Este resultado coincide con el obtenido al realizar esta misma comparación (aunque con estadísticos diferentes) en la subsección 7.2.3. En general, los resultados del resto de los cuadros del apéndice E indican que el modelo que minimiza el valor de  $\delta(\text{MDL}, \text{SR}, f)$  es el  $\text{MNoLIN}$  salvo por la excepción de la serie 2, en la que el mejor modelo fue el  $\text{MLIN}$ .

En la figura 9.1 se puede apreciar que los valores de la función  $\delta(\text{MDL}, \text{SR} = 1, f)$  tienden a crecer conforme aumenta el número de observaciones faltantes. Esto es lo intuitivamente esperado pues la función es la norma euclidiana del promedio de las desviaciones absolutas a lo largo de los doce valores pronosticados hacia adelante. Así, se espera que este valor se deteriore (crezca) conforme aumenta el número de observaciones faltantes.

Aunque los resultados del cuadro 9.1 parecen contundentes, es necesario hacer pruebas de hipótesis para saber si las diferencias son significativas. Una vez más, decidimos

utilizar la prueba de rangos con signo de Wilcoxon para comparar cada par de modelos y concluir si las diferencias observadas eran significativas. La prueba consistió en comparar cada par de modelos para saber si la media de uno era significativamente menor a la del otro. De esta forma, la hipótesis nula es  $H_0: \mu_1 = \mu_2$  contra la alternativa  $H_1: \mu_1 < \mu_2$ , siendo  $\mu_1$  la media del modelo con menor media muestral y  $\mu_2$  la media del modelo contra el que se compara. Los resultados de estas comparaciones se muestran en el cuadro 9.2.

**Cuadro 9.2** – Comparación de los 18 valores de  $\delta(\text{MDL}, \text{SR}, f)$ , para cada par de modelos con cada serie.

Series	MALtDB Vs.	MALtDB Vs.	MNoLIN Vs.
	MLIN	MNoLIN	MLIN
1	MLIN	MNoLIN	MNoLIN
2	MLIN	MNoLIN	MLIN
3	MALtDB	MNoLIN	MNoLIN
4	MALtDB	MNoLIN	MNoLIN
5	MALtDB	MNoLIN	MNoLIN
6	MALtDB	MNoLIN	MNoLIN

Los resultados del cuadro 9.2 son muy similares a los obtenidos cuando se pronosticaron las series con información regular y con información escasa. En general, el mejor modelo de pronóstico lo siguió siendo aunque hubiera pocos o muchos datos faltantes. El MNoLIN siguió siendo el mejor modelo de pronóstico en 11 de 12 comparaciones. El MALtDB fue mejor en 4 ocasiones. Finalmente, el MLIN fue mejor en tres ocasiones.

### 9.3. Conclusiones

Los modelos propuestos, al igual que la alternativa apta de la literatura, son capaces de pronosticar series con observaciones faltantes de manera directa, sin necesidad de sustituir las observaciones faltantes por estimados. Los tres modelos fueron comparados en un experimento en el que se mide la norma euclidiana del estadístico DESV % conforme aumenta el número de observaciones faltantes. Se observó que el modelo que mejor desempeño tuvo conforme aumentaba el número de observaciones faltantes fue el MNoLIN

salvo por la serie 2, en la que el modelo con mejor desempeño fue el MLIN.

En el siguiente capítulo, veremos un resultados teórico notable; el MNoLIN, que ha sido el mejor modelo para pronosticar series con estacionalidad creciente, también es capaz de pronosticar series con estacionalidad constante.

---

---

# CAPÍTULO 10

---

## El modelo no lineal para el pronóstico de series con estacionalidad constante

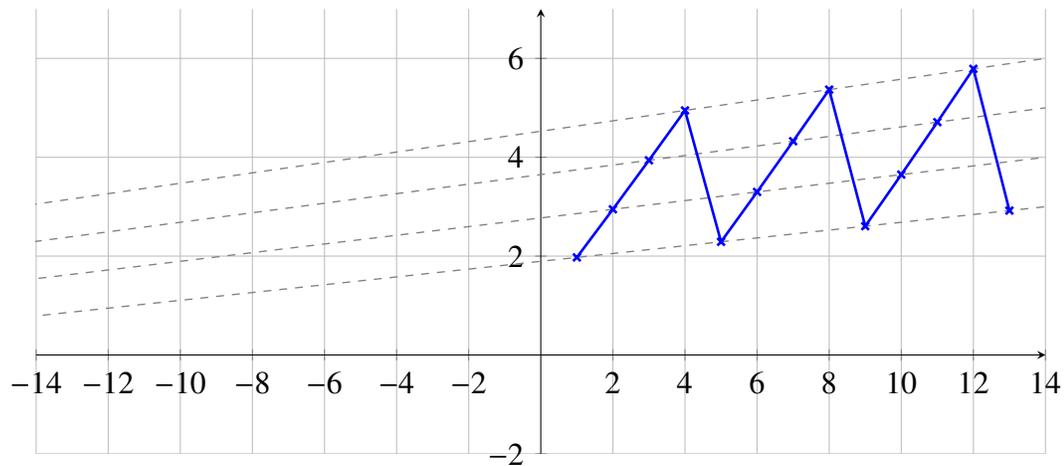
Matemáticamente, se demostrará que el MNoLIN tiende a un modelo para el pronóstico de series con estacionalidad constante conforme uno de sus parámetros (coordenada de intersección) tiende a infinito. Primero, veremos el comportamiento del MNoLIN conforme el parámetro tiende a infinito; luego, estudiaremos el comportamiento de los parámetros como una función de una de las coordenadas de intersección; también, se estudiará un modelo estándar para series temporales con tendencia lineal y estacionalidad constante; después de todo lo anterior, será posible demostrar matemáticamente que el MNoLIN tiende al modelo para el pronóstico de series con estacionalidad constante conforme una de las coordenadas de intersección tiende a infinito.

### 10.1. Comportamiento del modelo no lineal conforme el parámetro $\tau$ tiende en magnitud a infinito

El MNoLIN es capaz de pronosticar datos con tendencia lineal y estacionalidad creciente (como se demostró experimentalmente en los capítulos 7, 8 y 9) o constantes (lo que se demostrará matemáticamente en este capítulo). El MNoLIN puede pronosticar series con estacionalidad constante cuando el parámetro  $\tau$  tiende en magnitud a infinito. El

parámetro  $\tau$  es el desfase entre la serie y el eje de las ordenadas.

Recordemos que el MNoLIN ajusta una recta a cada conjunto de datos que representa una estación y que esas rectas se interceptan en un mismo punto del plano cartesiano. Pues bien, cuando  $\tau$ , que es una de las coordenadas del punto de intersección, tiende a estar «muy lejos» de la serie, las rectas que subyacen a cada estación tienden a comportarse como si fueran paralelas. En la figura 10.1 se ilustra esta situación.



**Figura 10.1** – El MNoLIN con intersección en  $(\tau = -100, \beta_0 = -6)$

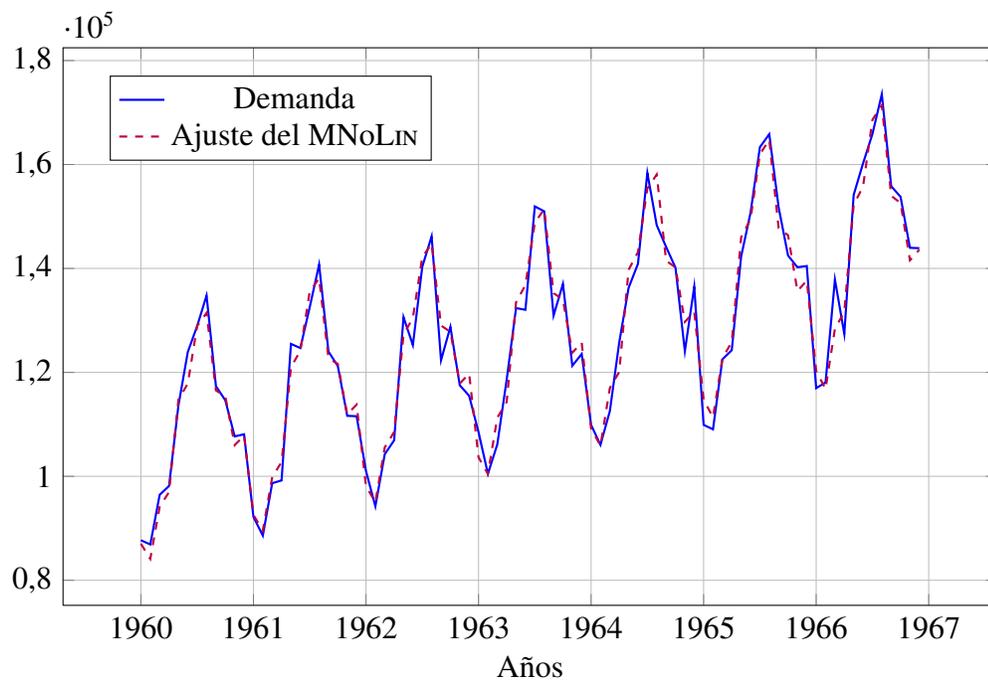
Si el punto de origen tiende a ser un número de gran magnitud entonces, como se ilustra en la figura 10.1, las rectas tenderán a ser paralelas. Para entender mejor este fenómeno, emplearemos una serie temporal con estacionalidad constante. En el cuadro 10.1 se muestra la demanda mensual de galones de gasolina de Ontario de enero de 1960 a diciembre de 1966. En la figura 10.2 se ilustra el comportamiento de los datos del cuadro 10.1 así como el ajuste del MNoLIN. La figura 10.2 muestra que la demanda de gasolina de Ontario exhibe patrones de tendencia lineal y estacionalidad constante. El ajuste del MNoLIN a los datos es bueno y el estimador  $\hat{\tau}$  tuvo un valor de  $-432,46$ . El hecho de que el MNoLIN se ajuste bien a la demanda de gasolina de Ontario es notable ya que el MNoLIN fue originalmente desarrollado para el pronóstico de datos con estacionalidad creciente. Para entender mejor lo que está ocurriendo, en la siguiente sección se explica lo que pasa con los parámetros del MNoLIN conforme la magnitud del estimador  $\hat{\tau}$  tiende a infinito.

**Cuadro 10.1** – Demanda de gasolina de Ontario de 1960 a 1966.  
(a) enero-junio.

Años	Enero	Feb.	Marzo	Abril	Mayo	Junio
1960	87695	86890	96442	98133	113615	123924
1961	92188	88591	98683	99207	125485	124677
1962	101007	94228	104255	106922	130621	125251
1963	108497	100482	106140	118581	132371	132042
1964	109894	106061	112539	125745	136251	140892
1965	109895	109044	122499	124264	142296	150693
1966	116963	118049	137869	127392	154166	160227

(b) julio-diciembre.

Años	Julio	Agosto	Sept.	Oct.	Nov.	Dic.
1960	128924	134775	117357	114626	107677	108087
1961	132543	140735	124008	121194	111634	111565
1962	140318	146174	122318	128770	117518	115492
1963	151938	150997	130931	137018	121271	123548
1964	158390	148314	144148	140138	124075	136485
1965	163331	165837	151731	142491	140229	140463
1966	165869	173522	155828	153771	143963	143898



**Figura 10.2** – Demanda de gasolina en Ontario y el ajuste del MNoLIN.

## 10.2. El comportamiento de los parámetros

Como ya ha sido mencionado, conforme  $|\hat{\tau}| \rightarrow \infty$ , las rectas que conforman al MNoLIN se comportan como si fuesen paralelas. Las pendientes de cada línea recta se comportan como si fueran una misma constante (estimadores  $\hat{\beta}_s$ ,  $s = 1, \dots, m$ ) y el desfase entre la serie y el eje de las ordenadas (estimador  $\hat{\beta}_0$ ) se comporta como una línea recta cuya variable independiente es  $\hat{\tau}$ . Para entender mejor estas afirmaciones, primero, estudiaremos el comportamiento de  $\hat{\beta}_0$  y la relación que éste guarda con  $\hat{\tau}$ . Después, haremos lo mismo con los estimadores  $\hat{\beta}_s \forall s$ . Demostraciones rigurosas serán empleadas en ambos casos. Observe el lector que el interés principal es el caso en el que  $|\hat{\tau}| \rightarrow \infty$ . Por ejemplo, no es necesario saber lo que ocurriría si  $|\hat{\tau}| \rightarrow 0$ , que es el caso en el que el MNoLIN tiende al MLIN. Por lo tanto, las demostraciones rigurosas serán realizadas tomando en cuenta únicamente el caso en el que  $\hat{\tau}$  tiende en magnitud a infinito, sin importar lo que se quiera demostrar.

### 10.2.1. Comportamiento lineal de $\hat{\beta}_0$ .

Para conocer el comportamiento de  $\hat{\beta}_0$  conforme varía  $\hat{\tau}$ , es necesario caracterizar al estimador  $\hat{\beta}_0$  como una función de  $\hat{\tau}$ . Recuérdese que los estimadores MCNL de  $\hat{\beta}_0$  y  $\hat{\beta}_s$  para  $s = 1, \dots, m$  son

$$\hat{\beta}_0 = \frac{1}{n} \sum_{s=1}^m \sum_{i=1}^{k_s} (y_{i,s} - \hat{\beta}_s(x_{i,s} - \hat{\tau})), \quad (10.1)$$

$$\hat{\beta}_s = \frac{\sum_{i=1}^{k_s} (y_{i,s} - \hat{\beta}_0)(x_{i,s} - \hat{\tau})}{\sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})^2}. \quad (10.2)$$

Sustituyendo los valores de la ecuación (10.2) en la igualdad (10.1) y despejando  $\hat{\beta}_0$  tenemos

$$\hat{\beta}_0 = \frac{\sum_{s=1}^m \sum_{i=1}^{k_s} y_{i,s} - \sum_{s=1}^m \frac{\sum_{i=1}^{k_s} y_{i,s} (x_{i,s} - \hat{\tau}) \sum_{j=1}^{k_s} (x_{j,s} - \hat{\tau})}{\sum_{l=1}^{k_s} (x_{l,s} - \hat{\tau})^2}}{n - \sum_{s=1}^m \frac{\sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau}) \sum_{j=1}^{k_s} (x_{j,s} - \hat{\tau})}{\sum_{l=1}^{k_s} (x_{l,s} - \hat{\tau})^2}}. \quad (10.3)$$

La ecuación (10.3) describe el comportamiento de  $\hat{\beta}_0$  como una función del estimador  $\hat{\tau}$ .

En la figura 10.3 se muestra el comportamiento de  $\hat{\beta}_0$  como una función de  $\hat{\tau}$ .

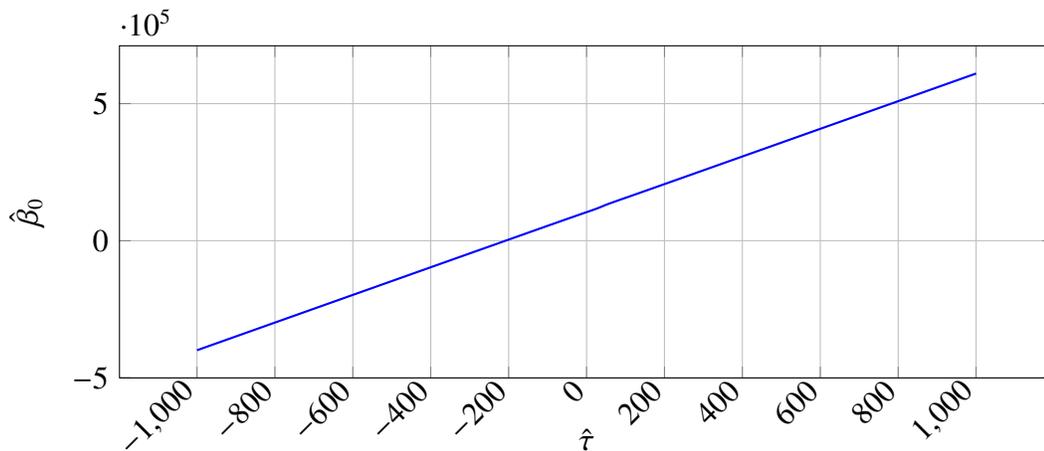


Figura 10.3 – Comportamiento de  $\hat{\beta}_0$  para los datos de la demanda de gasolina.

A pesar de la complejidad de la ecuación (10.3), la figura 10.3 revela que  $\hat{\beta}_0$  es aproximadamente una función lineal de  $\hat{\tau}$ . Esto se debe a que conforme  $|\hat{\tau}| \rightarrow \infty$ , el estimador  $\hat{\beta}_0$  tiende asintóticamente a una línea recta cuya variable independiente es  $\hat{\tau}$ . Esto se postula en la siguiente proposición.

**Proposición 6.** *Conforme  $|\hat{\tau}| \rightarrow \infty$  el estimador  $\hat{\beta}_0$  tiende asintóticamente a una recta cuya variable independiente es  $\hat{\tau}$ .*

*Demostración.* Del cuadro 3.1 (página 33) se deduce que  $n = \sum_{s=1}^m k_s$ . Esto significa que el número de observaciones es igual a la suma del número de datos contenidos en cada una de las  $m$  estaciones. Gracias a este hecho, la ecuación (10.3) puede ser reescrita de tal forma que la relación entre los estimadores  $\hat{\beta}_0$  y  $\hat{\tau}$  sea más «clara». Esto se logra de la siguiente manera:

$$\begin{aligned} \hat{\beta}_0 &= \frac{\sum_{s=1}^m \left[ \sum_{i=1}^{k_s} y_{i,s} - \frac{\sum_{i=1}^{k_s} y_{i,s} (x_{i,s} - \hat{\tau}) \sum_{j=1}^{k_s} (x_{j,s} - \hat{\tau})}{\sum_{l=1}^{k_s} (x_{l,s} - \hat{\tau})^2} \right]}{\sum_{s=1}^m \left[ k_s - \frac{\sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau}) \sum_{j=1}^{k_s} (x_{j,s} - \hat{\tau})}{\sum_{l=1}^{k_s} (x_{l,s} - \hat{\tau})^2} \right]} \\ &= \frac{\sum_{s=1}^m \left[ \frac{\sum_{i=1}^{k_s} y_{i,s} \sum_{l=1}^{k_s} (x_{l,s} - \hat{\tau})^2 - \sum_{i=1}^{k_s} y_{i,s} (x_{i,s} - \hat{\tau}) \sum_{j=1}^{k_s} (x_{j,s} - \hat{\tau})}{\sum_{l=1}^{k_s} (x_{l,s} - \hat{\tau})^2} \right]}{\sum_{s=1}^m \left[ \frac{k_s \sum_{l=1}^{k_s} (x_{l,s} - \hat{\tau})^2 - \sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau}) \sum_{j=1}^{k_s} (x_{j,s} - \hat{\tau})}{\sum_{l=1}^{k_s} (x_{l,s} - \hat{\tau})^2} \right]} \\ &= \frac{\sum_{s=1}^m \left[ \frac{\sum_{i=1}^{k_s} y_{i,s} \sum_{l=1}^{k_s} (x_{l,s}^2 - 2x_{l,s}\hat{\tau} + \hat{\tau}^2) - \sum_{i=1}^{k_s} y_{i,s} (x_{i,s} - \hat{\tau}) \sum_{j=1}^{k_s} (x_{j,s} - \hat{\tau})}{\sum_{l=1}^{k_s} (x_{l,s}^2 - 2x_{l,s}\hat{\tau} + \hat{\tau}^2)} \right]}{\sum_{s=1}^m \left[ \frac{k_s \sum_{l=1}^{k_s} (x_{l,s}^2 - 2x_{l,s}\hat{\tau} + \hat{\tau}^2) - \sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau}) \sum_{j=1}^{k_s} (x_{j,s} - \hat{\tau})}{\sum_{l=1}^{k_s} (x_{l,s}^2 - 2x_{l,s}\hat{\tau} + \hat{\tau}^2)} \right]} \end{aligned}$$

$$= \frac{\sum_{s=1}^m \left[ \frac{\sum_{i=1}^{k_s} y_{i,s} \left[ \sum_{l=1}^{k_s} x_{l,s}^2 - 2\hat{\tau} \sum_{l=1}^{k_s} x_{l,s} + k_s \hat{\tau}^2 \right] - \left[ \sum_{i=1}^{k_s} y_{i,s} x_{i,s} - \hat{\tau} \sum_{i=1}^{k_s} y_{i,s} \right] \left[ \sum_{j=1}^{k_s} x_{j,s} - k_s \hat{\tau} \right]}{\left[ \sum_{l=1}^{k_s} x_{l,s}^2 - 2\hat{\tau} \sum_{l=1}^{k_s} x_{l,s} + k_s \hat{\tau}^2 \right]} \right]}{\sum_{s=1}^m \left[ \frac{k_s \left[ \sum_{l=1}^{k_s} x_{l,s}^2 - 2\hat{\tau} \sum_{l=1}^{k_s} x_{l,s} + k_s \hat{\tau}^2 \right] - \left[ \sum_{i=1}^{k_s} x_{i,s} - k_s \hat{\tau} \right] \left[ \sum_{j=1}^{k_s} x_{j,s} - k_s \hat{\tau} \right]}{\left[ \sum_{l=1}^{k_s} x_{l,s}^2 - 2\hat{\tau} \sum_{l=1}^{k_s} x_{l,s} + k_s \hat{\tau}^2 \right]} \right]}.$$

El siguiente paso en la reescritura de  $\hat{\beta}_0$  consiste en multiplicar términos y agruparlos con respecto a  $\hat{\tau}$ . Para mantener estas operaciones «manejables», utilizaremos las constantes auxiliares  $f_s^I$ ,  $f_s^{II}$ ,  $g_s^I$ ,  $g_s^{II}$ ,  $g_s^{III}$  y  $h_s$  que se definen de la siguiente manera:

$$f_s^I = \left( \sum_{i=1}^{k_s} y_{i,s} \right) \left( \sum_{i=1}^{k_s} x_{i,s}^2 \right) - \left( \sum_{i=1}^{k_s} y_{i,s} x_{i,s} \right) \left( \sum_{i=1}^{k_s} x_{i,s} \right),$$

$$f_s^{II} = k_s \sum_{i=1}^{k_s} y_{i,s} x_{i,s} - \left( \sum_{i=1}^{k_s} y_{i,s} \right) \left( \sum_{i=1}^{k_s} x_{i,s} \right),$$

$$g_s^I = \sum_{i=1}^{k_s} x_{i,s}^2,$$

$$g_s^{II} = -2 \sum_{i=1}^{k_s} x_{i,s},$$

$$g_s^{III} = k_s,$$

$$h_s = k_s \sum_{i=1}^{k_s} x_{i,s}^2 - \left( \sum_{i=1}^{k_s} x_{i,s} \right)^2.$$

Gracias a las definiciones anteriores, es posible continuar con la reescritura de  $\hat{\beta}_0$  como se muestra a continuación

$$\hat{\beta}_0 = \frac{\sum_{s=1}^m \frac{f_s^I + f_s^{II} \hat{\tau}}{g_s^I + g_s^{II} \hat{\tau} + g_s^{III} \hat{\tau}^2}}{\sum_{s=1}^m \frac{h_s}{g_s^I + g_s^{II} \hat{\tau} + g_s^{III} \hat{\tau}^2}},$$

al multiplicar y dividir  $\hat{\beta}_0$  por  $\hat{\tau}^2$  tenemos

$$\hat{\beta}_0 = \frac{\sum_{s=1}^m \frac{f_s^I \hat{\tau}^2 + f_s^{II} \hat{\tau}^3}{g_s^I + g_s^{II} \hat{\tau} + g_s^{III} \hat{\tau}^2}}{\sum_{s=1}^m \frac{h_s \hat{\tau}^2}{g_s^I + g_s^{II} \hat{\tau} + g_s^{III} \hat{\tau}^2}}.$$

Aplicando las reglas de división entre polinomios y utilizando las constantes auxiliares  $l_s = f_s^{II} (g_s^{II})^2 - f_s^I g_s^{II} g_s^{III} - f_s^{II} g_s^I g_s^{III}$  para  $s = 1, \dots, m$ , podemos reescribir  $\hat{\beta}_0$  de la siguiente forma

$$\hat{\beta}_0 = \frac{\sum_{s=1}^m \frac{f_s^I g_s^{III} - f_s^{II} g_s^{II}}{(g_s^{III})^2} + \hat{\tau} \sum_{s=1}^m \frac{f_s^{II}}{g_s^{III}} + \sum_{s=1}^m \frac{l_s \hat{\tau} + f_s^{II} g_s^I g_s^{III} - f_s^I g_s^I g_s^{III}}{(g_s^{III})^2 (g_s^I + g_s^{II} \hat{\tau} + g_s^{III} \hat{\tau}^2)}}{\sum_{s=1}^m \frac{h_s}{g_s^{III}} - \sum_{s=1}^m \frac{h_s (g_s^{II} \hat{\tau} + g_s^I)}{(g_s^{III})(g_s^I + g_s^{II} \hat{\tau} + g_s^{III} \hat{\tau}^2)}}. \quad (10.4)$$

El tercer término del numerador en la ecuación (10.4) es una suma cuyo denominador es un grado mayor que su numerador. Por esta razón, este término se vuelve rápidamente despreciable para valores absolutos grandes de  $\hat{\tau}$ . Lo mismo aplica para el segundo término del denominador de la ecuación (10.4). De esta forma, para valores absolutos grandes de  $\hat{\tau}$ , la ecuación (10.4) tiende asintóticamente a la recta

$$\tilde{\beta}_0 = \frac{\sum_{s=1}^m \frac{f_s^I g_s^{III} - f_s^{II} g_s^{II}}{(g_s^{III})^2}}{\sum_{s=1}^m \frac{h_s}{g_s^{III}}} + \hat{\tau} \frac{\sum_{s=1}^m \frac{f_s^{II}}{g_s^{III}}}{\sum_{s=1}^m \frac{h_s}{g_s^{III}}}. \quad (10.5)$$

La relación (10.5) explica el comportamiento lineal de  $\hat{\beta}_0$  observado en la figura 10.3.  $\tilde{\beta}_0$  es la recta a la que tiende  $\hat{\beta}_0$  conforme  $|\hat{\tau}| \rightarrow \infty$ . Esto completa la demostración.  $\square$

Otra proposición importante que se relaciona con  $\hat{\beta}_0$  es la siguiente

**Proposición 7.** La razón  $\hat{\beta}_0/\hat{\tau}$  tiende a una asíntota vertical conforme  $|\hat{\tau}| \rightarrow \infty$ .

*Demostración.* Al dividir  $\hat{\beta}_0$  (ecuación (10.4)) entre  $\hat{\tau}$  tenemos

$$\frac{\hat{\beta}_0}{\hat{\tau}} = \frac{\frac{1}{\hat{\tau}} \sum_{s=1}^m \frac{f_s^I g_s^{III} - f_s^{II} g_s^{II}}{(g_s^{III})^2} + \frac{\hat{\tau}}{\hat{\tau}} \sum_{s=1}^m \frac{f_s^{II}}{g_s^{III}} + \sum_{s=1}^m \frac{l_s \hat{\tau} + f_s^{II} g_s^I g_s^{III} - f_s^I g_s^I g_s^{III}}{(g_s^{III})^2 (g_s^I + g_s^{II} \hat{\tau} + g_s^{III} \hat{\tau}^2)}}{\sum_{s=1}^m \frac{h_s}{g_s^{III}} - \sum_{s=1}^m \frac{h_s (g_s^{II} \hat{\tau} + g_s^I)}{(g_s^{III})(g_s^I + g_s^{II} \hat{\tau} + g_s^{III} \hat{\tau}^2)}}.$$

y conforme  $|\hat{\tau}| \rightarrow \infty$

$$\begin{aligned} \lim_{|\hat{\tau}| \rightarrow \infty} \frac{\hat{\beta}_0}{\hat{\tau}} &= \frac{0 + \sum_{s=1}^m \frac{f_s^{II}}{g_s^{III}} + 0}{\sum_{s=1}^m \frac{h_s}{g_s^{III}} - 0} \\ &= \frac{\sum_{s=1}^m \frac{f_s^{II}}{g_s^{III}}}{\sum_{s=1}^m \frac{h_s}{g_s^{III}}} \end{aligned}$$

Esto completa la demostración. □

### 10.2.2. Comportamiento constante de las $\hat{\beta}_s$ 's

Conforme  $\hat{\tau}$  tiende a un número muy grande, las líneas rectas subyacentes a cada estación se comportan como si fueran paralelas. Esto se refleja en los estimadores  $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m$ , los cuales tienden a una misma constante, como se muestra en la figura 10.4, donde se observa el comportamiento de cuatro de estos parámetros conforme  $\hat{\tau}$  variaba desde  $-1000$  hasta  $1000$ .

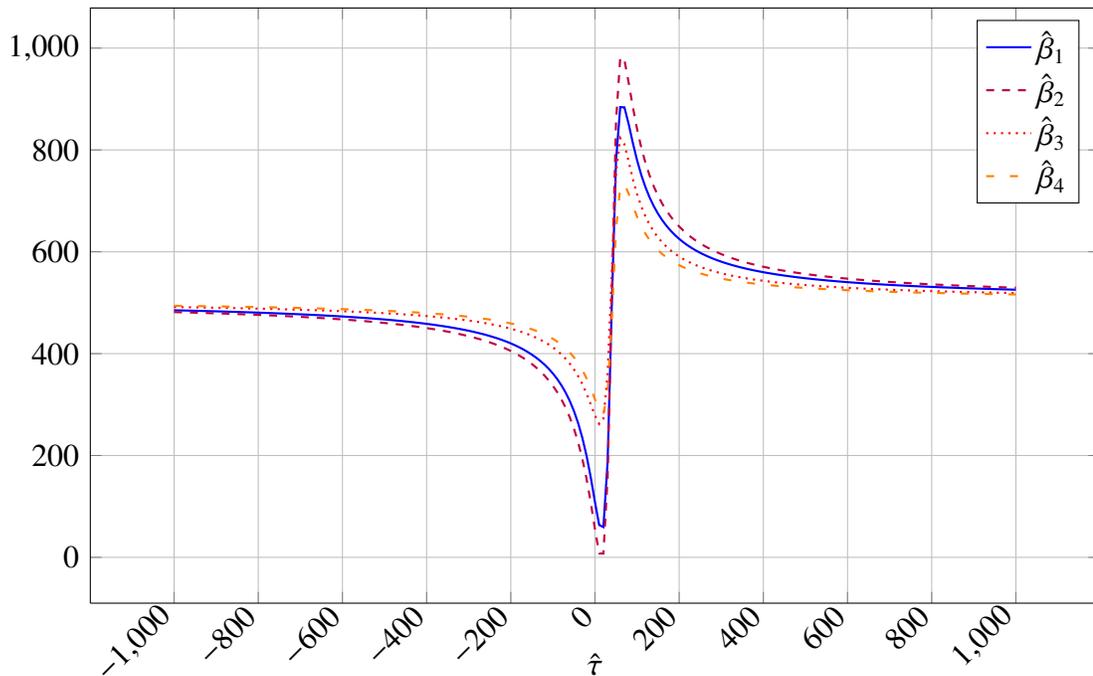


Figura 10.4 – Comportamiento de  $\hat{\beta}_1, \dots, \hat{\beta}_4$  para los datos del cuadro 10.1.

En la figura 10.4 se observa que los estimadores  $\hat{\beta}_1, \dots, \hat{\beta}_4$  tienden a una misma constante, un número cercano a 500. El comportamiento es similar para el resto de las  $\beta_s$ 's. En la siguiente proposición se explica esta relación

**Proposición 8.** *Los estimadores  $\hat{\beta}_s$  para  $s = 1, \dots, m$  se comportan como una misma constante conforme  $|\hat{\tau}| \rightarrow \infty$ .*

*Demostración.* La igualdad (10.2) puede reescribirse de la siguiente manera

$$\begin{aligned}\hat{\beta}_s &= \frac{\sum_{i=1}^{k_s} y_{i,s}(x_{i,s} - \hat{\tau})}{\sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})^2} - \hat{\beta}_0 \frac{\sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})}{\sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})^2}, \\ &= \frac{\sum_{i=1}^{k_s} y_{i,s}(x_{i,s} - \hat{\tau})}{\sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})^2} - \frac{\hat{\beta}_0 \hat{\tau} \sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})}{\hat{\tau} \sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})^2}\end{aligned}$$

y conforme  $\hat{\tau}$  crece en magnitud tenemos

$$\lim_{|\hat{\tau}| \rightarrow \infty} \hat{\beta}_s = 0 - \left[ \frac{\sum_{s=1}^m \frac{f_s''}{g_s''}}{\sum_{s=1}^m \frac{h_s}{g_s''}} \right] (-1) \quad (10.6)$$

$$= \frac{\sum_{s=1}^m \sum_{i=1}^{k_s} y_{i,s} x_{i,s} - \sum_{s=1}^m \left[ \frac{1}{k_s} \right] \left[ \sum_{i=1}^{k_s} y_{i,s} \right] \left[ \sum_{i=1}^{k_s} x_{i,s} \right]}{\sum_{s=1}^m \sum_{i=1}^{k_s} x_{i,s}^2 - \sum_{s=1}^m \left[ \frac{1}{k_s} \right] \left[ \sum_{i=1}^{k_s} x_{i,s} \right]^2}. \quad (10.7)$$

El límite (10.7) explica el comportamiento constante, conforme  $|\hat{\tau}|$  crece, de  $\hat{\beta}_s$  para  $s = 1, \dots, m$  observado en la figura 10.4. La constante a la que estos estimadores tienden es idéntica a la asíntota horizontal de  $\hat{\beta}_0/\hat{\tau}$ . Esto completa la demostración.  $\square$

### 10.3. El modelo lineal para estacionalidad constante.

Uno de los modelos básicos para el pronóstico de datos con tendencia lineal y estacionalidad constante es el siguiente

$$y_t = \sum_{s=1}^m \gamma_s D_{s,t} + \gamma_0 t + \mu_t, \quad (10.8)$$

ó utilizando la notación en función de períodos y estaciones

$$y_{i,s} = \gamma_s + \gamma_0 x_{i,s} + \mu_{i,s}. \quad (10.9)$$

La ecuación (10.9) es un modelo de rectas para estacionalidad constante al que haremos referencia como MCONS (léase modelo constante). El MCONS se encuentra en el capítulo de antecedentes en la sección sobre modelos ARMA con filtros diferenciadores y componentes deterministas (ecuación (2.13), página 18). El funcionamiento del modelo es muy simple; ajusta una línea recta a cada conjunto de datos que representa una estación y supone que todas las rectas tienen la misma pendiente. Obviamente, las líneas que ajusta el modelo son paralelas entre sí. Los estimadores MCO de  $\gamma_0$  y  $\gamma_s$  para  $s = 1, \dots, m$  son

$$\hat{\gamma}_0 = \frac{\sum_{s=1}^m \sum_{i=1}^{k_s} y_{i,s} x_{i,s} - \sum_{s=1}^m \left[ \frac{1}{k_s} \right] \left[ \sum_{i=1}^{k_s} y_{i,s} \right] \left[ \sum_{i=1}^{k_s} x_{i,s} \right]}{\sum_{s=1}^m \sum_{i=1}^{k_s} x_{i,s}^2 - \sum_{s=1}^m \left[ \frac{1}{k_s} \right] \left[ \sum_{i=1}^{k_s} x_{i,s} \right]^2}, \quad (10.10)$$

$$\hat{\gamma}_s = \frac{1}{k_s} \sum_{i=1}^{k_s} y_{i,s} - \frac{\hat{\gamma}_0}{k_s} \sum_{i=1}^{k_s} x_{i,s}. \quad (10.11)$$

Notemos que el estimador  $\hat{\gamma}_0$  es el límite al cual tiende  $\hat{\beta}_s$  conforme  $|\hat{\tau}| \rightarrow \infty$  (ecuación (10.7)). Este es un hecho clave para la demostración que veremos en la siguiente sección.

## 10.4. El MNoLIN para series con estacionalidad constante.

El comportamiento de los estimadores y el modelo explicado en la sección anterior son piezas claves para entender cómo el MNoLIN es capaz de capturar los efectos de tendencia lineal y no sólo estacionalidad creciente, sino también constante. Este es el objeto de estudio de este capítulo y de la siguiente proposición.

**Proposición 9.** Conforme  $|\hat{\tau}| \rightarrow \infty$  el MNoLIN dado por la ecuación  $\hat{y}_{i,s} = \hat{\beta}_0 + \hat{\beta}_s(x_{s,i} - \hat{\tau}) + \hat{\mu}_{i,s}$  tiende al MCons dado por  $\hat{y}_{i,s} = \hat{\gamma}_s + \hat{\gamma}_0 x_{i,s} + \hat{\mu}_{i,s}$ .

*Demostración.* El MNoLIN con estimadores MCNL puede reescribirse convenientemente de la siguiente manera

$$\begin{aligned}
\hat{y}_{i,s} &= \hat{\beta}_0 + \hat{\beta}_s x_{s,i} - \hat{\tau} \hat{\beta}_s + \hat{\mu}_{i,s}, \\
&= \hat{\beta}_0 + \hat{\beta}_s x_{s,i} - \frac{\hat{\tau} \sum_{i=1}^{k_s} y_{i,s} (x_{i,s} - \hat{\tau})}{\sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})^2} + \hat{\beta}_0 \frac{\hat{\tau} \sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})}{\sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})^2} + \hat{\mu}_{i,s}, \\
&= \hat{\beta}_0 \left[ 1 + \frac{\hat{\tau} \sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})}{\sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})^2} \right] - \frac{\hat{\tau} \sum_{i=1}^{k_s} y_{i,s} (x_{i,s} - \hat{\tau})}{\sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})^2} + \hat{\beta}_s x_{s,i} + \hat{\mu}_{i,s}, \\
&= \hat{\beta}_0 \left[ \frac{\sum_{i=1}^{k_s} x_{i,s}^2 - \hat{\tau} \sum_{i=1}^{k_s} x_{i,s}}{\sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})^2} \right] - \frac{\hat{\tau} \sum_{i=1}^{k_s} y_{i,s} (x_{i,s} - \hat{\tau})}{\sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})^2} + \hat{\beta}_s x_{s,i} + \hat{\mu}_{i,s}, \\
&= \frac{\hat{\beta}_0}{\hat{\tau}} \left[ \frac{\hat{\tau} \sum_{i=1}^{k_s} x_{i,s}^2 - \hat{\tau}^2 \sum_{i=1}^{k_s} x_{i,s}}{\sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})^2} \right] - \frac{\hat{\tau} \sum_{i=1}^{k_s} y_{i,s} (x_{i,s} - \hat{\tau})}{\sum_{i=1}^{k_s} (x_{i,s} - \hat{\tau})^2} + \hat{\beta}_s x_{s,i} + \hat{\mu}_{i,s}.
\end{aligned}$$

Conforme  $|\hat{\tau}|$  tiende a infinito, tenemos que  $\hat{y}_{i,s}$  se comporta así

$$\begin{aligned}
\lim_{|\hat{\tau}| \rightarrow \infty} \hat{y}_{i,s} &= -\hat{\gamma}_0 \frac{1}{k_s} \sum_{i=1}^{k_s} x_{i,s} + \frac{1}{k_s} \sum_{i=1}^{k_s} y_{i,s} + \hat{\gamma}_0 x_{i,s} + \hat{\mu}_{i,s}, \\
&= \hat{\gamma}_s + \hat{\gamma}_0 x_{i,s} + \hat{\mu}_{i,s}.
\end{aligned} \tag{10.12}$$

El límite (10.12) comprueba que conforme  $|\hat{\tau}| \rightarrow \infty$  el MNoLIN con estimadores MCNL se comporta como el MCons con estimadores MCO. Esto completa la demostración.  $\square$

## 10.5. Caso de ejemplo

Para saber que tan eficaz puede ser el MNoLIN al pronosticar series temporales con tendencia lineal y estacionalidad constante, decidimos pronosticar la demanda mensual de gasolina de Ontario de 1960 a 1966 (cuadro 10.1). Los lineamientos observados durante el cotejo fueron los mismos que aquellos observados al comparar el desempeño de los

modelos para series con estacionalidad creciente de la subsección 7.1 (página 87). Los resultados obtenidos con el MNoLIN se muestran en el cuadro 10.1b mientras que los resultados obtenidos con el MCons se muestran en el cuadro 10.1a.

**Cuadro 10.2** – Resultados de pronosticar la demanda de gasolina de Ontario de 1960 a 1966.  
(a) MCons; AR(1, 3). (b) MNoLIN; AR(1, 3).

<i>h</i>	COB	EST	DESV	DESV %	<i>h</i>	COB	EST	DESV	DESV %
1	0,7083	8026	3382,1	2,4512	1	0,7083	7796	3190,7	2,3132
2	0,7083	8660	3415,5	2,4947	2	0,7500	8461	3275,7	2,3904
3	0,6666	8769	3642,7	2,6518	3	0,6666	8583	3345,1	2,4177
4	0,5833	9438	4189,6	3,0011	4	0,6250	9208	4027,1	2,8804
5	0,6250	9914	4180,8	2,9968	5	0,6250	9702	4013,0	2,8814
6	0,6250	10110	4032,8	2,8823	6	0,6250	9919	3924,1	2,8093
7	0,5833	10165	4533,9	3,2146	7	0,5833	9961	4397,2	3,1148
8	0,6250	10339	4249,5	3,0385	8	0,5833	10133	4321,5	3,0830
9	0,6666	10493	4367,6	3,1235	9	0,6250	10302	4229,6	3,0277
10	0,6666	10508	4344,6	3,0915	10	0,6666	10320	4189,9	2,9861
11	0,7083	10542	4269,4	3,0284	11	0,7083	10349	4093,5	2,9085
12	0,6666	10614	4573,7	3,2215	12	0,7500	10424	4325,6	3,0652

Para saber si las diferencias observadas en el cuadro 10.2 son significativas, utilizamos la prueba de rangos con signo de Wilcoxon de manera similar a lo hecho en la sección 7.2.1 (página 92) y a un nivel de significancia  $\alpha = 0,05$ . A continuación, se resumen los resultados de cada estadístico:

- En cuanto a los resultados del estadístico  $COB_{dist}$ , éstos fueron favorables para el MNoLIN, sin embargo, no fue posible rechazar  $H_0$ .
- Los resultados al comparar los modelos de acuerdo al estadístico Est, fueron favorables para el MNoLIN. La prueba de Wilcoxon resultó significativa en esta ocasión.
- En cuanto al pronóstico, los resultados de comparar los modelos de acuerdo al estadístico DESV % fueron favorables para el MNoLIN, siendo estos resultados significativos de acuerdo a la prueba de Wilcoxon.

Los resultados anteriores, no sólo indican que el MNoLIN puede pronosticar series con estacionalidad constante, sino que también es capaz de compararse, en cuanto a desempeño

de pronóstico, al MCONS e incluso puede superar a este modelo en su propio terreno.

## 10.6. Conclusiones

En este capítulo, se ha demostrado matemáticamente la habilidad del MNoLIN de pronosticar no sólo series temporales con estacionalidad creciente, sino también series con estacionalidad constante. Esto ocurre cuando el punto en el que se interceptan las rectas tiende a ser muy grande en magnitud.

En la práctica, el MNoLIN demostró ser efectivo al pronosticar la demanda de gasolina de Ontario mejorando a su contraparte; el MCONS.

De nuestro conocimiento es que no hay ningún método en la literatura capaz de pronosticar tanto series con estacionalidad creciente como constante. Esta es la razón por la cual es notable que el MNoLIN si pueda hacer esta labor. Esto abre la posibilidad de no tener que decidir a priori si la estacionalidad de un modelo es creciente o constante para poder proponer un modelo adecuado para su pronóstico. Con el MNoLIN ya no es necesario elegir entre un modelo para estacionalidad creciente o constante ya que este modelo es capaz de pronosticar bien ambos efectos.

---

---

# CAPÍTULO 11

---

## Resultados y contribuciones

**R**esumiremos los resultados y las contribuciones teóricas y prácticas del presente trabajo. Las contribuciones y los resultados van, desde los modelos propuestos, pasando por la resolución de los problemas de optimización asociados a cada uno, hasta los resultados prácticos de los experimentos que incluyen la implementación computacional así como los casos de ejemplo y por último, la capacidad del modelo no lineal de pronosticar series con estacionalidad constante o creciente.

### 11.1. Modelos propuestos

Los modelos propuestos para el pronóstico de STTLEC son dos: uno es de regresión lineal y el otro es de regresión no lineal. Como característica principal, los modelos propuestos son altamente deterministas, capaces de pronosticar series con observaciones faltantes y utilizan muy pocos parámetros reduciendo así el riesgo de sobreparametrización. A continuación, describimos más específicamente cada modelo.

#### El $ML_{LN}$

Este es un modelo de regresión lineal para el pronóstico de STTLEC. Tres notaciones fueron propuestas y utilizadas para este modelo, siendo cada notación más adecuada para cierta labor. Por ejemplo, la estimación de parámetros se facilitó con la notación estacional

y la vectorial mientras que la caracterización estadística se facilitó con la notación de variables binarias.

El  $MLIN$  supone que las STTLEC pueden ser modeladas con una recta cuya pendiente varía en cada estación. Este modelo es equivalente a ajustar una recta a cada estación del año y suponer que las rectas se interceptan en algún punto del eje de las ordenadas. La ventaja de este modelo radica en su simplicidad y sobretodo en su linealidad. La desventaja de este modelo se encuentra en lo limitado de las amplitudes que pueden ser modeladas con él, pues supone que las líneas rectas que subyacen a cada estación se interceptan siempre en el eje de las ordenadas.

### **El $MNoLIN$**

Este es un modelo de regresión no lineal capaz de pronosticar STTLEC. Se propusieron tres notaciones para representar este modelo y cada notación resultó más adecuada dependiendo de lo que se hacía con ella. La estimación de parámetros fue más simple con la notación vectorial, la caracterización estadística fue más simple con la notación de variables binarias y la prueba de la capacidad de este modelo para pronosticar series con estacionalidad constantes fue más simple con la notación estacional.

La interpretación geométrica reveló que este modelo equivale a ajustar una recta a cada estación y que las rectas tienen un punto común de intersección en algún lugar del plano. La ventaja de este modelo es su capacidad de capturar una gran variedad de amplitudes estacionales. La desventaja de este modelo radica en su no linealidad pues dificulta entre otras cosas, la estimación de parámetros.

## **11.2. Estimación de parámetros**

A continuación, se presentan los procedimientos de estimación paramétrica para cada modelo.

### **Estimadores MCO para el $MLIN$**

Para el caso del  $MLIN$ , fue relativamente sencillo estimar sus parámetros, ya que éste es un modelo de regresión lineal cuyos parámetros pueden ser estimados con simplicidad utilizando los clásicos estimadores MCO.

### **Estimadores MCNL para el $MNoLIN$**

La estimación de parámetros del  $MNoLIN$  no fue tan simple. Debido a la no linealidad de este modelo, no fue posible obtener estimadores MCO. En el contexto de modelos no lineales, los estimadores que minimizan la SDC son los estimadores MCNL. Para encontrar los estimadores MCNL, propusimos un algoritmo que saca ventaja de la estructura del modelo. También, utilizamos la técnica estándar para realizar esta labor; el algoritmo de Levenberg-Marquardt.

Para saber cual de las técnicas anteriores para minimizar la SDC es mejor, comparamos ambos procedimientos. El experimento indica que la técnica propuesta, el algoritmo de BLSD, mejoraba ampliamente al algoritmo de Levenberg-Marquardt.

## **11.3. Caracterización estadística**

A continuación, se describen las aportaciones estadísticas de este trabajo.

### **Estimadores MELI o insesgados y consistentes para el $MLIN$**

Se demostró que los estimadores MCO del  $MLIN$  son MELI si la fuente de variación sigue un proceso  $NID(0, \sigma^2)$ . También, se demostró que los estimadores MCO son insesgados y consistentes si las desviaciones siguen un proceso  $ARMA(p, q)$ .

Además, se dedujeron los intervalos predictivos para este modelo bajo los dos supuestos anteriores sobre la fuente de variación.

### **Estimadores condicionalmente MELI o condicionalmente insesgados y consistentes para el MNoLIN**

Se demostró que los estimadores MCNL del MNoLIN son MELI bajo las suposiciones de que el parámetro  $\tau$  es conocido y que la fuente de variación sigue un proceso NID( $0, \sigma^2$ ). Si la fuente de variación sigue un proceso ARMA( $p, q$ ) y  $\tau$  es conocido entonces, los estimadores MCNL son insesgados y consistentes.

Además, se dedujeron los intervalos predictivos para este modelo bajo los dos supuestos anteriores sobre la fuente de variación y bajo el supuesto de que  $\tau$  es conocido.

### **Estimadores incondicionalmente MELI o incondicionalmente insesgados y consistentes para el MNoLIN**

Se demostró que era posible estimar los parámetros del MNoLIN mediante una técnica analítica. Los estimadores obtenidos mediante este procedimiento eran MELI si la fuente de variación seguía un proceso NID( $0, \sigma^2$ ) o insesgados y consistentes si la fuente de variación seguía un proceso ARMA( $p, q$ ). Una de las ventajas de esta estimación es el no necesitar de la implementación de código computacional pues bajo este procedimiento, los estimadores se obtienen de forma analítica. Otra ventaja de este procedimiento radica en no necesitar suponer que  $\tau$  es conocido para dar estimadores MELI o insesgados y consistentes. La desventaja de este procedimiento es que sólo se disponen de estimadores para  $\beta_s \forall s$  y no hay estimaciones para  $\tau$  o  $\beta_0$ . Sin embargo, esta técnica de estimación sirve para saber que tan buenos son los estimadores MCNL frente a técnicas alternativas de estimación.

Se realizó un experimento en el que se comparó la eficacia de los estimadores analíticos contra los estimadores MCNL obtenidos mediante el algoritmo de búsqueda lineal con sección dorada. Los resultados fueron muy favorables para los estimadores obtenidos mediante el segundo procedimiento.

## 11.4. Códigos computacionales

A continuación, se describen los códigos desarrollados para este trabajo.

### **Código desarrollado para el *software* Matlab/Octave para todo lo anterior**

La estimación de parámetros así como la caracterización estadística para los modelos propuestos y también para los modelos comparativos (el MALTDB y el MCONS) fueron tareas programadas en el software Matlab/Octave. Este software fue utilizado debido a las facilidades que ofrece para manipular arreglos matriciales y vectoriales.

### **Código desarrollado para el software Matlab/Octave para la eliminación de ruido autorregresivo**

También se implementó en el software Matlab/Octave un desarrollo para estimar los parámetros de los procesos  $AR(p)$ . Este algoritmo fue diseñado para eliminar el ruido autorregresivo de las desviaciones y tiene como objetivo maximizar la función de verosimilitud. Esta labor se realiza con *el algoritmo simplex de Nelder y Mead*, que es una de las muchas funciones de Octave.

## 11.5. Experimentación: casos de ejemplo

Se compararon los dos modelos propuestos (el MLIN y el MNoLIN) contra la alternativa *apta* de la literatura (el MALTDB). Se utilizaron seis series de ejemplo en tres diferentes escenarios de pronóstico: el número de observaciones es regular, el número de observaciones es escaso y hay observaciones faltantes. Se utilizó la prueba de rangos con signo de Wilcoxon para saber si las diferencias en los estadísticos, observadas al contrastar cada par de modelos, eran significativas al nivel  $\alpha = 0,05$ . Los resultados obtenidos se resumen a continuación.

### **El número de observaciones es regular**

En términos generales, los resultados en cuanto a cobertura de los intervalos predictivos fueron favorables para el  $MLIN$ . Los resultados en cuanto a la estrechez de los intervalos predictivos fueron favorables para el  $MALTDDB$ . Finalmente, el modelo que mejor pronosticó en términos del estadístico  $DES\%$  fue el  $MNoLIN$ .

### **El número de observaciones es escaso**

Los modelos propuestos pueden pronosticar series temporales con  $m + 1$  y  $m + 2$  observaciones. Sin embargo, el modelo comparativo, el  $MALTDDB$  necesita de  $2m$  observaciones para dar un pronóstico y de  $2m + 1$  observaciones para deducir sus intervalos predictivos. Para comparar los modelos propuestos con el  $MALTDDB$ , tomamos los primeros 25 datos de los seis casos de ejemplo y procedimos a pronosticar las siguientes 12 observaciones 24 veces, aumentando la serie con el siguiente dato en cada ocasión. La única suposición factible sobre la fuente de variación, fue que ésta seguía un proceso  $NID(0, \sigma^2)$ . Los resultados indican que el  $MNoLIN$  arroja el mejor pronóstico. La cobertura de los intervalos predictivos del  $MLIN$  fue la mejor en este escenario. Finalmente, el modelo con la mejor estrechez de intervalos predictivos fue el  $MNoLIN$ .

### **Pronóstico de series temporales con observaciones faltantes**

Los modelos propuestos tienen la capacidad de modelar series con observaciones faltantes de manera directa y sin necesidad de pronosticar el dato faltante como se sugiere en algunas otras metodologías. Para conocer el comportamiento de los modelos conforme aumentaba el número de datos faltantes en la serie, diseñamos un experimento en el cual se extraían observaciones de la serie, una por una, y se registraba el desempeño del pronóstico con cada número de observaciones faltantes, el cual varió desde 0 hasta 18. Los resultados de este experimento indican que el modelo que mejor pronosticó cuando no

había datos faltantes, fue también el modelo que mejor desempeño tuvo cuando el número de observaciones faltantes fue en aumento. En 5 de los 6 casos de ejemplo, el mejor modelo fue el MNoLIN, mientras que para la serie 2, el mejor modelo fue el MLIN.

## **11.6. Innovación, un modelo para el pronóstico de series con estacionalidad constante o creciente**

Esta es una de las contribuciones teóricas más importantes de este trabajo. El MNoLIN es capaz de pronosticar series con estacionalidad creciente muy bien, según los resultados de los seis casos de ejemplo. Sin embargo, la eficacia de este modelo, no se limita al pronóstico de series con estacionalidad creciente, sino que también pronostica series con estacionalidad constante.

La manera en que se demostró la capacidad del MNoLIN de pronosticar series con estacionalidad constante fue matemática más que comparativa. Sin embargo, los resultados del caso de ejemplo demuestran que hay series con estacionalidad constante que pueden ser mejor pronosticadas con el MNoLIN que con la técnica *apta* de la literatura; el MCons.

---

---

# CAPÍTULO 12

---

## Conclusiones y trabajos futuros.

Tres aspectos importantes, sobre los que se puede concluir y sobre los que aún hay trabajo para futuros proyectos son: pronóstico e intervalos predictivos, estimación paramétrica y series temporales con estacionalidad constante o creciente. A continuación, se discute de manera detallada cada uno de estos aspectos.

### Pronóstico e intervalos predictivos

El MNoLIN es el modelo que mejor pronosticó los casos de ejemplo cuando había información regular, información escasa y observaciones faltantes.

En el primero de los escenarios anteriores, el mejor modelo de pronóstico fue el MNoLIN con la suposición de ruido  $AR(p)$  mientras que en los otros dos escenarios, el mejor modelo fue el MNoLIN con la suposición de ruido  $NID(0, \sigma^2)$  (la única suposición posible en estos escenarios).

En cuanto a los intervalos predictivos, si bien, el MLIN tuvo mejor cobertura, esto se debió al hecho de que la varianza del error de pronóstico estimada con este modelo fue muy grande; dicho de otra manera, si la cobertura de este modelo fue buena es debido a que su pronóstico fue malo. En esta situación, el MNoLIN logró un equilibrio entre buen pronóstico, cobertura y estrechez de los intervalos predictivos. No sólo se trata de que los intervalos predictivos tengan una cobertura próxima a la teórica, sino de que también sean tan estrechos como sea posible.

El trabajo futuro en cuanto a modelado y pronóstico se refiere es, como lo habrá podido advertir el lector, extender la capacidad del MNoLIN para hacerlo capaz de pronosticar series con tendencia polinomial. Quizá se podría pensar que no es tan difícil extender el MNoLIN a casos de series con tendencia polinomial. Basta con sustituir el término  $(x_{i,s} - \tau)$  por algo como  $(\alpha_3 x_{i,s}^3 + \alpha_2 x_{i,s}^2 + \alpha_1 x_{i,s} - \tau)$  por decir algo. Ciertamente, lo anterior no tiene mayor grado de dificultad. Sin embargo, lo que si resulta difícil, no es el modelado, sino la estimación paramétrica, la caracterización estadística, compararlo con otras metodologías usando series temporales con tendencia polinomial y finalmente, demostrar si el modelo es capaz de pronosticar series con tendencia lineal y estacionalidad constante o creciente. Lo anterior puede perfectamente ser el tema de otra tesis.

### **Estimación paramétrica**

De las tres estimaciones paramétricas propuestas para el MNoLIN, la mejor opción fue la de minimizar la *suma de desviaciones cuadradas* (SDC) mediante el algoritmo de *búsqueda lineal con sección dorada* (BLSD). Este algoritmo fue mejor que el procedimiento de Levenber-Marquardt cuyo objetivo es el mismo. Los resultados del experimento comparativo indican que el algoritmo de BLSD minimiza el objetivo hasta 10 veces más que el algoritmo de Levenberg-Marquardt.

El algoritmo de BLSD fue también mejor que los estimadores analíticos. Los estimadores obtenidos con el algoritmo de BLSD estuvieron en promedio 50 % más cerca de los parámetros que los estimadores analíticos.

En cuanto a estimación paramétrica se refiere, el trabajo futuro es proponer otras funciones objetivo. Por ejemplo, si las desviaciones de las series exhiben heterocedasticidad, los estimadores de *mínimos cuadrados ordinarios* (MCO) y *mínimos cuadrados no lineales* (MCNL) siguen siendo insesgados y consistentes. Sin embargo, puede haber estimaciones paramétricas más eficientes que las que resultan de minimizar la SDC. Los estimadores *Mínimos Cuadrados Generalizados* pueden ser una técnica que arroje estimadores más

eficientes. Sería entonces necesario crear un algoritmo para obtener los estimadores antes mencionados en el contexto de modelos no lineales, un tema poco tratado en la literatura.

### **Series temporales con estacionalidad constante o creciente**

Por último, el MNoLIN tiene una capacidad nunca antes vista en la literatura; pronostica series con estacionalidad creciente o constante. La capacidad de este modelo de pronosticar series con estacionalidad constante se debe a que conforme uno de sus parámetros tiende a infinito, el modelo, con estimadores MCNL, tiende al MCons, con estimadores MCO. El MCons es un modelo *apto* para el pronóstico de series con estacionalidad constante y por tanto, el MNoLIN puede pronosticar series de este tipo. El caso de ejemplo demostró que hay series con estacionalidad constante que pueden ser pronosticadas mejor con el MNoLIN que con el MCons.

El trabajo futuro en este tema es muy extenso; por ejemplo, ahora que sabemos que hay modelos capaces de pronosticar series con estacionalidad creciente o constante podríamos explorar si habrá modelos en otras metodologías capaces de pronosticar series con ambos tipos de estacionalidad y si será mejor usar modelos para un tipo específico de estacionalidad o modelos que no supongan nada sobre el tipo de estacionalidad. Las cuestiones anteriores son sólo algunas de las muchas posibles incógnitas generadas por el MNoLIN y cuyas respuestas las irá dando la profunda investigación de cada una de ellas.

---

---

# APÉNDICE A

---

## Optimalidad de los puntos críticos

Para saber si los valores críticos de los parámetros  $\hat{\beta}_0$  y  $\hat{\beta}_s \forall s$  minimizan la SDC, necesitaremos del siguiente teorema

**Teorema 3.** Si  $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$  es de clase  $C_2$ ,  $\mathbf{z}_0 \in U$  es un punto crítico de  $f$  y el hessiano  $Hf(\mathbf{z}_0)$  es definido positivo, entonces  $\mathbf{z}_0$  es un mínimo relativo de  $f$ .

Para más detalles sobre el teorema anterior, consulte [Marsden y Tromba \(1998, p. 195\)](#) o cualquier libro de calculo vectorial. Para nuestro problema  $f = \text{SDC}$ ;  $\mathbf{z}_0 = [\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m, \hat{\beta}_0]'$  y  $Hf(\mathbf{z}_0)$  tiene la forma

$$Hf(\mathbf{z}_0) = \begin{bmatrix} g_1 & 0 & \dots & 0 & d_1 \\ 0 & g_2 & \dots & 0 & d_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & g_m & d_m \\ d_1 & d_2 & \dots & d_m & 2n \end{bmatrix},$$

donde  $g_s = 2 \sum_{i=1}^{k_s} x_{i,s}^2$  y  $d_s = 2 \sum_{i=1}^{k_s} x_{i,s}$ .

Es evidente que  $Hf(\mathbf{z}_0)$  es una matriz simétrica y para saber si es definida positiva necesitaremos del siguiente teorema

**Teorema 4.** Cualquier matriz  $B$  de  $n \times n$  que sea simétrica es definida positiva si y sólo si los determinantes de las submatrices diagonales de  $B$  son todos mayores que cero.

Para nuestro caso, es evidente que los determinantes de las primeras  $m$  submatrices son mayores que cero ya que cada una de estas submatrices es diagonal y el determinante de una matriz diagonal es igual al producto de los elementos de la diagonal. Estos elementos diagonales son sumas de cantidades elevadas al cuadrado y por esta razón, su producto será positivo siempre y cuando exista una  $s$  entre 1 y  $m$  tal que para toda  $i$  entre 1 y  $k_s$ ,  $x_{i,s} > 0$ . Es difícil imaginar un caso en el que esto no ocurra dadas las suposiciones del MLIN.

El calculo del último determinante no es tan fácil y requiere de operaciones elementales de renglón de la matriz  $Hf(\mathbf{z}_0)$ . Después de tales operaciones, tenemos que el determinante de la última submatriz diagonal de  $Hf(\mathbf{z}_0)$  es equivalente al determinante de esta matriz y está dado por

$$\det Hf(\mathbf{z}_0) = \left( 2n - \sum_{s=1}^m \frac{d_s^2}{g_s} \right) \prod_{s=1}^m g_s \quad (\text{A.1})$$

El producto del lado derecho de la ecuación A.1 es positivo por la misma razón que los determinantes de las primeras  $m$  submatrices también lo son. Sin embargo, para que el lado derecho de esta ecuación sea positivo, es necesario que se cumpla la siguiente relación

$$\sum_{s=1}^m \frac{(\sum_{i=1}^{k_s} x_{i,s})^2}{\sum_{i=1}^{k_s} x_{i,s}^2} < n \quad (\text{A.2})$$

Si la relación de la ecuación A.2 se cumple entonces, es posible asegurar que  $\mathbf{z}_0$  minimiza la SDC. De lo contrario, no es posible garantizar nada.

Para darnos una idea de que tan posible es que la desigualdad A.2 no se cumpla (esto es, que  $\mathbf{z}_0$  no sea el óptimo) supondremos un peor caso y veremos lo que pasa. Este peor caso ocurre cuando la suma

$$\sum_{s=1}^m \frac{(\sum_{i=1}^{k_s} x_{i,s})^2}{\sum_{i=1}^{k_s} x_{i,s}^2} \quad (\text{A.3})$$

toma su mayor valor. Esto debido a que ese mayor valor puede ser superior a  $n$  y esto impediría el cumplimiento de la desigualdad A.2. El peor caso de la suma A.3 ocurre

cuando cada cociente

$$\frac{(\sum_{i=1}^{k_s} x_{i,s})^2}{\sum_{i=1}^{k_s} x_{i,s}^2} \quad (\text{A.4})$$

toma su mayor valor. Supongamos ahora que  $\sum_{i=1}^{k_s} x_{i,s} = \alpha_s$ . El mayor valor que el cociente A.4 puede tomar ocurre cuando  $\sum_{i=1}^{k_s} x_{i,s}^2$  toma su valor mínimo (sujeto a la restricción  $\sum_{i=1}^{k_s} x_{i,s} = \alpha_s$ ) pues, al ser el denominador, entre más pequeño es este, más grande es el valor del cociente A.4. Más aún, el cociente A.4 es independiente de  $\alpha_s$  cuando  $\sum_{i=1}^{k_s} x_{i,s}^2$  toma su valor mínimo, como se demuestra en el Apéndice B. En este mismo apéndice, se muestra que el mayor valor que puede tomar el cociente A.4 es  $k_s$ , lo cual nos conduce a violar la desigualdad A.2 pues

$$\sum_{s=1}^m \frac{(\sum_{i=1}^{k_s} x_{i,s})^2}{\sum_{i=1}^{k_s} x_{i,s}^2} = \sum_{s=1}^m k_s = n \not\leq n \quad (\text{A.5})$$

La desigualdad A.2 apenas es violada en nuestro peor caso pues, evidentemente  $n \not\leq n$ . En cualquier otro caso, la violación A.5 no ocurriría pues, el cociente A.4 sería menor a  $k_s$  y por lo tanto, la suma A.3 sería menor a  $n$ , cumpliendo la desigualdad A.2.

Según el Apéndice B, el peor caso ocurre cuando  $x_{i,s} = \alpha_s/k_s \forall s$ , así que, mientras esto no ocurra, (es decir, mientras alguna  $x_{i,s} \neq \alpha_s/k_s$ ) entonces, podremos garantizar que  $\mathbf{z}_0$  minimiza la SDC. Afortunadamente, el peor caso rara vez ocurre en la practica, a menos que tengamos un ciclo o menos de observaciones ( $n \leq m$ ), en cuyo caso, la desigualdad A.2 deberá ser revisada. Si tenemos más de un ciclo de observaciones,  $\mathbf{z}_0$  será un valor óptimo que satisface las condiciones del teorema 3 y por ser único, minimiza la SDC de manera global.

---

---

# APÉNDICE B

---

## El valor del coeficiente en un peor caso

El valor más grande que puede tomar el coeficiente A.2 de la página 143 se alcanza cuando su denominador toma un valor mínimo. Suponga que  $\sum_{i=1}^{k_s} x_{i,s} = \alpha_s$ . Entonces, para conocer los valores de las  $x_{i,s}$  que minimizan el denominador, es necesario resolver el siguiente problema de optimización

$$\begin{array}{ll} \text{maximizar} & \sum_{i=1}^{k_s} x_{i,s}^2, \\ \text{sujeto a} & \sum_{i=1}^{k_s} x_{i,s} = \alpha_s. \end{array}$$

La solución óptima de este problema se alcanza cuando  $x_{i,s} = \alpha_s/k_s \forall s$ . Esta solución satisface las condiciones de Karush-Kuhn-Tucker, lo que significa que constituyen un óptimo global.

Sustituyendo el valor óptimo en el denominador y el de la restricción en el numerador tenemos

$$\frac{(\sum_{i=1}^{k_s} x_{i,s})^2}{\sum_{i=1}^{k_s} x_{i,s}^2} = \frac{(\alpha_s)^2}{\sum_{i=1}^{k_s} (\alpha_s/k_s)^2} = k_s$$

Lo anterior demuestra que en un peor caso el cociente, no sólo es independiente de  $\alpha_s$ , sino que es igual al índice  $k_s$ .

---

---

# APÉNDICE C

---

## Resultados de pronosticar series con información regular

A continuación se muestran los resultados del experimento con el primer conjunto de datos en el que la cantidad de observaciones es regular. Los lineamientos observados así como el significado de los estadísticos COB, EST, DESV y DESV % y también el significado de la descripción de cada cuadro de este apéndice, se encuentran en la sección 7.1 (página 87).

### C.1. Resultados con el MLIN

**Cuadro C.1** – Serie 1; MLIN; AR(1).

$h$	COB	EST	DESV	DESV %
1	0,750	163,235	52,468	5,898
2	0,833	171,842	52,764	5,821
3	0,791	172,797	52,038	5,673
4	0,833	172,910	50,133	5,441
5	0,875	172,924	48,551	5,206
6	0,916	172,925	45,559	4,819
7	0,875	172,925	46,098	4,817
8	0,875	172,926	46,412	4,810
9	0,875	172,926	45,304	4,643
10	0,875	172,926	45,980	4,708
11	0,875	172,926	45,041	4,576
12	0,875	172,926	46,865	4,745

**Cuadro C.2** – Serie 2; MLIN; AR(1, 3, 7).

$h$	COB	EST	DESV	DESV %
1	0,625	8636,1	3128,5	8,836
2	0,666	8924,4	3565,5	10,13
3	0,708	8945,1	3544,2	9,995
4	0,708	9126,3	3591,8	10,07
5	0,791	9171,1	3284,1	9,212
6	0,833	9178,3	3093,1	8,651
7	0,875	9189,5	2940,9	8,256
8	0,833	9455,3	3097,4	8,838
9	0,833	9521,9	3003,8	8,504
10	0,833	9529,4	3074,9	8,570
11	0,833	9586,3	3192,4	8,868
12	0,833	9614,9	3283,9	9,071

**Cuadro C.3** – Serie 3; MLIN; AR(1, 2, 3).

<i>h</i>	COB	EST	DESV	DESV %
1	0,541	89,834	47,430	14,371
2	0,708	114,253	55,509	16,026
3	0,541	139,431	60,871	17,861
4	0,583	144,212	60,478	17,827
5	0,625	145,149	63,227	18,299
6	0,625	146,531	61,844	17,559
7	0,708	149,526	59,540	16,586
8	0,666	153,963	55,575	15,705
9	0,750	156,428	51,317	14,635
10	0,791	157,442	51,347	15,102
11	0,791	157,560	48,106	13,977
12	0,791	157,813	46,472	12,950

**Cuadro C.4** – Serie 4; MLIN; AR(1, 3)\*.

<i>h</i>	COB	EST	DESV	DESV %
1	0,833	111,401	29,412	3,468
2	0,708	135,041	35,590	4,169
3	0,833	144,818	36,860	4,287
4	0,833	144,856	35,803	4,131
5	0,875	146,299	36,997	4,283
6	0,875	149,235	37,621	4,305
7	0,833	150,713	39,794	4,534
8	0,833	150,948	39,980	4,558
9	0,833	150,979	40,120	4,541
10	0,833	151,204	39,020	4,413
11	0,833	151,422	39,815	4,511
12	0,833	151,494	40,909	4,624

**Cuadro C.5** – Serie 5; MLIN; AR(1).

<i>h</i>	COB	EST	DESV	DESV %
1	0,833	16,319	4,524	3,458
2	0,875	17,282	4,955	3,789
3	0,833	17,396	5,018	3,794
4	0,833	17,410	5,148	3,874
5	0,833	17,412	5,186	3,898
6	0,875	17,412	4,888	3,654
7	0,875	17,412	4,895	3,645
8	0,875	17,412	4,922	3,658
9	0,875	17,412	4,873	3,601
10	0,875	17,412	4,835	3,549
11	0,875	17,412	4,701	3,443
12	0,916	17,412	4,577	3,334

**Cuadro C.6** – Serie 6; MLIN; NID(0,  $\sigma^2$ ).

<i>h</i>	COB	EST	DESV	DESV %
1	0,875	337,284	74,402	12,092
2	0,875	336,621	72,577	11,827
3	0,875	335,973	71,565	11,675
4	0,875	335,339	71,026	11,588
5	0,875	334,719	68,795	11,278
6	0,875	334,113	66,187	10,901
7	0,875	333,520	61,766	10,341
8	0,875	332,940	62,519	10,354
9	0,875	332,373	63,553	10,374
10	0,916	331,818	64,439	10,462
11	0,916	331,274	64,862	10,492
12	0,916	330,742	63,456	10,303

## C.2. Resultados con el MNoLIN

**Cuadro C.7** – Serie 1; MNoLIN; AR(12).

$h$	COB	EST	DESV	DESV %
1	0,7916	141,480	43,282	4,773
2	0,8333	141,480	41,313	4,502
3	0,8333	141,480	41,667	4,497
4	0,8750	141,480	40,509	4,352
5	0,8750	141,480	42,319	4,486
6	0,9166	141,480	39,516	4,113
7	0,8750	141,480	41,089	4,265
8	0,8750	141,480	42,822	4,417
9	0,8750	141,480	43,290	4,444
10	0,8750	141,480	45,103	4,621
11	0,8750	141,480	43,323	4,414
12	0,8750	141,480	42,669	4,325

**Cuadro C.8** – Serie 2; MNoLIN; AR(1, 3, 7).

$h$	COB	EST	DESV	DESV %
1	0,625	8448,4	3194,9	8,951
2	0,541	8754,2	3747,0	10,56
3	0,625	8777,6	3752,4	10,49
4	0,666	8993,1	3772,3	10,49
5	0,708	9049,3	3497,7	9,715
6	0,750	9058,8	3346,7	9,253
7	0,791	9075,5	3199,6	8,861
8	0,791	9315,9	3242,7	9,142
9	0,791	9381,7	3182,9	8,924
10	0,791	9389,0	3270,5	9,028
11	0,791	9456,4	3330,5	9,178
12	0,791	9492,6	3368,1	9,258

**Cuadro C.9** – Serie 3; MNoLIN; AR(1, 2).

$h$	COB	EST	DESV	DESV %
1	0,6666	85,236	39,598	10,477
2	0,6666	97,726	42,705	10,584
3	0,5833	105,966	47,793	12,379
4	0,6250	110,157	50,354	13,402
5	0,5416	112,661	52,170	13,574
6	0,5000	114,099	52,832	13,226
7	0,5416	114,956	54,080	13,481
8	0,5416	115,464	50,357	12,536
9	0,6250	115,768	46,537	11,838
10	0,6666	115,951	46,310	12,210
11	0,6666	116,061	45,532	11,605
12	0,7083	116,127	47,818	11,851

**Cuadro C.10** – Serie 4; MNoLIN; AR(1, 2, 3, 4).

$h$	COB	EST	DESV	DESV %
1	0,8333	45,429	14,859	1,844
2	0,7500	47,228	15,093	1,885
3	0,7916	48,848	13,749	1,684
4	0,7916	49,353	13,350	1,623
5	0,8750	50,400	12,335	1,489
6	0,8333	51,351	13,655	1,645
7	0,7500	51,578	14,121	1,693
8	0,7500	51,585	13,792	1,663
9	0,8333	51,717	13,904	1,664
10	0,8333	51,881	12,911	1,539
11	0,8750	51,943	13,502	1,608
12	0,8333	51,945	13,849	1,649

**Cuadro C.11** – Serie 5; MNoLIN; AR(1).

$h$	COB	EST	DESV	DESV %
1	0,8333	8,223	2,221	1,744
2	0,9166	8,506	2,154	1,652
3	0,9166	8,525	2,074	1,573
4	0,8750	8,527	2,161	1,615
5	0,8750	8,527	2,249	1,678
6	0,8333	8,527	2,325	1,713
7	0,7916	8,527	2,470	1,801
8	0,7916	8,527	2,599	1,892
9	0,7083	8,527	2,863	2,063
10	0,7083	8,527	2,900	2,070
11	0,7083	8,527	2,801	1,998
12	0,7500	8,527	2,716	1,918

**Cuadro C.12** – Serie 6; MNoLIN; NID(0,  $\sigma^2$ ).

$h$	COB	EST	DESV	DESV %
1	0,7500	152,616	42,682	5,996
2	0,7916	152,553	41,160	5,788
3	0,7916	152,491	43,009	5,948
4	0,7916	152,430	43,510	6,020
5	0,7916	152,370	43,633	5,992
6	0,8333	152,310	41,357	5,623
7	0,8750	152,252	39,870	5,396
8	0,9166	152,196	37,582	5,006
9	0,9166	152,141	36,698	4,666
10	0,9166	152,088	38,049	4,847
11	0,9166	152,037	38,519	4,897
12	0,9583	151,988	36,175	4,596

### C.3. Resultados con el MAltDB

**Cuadro C.13** – Serie 1; MAltDB; AR(1, 3).

$h$	COB	EST	DESV	DESV %
1	0,5833	123,962	51,040	5,652
2	0,5833	123,962	48,557	5,280
3	0,6250	123,962	47,715	5,131
4	0,6250	123,962	49,561	5,292
5	0,6250	123,962	47,993	5,069
6	0,6666	123,962	44,104	4,614
7	0,6250	123,962	45,572	4,731
8	0,6250	123,962	45,253	4,636
9	0,6250	123,962	46,240	4,757
10	0,6250	123,962	45,382	4,635
11	0,6666	123,962	43,193	4,352
12	0,6666	123,962	45,016	4,521

**Cuadro C.14** – Serie 2; MAltDB; AR(3, 7, 12).

$h$	COB	EST	DESV	DESV %
1	0,625	8240,3	3828,5	10,665
2	0,625	8240,3	3784,3	10,578
3	0,625	8240,3	3703,6	10,281
4	0,666	8454,1	3728,8	10,285
5	0,708	8454,1	3534,5	9,7451
6	0,750	8454,1	3413,6	9,3972
7	0,750	8466,3	3344,8	9,2293
8	0,708	8620,6	3564,4	10,013
9	0,750	8620,6	3502,8	9,7354
10	0,791	8621,4	3497,0	9,5890
11	0,791	8655,5	3586,9	9,8077
12	0,791	8655,5	3681,1	9,9977

**Cuadro C.15** – Serie 3; MAltDB;  
AR(1, 2, 3, 12, 14).

$h$	COB	EST	DESV	DESV %
1	0,3333	65,628	54,355	15,105
2	0,4583	77,027	59,474	15,618
3	0,4166	92,056	63,040	16,426
4	0,4583	97,609	65,953	17,386
5	0,4166	102,519	72,041	18,883
6	0,4583	104,619	74,219	18,829
7	0,3750	106,234	80,467	20,594
8	0,3750	107,088	72,786	18,988
9	0,3750	107,706	68,827	18,165
10	0,3750	108,107	66,895	17,827
11	0,5000	108,385	60,524	15,477
12	0,5000	108,581	60,494	14,899

**Cuadro C.16** – Serie 4; MAltDB; AR(1, 3).

$h$	COB	EST	DESV	DESV %
1	0,6666	39,695	17,003	2,116
2	0,5416	41,087	18,686	2,334
3	0,5833	41,192	17,701	2,192
4	0,6250	43,058	16,653	2,028
5	0,7083	43,593	16,618	2,012
6	0,7083	43,685	17,171	2,078
7	0,7083	43,828	17,080	2,061
8	0,7500	43,943	16,415	1,988
9	0,7083	43,979	16,779	2,012
10	0,7500	43,987	15,801	1,889
11	0,7500	44,005	16,261	1,941
12	0,7500	44,015	16,589	1,983

**Cuadro C.17** – Serie 5; MAltDB; AR(1).

$h$	COB	EST	DESV	DESV %
01	0,7916	7,465	3,168	2,472
02	0,7916	7,819	2,987	2,275
03	0,7916	7,854	2,940	2,226
04	0,7916	7,858	3,042	2,286
05	0,7916	7,859	3,076	2,311
06	0,7916	7,859	2,986	2,216
07	0,7083	7,859	3,172	2,335
08	0,7083	7,859	3,248	2,386
09	0,6666	7,859	3,377	2,456
10	0,7083	7,859	3,208	2,304
11	0,7083	7,859	3,109	2,224
12	0,6666	7,859	3,277	2,329

**Cuadro C.18** – Serie 6; MAltDB; NID(0,  $\sigma^2$ ).

$h$	COB	EST	DESV	DESV %
01	0,8333	174,473	50,873	6,944
02	0,8750	174,056	48,463	6,607
03	0,8750	173,638	49,936	6,735
04	0,8750	173,220	49,243	6,657
05	0,8750	172,800	49,563	6,648
06	0,8750	172,379	46,827	6,295
07	0,8750	171,966	47,070	6,264
08	0,8750	171,556	45,063	5,917
09	0,8750	171,147	43,626	5,453
10	0,8750	170,738	44,068	5,504
11	0,8750	170,335	44,086	5,494
12	0,8750	169,931	43,743	5,433

---

---

# APÉNDICE D

---

## Resultados obtenidos al pronosticar series con información escasa

A continuación se presentan los resultados del experimento con el segundo conjunto de datos en el que la cantidad de observaciones es escasa. Los lineamientos observados así como el significado de la descripción de cada cuadro de este apéndice se encuentran en la sección 8.1 (página 101). El significado de los estadísticos COB, EST, DESV y DESV % se encuentra en la sección 7.1 (página 87).

### D.1. Resultados obtenidos con el MLIN

**Cuadro D.1** – Serie 1; MLIN; NID(0,  $\sigma^2$ ).

<i>h</i>	COB	EST	DESV	DESV %
1	0,9583	358,907	68,787	8,679
2	0,9583	353,025	70,940	8,900
3	0,9583	347,603	75,506	9,412
4	0,9583	342,605	76,991	9,507
5	0,9583	337,993	79,518	9,791
6	0,9583	333,730	83,285	10,228
7	0,9583	329,778	81,723	9,973
8	0,9583	326,104	80,685	9,737
9	0,9583	322,680	77,233	9,204
10	0,9583	319,477	73,561	8,728
11	0,9583	316,475	71,896	8,516
12	0,9583	313,651	69,251	8,202

**Cuadro D.2** – Serie 2; MLIN; NID(0,  $\sigma^2$ ).

<i>h</i>	COB	EST	DESV	DESV %
1	0,8333	13297,0	3849,4	26,588
2	0,8333	13094,0	4009,4	27,149
3	0,8750	12906,0	4041,8	26,985
4	0,8333	12732,0	4066,7	26,548
5	0,7916	12572,0	4091,7	26,302
6	0,8333	12423,0	4122,4	26,189
7	0,7916	12284,0	4101,8	25,515
8	0,7916	12155,0	3917,8	23,960
9	0,7916	12034,0	3814,4	22,755
10	0,7916	11920,0	3856,1	22,845
11	0,7916	11814,0	3681,6	21,419
12	0,7916	11714,0	3427,2	19,760

**Cuadro D.3** – Serie 3; MLIN; NID(0,  $\sigma^2$ ).

<i>h</i>	COB	EST	DESV	DESV %
1	0,9583	310,857	57,987	28,715
2	0,8750	305,715	65,155	30,422
3	0,8750	300,976	67,989	30,495
4	0,8750	296,611	69,811	29,985
5	0,9166	292,585	70,836	28,928
6	0,9166	288,865	70,888	28,262
7	0,9166	285,420	68,875	26,821
8	0,9583	282,220	68,885	26,652
9	0,9583	279,239	69,210	26,383
10	0,9583	276,454	68,544	26,063
11	0,9583	273,843	68,989	26,020
12	0,9166	271,388	72,521	26,598

**Cuadro D.4** – Serie 4; MLIN; NID(0,  $\sigma^2$ ).

<i>h</i>	COB	EST	DESV	DESV %
1	1,0000	344,512	56,519	8,689
2	1,0000	338,813	55,684	8,466
3	1,0000	333,561	56,080	8,513
4	1,0000	328,723	55,727	8,447
5	1,0000	324,262	55,418	8,384
6	1,0000	320,137	53,219	8,031
7	1,0000	316,313	49,855	7,524
8	1,0000	312,757	47,217	7,093
9	1,0000	309,443	46,189	6,918
10	1,0000	306,345	46,308	6,940
11	1,0000	303,441	46,576	6,916
12	1,0000	300,712	47,685	7,059

**Cuadro D.5** – Serie 5; MLIN; NID(0,  $\sigma^2$ ).

<i>h</i>	COB	EST	DESV	DESV %
1	0,9583	38,967	6,233	5,102
2	0,9583	38,321	6,536	5,373
3	0,9583	37,725	6,448	5,277
4	0,9583	37,175	6,416	5,244
5	0,9583	36,669	6,537	5,320
6	0,9583	36,201	6,167	4,982
7	0,9583	35,768	6,092	4,905
8	0,9583	35,365	6,222	4,992
9	0,9583	34,990	6,496	5,206
10	0,9583	34,640	6,444	5,141
11	0,9583	34,311	6,376	5,066
12	0,9583	34,002	6,142	4,847

**Cuadro D.6** – Serie 6; MLIN; NID(0,  $\sigma^2$ ).

<i>h</i>	COB	EST	DESV	DESV %
1	0,9166	485,708	73,758	17,513
2	0,9166	477,781	73,431	17,512
3	0,9166	470,475	71,421	17,120
4	0,9166	463,741	72,975	17,321
5	0,9166	457,527	74,070	17,477
6	0,9166	451,774	75,711	17,584
7	0,9166	446,432	70,620	16,714
8	0,9166	441,463	65,448	14,678
9	0,9166	436,826	63,306	14,360
10	0,9166	432,490	60,714	14,025
11	0,9166	428,422	59,234	13,867
12	0,9583	424,597	59,788	13,939

## D.2. Resultados obtenidos con el MNoLIN

**Cuadro D.7** – Serie 1; MNoLIN; NID(0,  $\sigma^2$ ).

$h$	COB	EST	DESV	DESV %
1	1,0000	246,081	59,089	7,436
2	0,9583	245,249	59,650	7,391
3	0,9583	244,435	60,421	7,401
4	0,9583	243,650	62,106	7,550
5	0,9583	242,920	61,826	7,480
6	0,9166	242,181	65,321	7,918
7	0,9166	241,503	66,016	7,946
8	0,9166	240,866	67,139	7,999
9	0,9166	240,289	68,382	8,134
10	0,8750	239,796	66,295	7,847
11	0,8750	239,502	66,844	7,861
12	0,8750	239,332	63,592	7,432

**Cuadro D.8** – Serie 2; MNoLIN; NID(0,  $\sigma^2$ ).

$h$	COB	EST	DESV	DESV %
1	0,7916	9902,10	2994,5	20,89
2	0,7500	9816,40	3246,7	21,90
3	0,7083	9748,00	3406,9	22,51
4	0,6666	9685,20	3572,1	23,05
5	0,6666	9628,70	3831,6	24,23
6	0,5833	9577,60	3973,9	24,86
7	0,5416	9530,10	4022,2	24,60
8	0,5833	9486,20	3956,8	23,75
9	0,6666	9446,70	3816,6	22,43
10	0,6666	9410,30	3788,3	22,15
11	0,6666	9377,60	3682,3	21,13
12	0,6250	9349,40	3648,2	20,37

**Cuadro D.9** – Serie 3; MNoLIN; NID(0,  $\sigma^2$ ).

$h$	COB	EST	DESV	DESV %
1	0,7500	146,320	38,419	14,10
2	0,7083	145,758	46,717	15,53
3	0,6666	145,195	52,693	17,97
4	0,5833	144,634	56,624	19,99
5	0,6250	144,083	59,339	21,18
6	0,6666	143,551	60,255	21,51
7	0,6666	143,043	59,031	20,39
8	0,6666	142,560	58,845	20,12
9	0,6666	142,108	60,423	20,31
10	0,6666	141,694	59,691	20,02
11	0,7083	141,330	59,020	19,82
12	0,6250	141,031	64,834	20,62

**Cuadro D.10** – Serie 4; MNoLIN; NID(0,  $\sigma^2$ ).

$h$	COB	EST	DESV	DESV %
1	0,9583	41,749	7,762	1,231
2	0,9583	41,667	8,538	1,372
3	0,9166	41,590	8,034	1,275
4	0,9166	41,518	8,047	1,273
5	0,9166	41,454	8,419	1,326
6	0,9583	41,397	8,707	1,362
7	0,9583	41,347	8,812	1,377
8	0,9583	41,305	8,786	1,370
9	0,9583	41,274	8,432	1,301
10	0,9583	41,257	8,644	1,330
11	0,9583	41,260	8,779	1,350
12	0,9166	41,289	9,473	1,442

**Cuadro D.11** – Serie 5; MNoLIN; NID(0,  $\sigma^2$ ).

$h$	COB	EST	DESV	DESV %
1	0,9583	11,930	3,171	2,756
2	0,9166	11,907	3,226	2,776
3	0,9166	11,886	3,262	2,794
4	0,9166	11,867	3,350	2,869
5	0,8750	11,848	3,551	3,032
6	0,7916	11,831	3,471	2,955
7	0,8333	11,817	3,475	2,954
8	0,8333	11,805	3,602	3,060
9	0,8333	11,796	3,739	3,173
10	0,7916	11,791	3,823	3,234
11	0,7083	11,792	3,969	3,342
12	0,8333	11,800	4,206	3,559

**Cuadro D.12** – Serie 6; MNoLIN; NID(0,  $\sigma^2$ ).

$h$	COB	EST	DESV	DESV %
1	0,8333	137,849	33,117	5,672
2	0,8333	137,499	31,114	4,953
3	0,8333	137,179	30,685	4,929
4	0,8333	136,889	29,844	4,780
5	0,8750	136,613	29,888	4,784
6	0,8333	136,357	34,124	5,522
7	0,8750	136,135	32,154	5,201
8	0,8750	135,937	32,802	5,179
9	0,8750	135,780	31,756	5,006
10	0,8750	135,657	30,393	4,792
11	0,9166	135,573	29,157	4,600
12	0,9166	135,545	30,701	4,834

### D.3. Resultados obtenidos con el MALtDB

**Cuadro D.13** – Serie 1; MALtDB; NID(0,  $\sigma^2$ ).

$h$	COB	EST	DESV	DESV %
1	0,9166	459,593	78,951	9,818
2	0,9166	453,104	80,180	9,878
3	0,9166	446,368	75,555	9,276
4	0,9166	439,560	75,097	9,114
5	0,9166	432,439	75,534	9,141
6	0,9166	424,958	75,326	9,024
7	0,9166	417,054	77,576	9,268
8	0,9583	408,916	76,055	8,998
9	0,9583	400,282	77,138	9,107
10	0,9583	391,339	76,722	9,019
11	1,0000	381,612	75,951	8,916
12	1,0000	369,228	70,972	8,319

**Cuadro D.14** – Serie 2; MALtDB; NID(0,  $\sigma^2$ ).

$h$	COB	EST	DESV	DESV %
1	0,8750	15994,0	3666,6	24,090
2	0,8750	15750,0	3817,9	24,713
3	0,8750	15514,0	3834,1	24,554
4	0,8750	15272,0	3891,5	24,432
5	0,8333	15029,0	3981,2	24,791
6	0,8333	14805,0	4079,0	25,213
7	0,7916	14572,0	4206,2	25,719
8	0,8333	14327,0	4085,7	24,528
9	0,8333	14065,0	3939,3	23,194
10	0,8333	13786,0	3979,8	23,335
11	0,8750	13475,0	3811,9	21,912
12	0,8750	13085,0	3554,8	20,163

**Cuadro D.15** – Serie 3; MALtDB; NID(0,  $\sigma^2$ ).

$h$	COB	EST	DESV	DESV %
1	0,7916	190,338	60,131	26,219
2	0,7083	187,209	64,423	26,210
3	0,6666	184,029	66,881	27,030
4	0,6666	180,815	68,006	26,865
5	0,6666	177,529	67,319	24,971
6	0,6666	174,201	67,194	24,431
7	0,7083	171,637	63,027	21,289
8	0,7083	169,002	63,361	21,067
9	0,7500	166,565	63,798	21,047
10	0,7083	164,087	63,173	20,826
11	0,6666	161,446	61,840	20,542
12	0,7083	159,363	59,236	19,134

**Cuadro D.16** – Serie 4; MALtDB; NID(0,  $\sigma^2$ ).

$h$	COB	EST	DESV	DESV %
1	1,0000	62,946	12,416	2,013
2	1,0000	61,899	11,354	1,806
3	1,0000	60,839	10,645	1,671
4	1,0000	59,829	10,750	1,686
5	1,0000	58,803	11,312	1,771
6	1,0000	57,757	11,291	1,767
7	1,0000	56,748	11,041	1,729
8	1,0000	55,671	11,395	1,767
9	1,0000	54,495	11,187	1,730
10	1,0000	53,173	10,562	1,619
11	1,0000	51,765	10,166	1,539
12	1,0000	50,653	10,208	1,542

**Cuadro D.17** – Serie 5; MALtDB; NID(0,  $\sigma^2$ ).

$h$	COB	EST	DESV	DESV %
1	0,9583	22,081	4,568	3,992
2	0,9583	21,739	4,125	3,518
3	0,9583	21,393	4,100	3,489
4	1,0000	21,035	4,064	3,457
5	1,0000	20,665	4,150	3,511
6	0,9583	20,283	3,833	3,230
7	0,9583	19,876	3,810	3,208
8	1,0000	19,483	3,889	3,265
9	0,9583	19,082	4,062	3,392
10	0,9166	18,640	4,077	3,394
11	0,9166	18,126	3,935	3,270
12	0,9583	17,784	3,981	3,298

**Cuadro D.18** – Serie 6; MALtDB; NID(0,  $\sigma^2$ ).

$h$	COB	EST	DESV	DESV %
1	0,9166	200,377	50,194	7,562
2	0,9166	196,509	49,502	7,453
3	0,9166	192,647	48,508	7,300
4	0,8750	189,157	47,059	7,084
5	0,8750	185,594	46,327	6,981
6	0,8333	181,844	47,865	7,179
7	0,8750	178,892	43,526	6,478
8	0,8750	175,838	43,040	6,273
9	0,8750	173,025	42,483	6,160
10	0,9166	170,161	40,087	5,841
11	0,7916	167,311	43,126	6,288
12	0,7916	163,861	43,711	6,352

---

---

# APÉNDICE E

---

## Resultados de pronosticar series con información faltante

En este apéndice, se muestran los resultados obtenidos al pronosticar series con información faltante. Los lineamientos observados así como la descripción de cada cuadro se explican en la sección 9.2 (página 112).

**Cuadro E.1** –  $\delta(\text{MDL}, \text{SR} = 1, f); \forall \text{MDL}; \forall f.$

$f$	MLIN	MNoLIN	MAltDB
0	17,90	15,08	17,30
1	18,69	15,40	19,84
2	20,55	16,52	20,43
3	18,87	14,93	20,05
4	18,96	15,93	19,07
5	19,31	15,22	20,50
6	19,99	16,60	20,95
7	20,18	15,92	20,22
8	19,32	16,11	20,54
9	19,20	15,72	20,52
10	19,20	15,35	20,08
11	20,24	16,11	20,77
12	19,47	15,78	20,08
13	20,11	15,96	20,19
14	19,76	16,31	20,37
15	19,51	15,86	20,55
16	20,53	16,42	20,51
17	18,82	15,43	19,72
18	20,83	16,51	20,41

**Cuadro E.2** –  $\delta(\text{MDL}, \text{SR} = 2, f); \forall \text{MDL}; \forall f.$

$f$	MLIN	MNoLIN	MAltDB
0	30,52	31,41	33,43
1	30,59	31,44	33,78
2	31,28	32,40	34,42
3	31,21	31,73	34,82
4	32,08	32,82	35,33
5	31,26	31,80	34,60
6	30,27	30,98	33,57
7	31,14	31,49	34,61
8	30,88	31,82	33,96
9	31,05	31,82	33,95
10	31,20	32,00	34,38
11	31,17	32,35	34,15
12	31,04	31,77	34,23
13	31,06	31,94	34,70
14	31,38	32,02	34,37
15	30,76	31,38	33,87
16	30,66	31,37	33,97
17	31,15	31,93	34,35
18	30,96	31,93	33,86

**Cuadro E.3** –  $\delta(\text{MDL}, \text{SR} = 3, f); \forall \text{MDL}; \forall f.$

$f$	MLIN	MNoLIN	MAltDB
0	56,23	43,94	51,37
1	69,04	48,76	62,50
2	64,56	45,24	60,27
3	68,17	46,01	59,13
4	68,52	48,05	58,40
5	64,95	47,70	60,11
6	68,19	46,51	59,42
7	64,86	47,72	57,90
8	63,12	45,42	56,05
9	67,82	46,91	62,11
10	67,24	47,07	65,79
11	69,10	47,42	62,45
12	64,69	47,05	60,70
13	69,90	48,57	62,88
14	64,11	47,07	62,49
15	62,35	46,83	57,84
16	68,53	47,17	59,54
17	68,38	47,06	60,53
18	67,95	46,93	57,51

**Cuadro E.4** –  $\delta(\text{MDL}, \text{SR} = 4, f); \forall \text{MDL}; \forall f.$

$f$	MLIN	MNoLIN	MAltDB
0	16,04	5,77	7,03
1	16,31	5,82	7,21
2	16,10	5,83	7,09
3	16,16	5,78	7,20
4	16,22	5,80	7,25
5	16,27	5,85	7,25
6	16,16	5,90	7,08
7	16,29	5,86	7,13
8	16,07	5,85	7,16
9	16,24	5,79	7,20
10	16,24	5,83	7,20
11	16,19	5,83	7,15
12	16,10	5,87	7,22
13	16,05	5,81	7,04
14	16,17	5,85	7,20
15	16,05	5,87	7,24
16	16,24	5,83	7,13
17	15,98	5,84	7,15
18	16,25	5,81	7,16

**Cuadro E.5** –  $\delta(\text{MDL}, \text{SR} = 5, f); \forall \text{MDL}; \forall f.$

$f$	MLIN	MNoLIN	MAltDB
0	12,63	6,31	8,05
1	14,11	6,52	9,77
2	13,38	6,58	9,06
3	13,45	6,76	9,44
4	13,66	6,54	8,83
5	13,23	6,49	9,07
6	13,63	6,50	9,61
7	14,07	6,67	9,99
8	13,40	6,61	8,88
9	13,49	6,37	9,37
10	13,91	6,59	9,87
11	13,66	6,77	8,98
12	13,26	6,38	8,97
13	13,95	6,54	9,70
14	13,26	6,41	9,15
15	14,09	6,41	9,23
16	14,02	6,73	9,33
17	13,90	6,56	9,37
18	13,62	6,20	8,41

**Cuadro E.6** –  $\delta(\text{MDL}, \text{SR} = 6, f); \forall \text{MDL}; \forall f.$

$f$	MLIN	MNoLIN	MAltDB
0	38,08	18,79	21,43
1	38,01	19,22	22,02
2	41,76	19,21	21,82
3	39,47	19,16	22,59
4	39,13	19,82	22,90
5	38,88	19,03	22,29
6	39,13	19,32	22,77
7	38,01	19,47	22,72
8	39,78	19,23	22,49
9	38,81	19,02	22,53
10	38,49	19,38	22,04
11	39,03	19,61	22,80
12	38,58	18,98	21,81
13	39,16	19,60	22,70
14	39,03	19,16	22,15
15	37,72	19,20	21,86
16	39,80	19,15	21,56
17	38,97	19,51	22,59
18	42,16	19,19	21,79

---

# BIBLIOGRAFÍA

- Akaike, H., 1973. Discussion of paper by Dr. Chatfield and Dr. Prothero. *Journal of the Royal Statistical Society A* 136 Part 3, 330–331.
- Anderson, O., Nochmals, U., 1914. The elimination of spurious correlation due to position in time or space. *Biometrika* 10, 269–276.
- Birchenhall, C. R., Blanden-Hovell, R. C., Chui, A. P. L., Osborn, D. R., Smith, J. P., 1989. A Seasonal Model of Consumption. *Economic Journal* 99, 837–843.
- Bowerman, B. L., Koehler, A. B., Pack, D. J., 1990. Forecasting time series with increasing seasonal time variation. *Journal of Forecasting* 9, 419–436.
- Bowerman, B. L., O’Connell, R., 1993. Time Series Forecasting: Unified Concepts and Computer Implementation, 3ª edición. Duxbury Press, Boston.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., 2008. Time series analysis: *Forecasting and Control*, 4ª edición. WILEY.
- Brown, R. G., 1959. *Statistical forecasting for inventory control*. New York: McGraw-Hill.
- Chatfield, C., Prothero, D. L., 1973. Box Jenkins Seasonal Forecasting: Problems in a Case Study (with discussion). *Journal of The Royal Statistical Society A* 136, 295–336.
- Cleveland, W. S., 1983. Seasonal and calendar adjustment, in *Handbook of Statistics*. Vol. 3. Elsevier Science Publishers B.V.

- Dagum, E. B., 1982. Revisions of time varying seasonal filters. *Journal of Forecasting* 1, 20–28.
- Flores, R., Novales, A., 1997. A General Test for Univariate Seasonality. *Journal of Time Series Analysis* 18, 29–48.
- Franses, P., 1996. Recent Advances in Modeling Seasonality. *Journal of Economic Surveys* 10, 299–345.
- Franses, P. H., 1993. A Method to Select Between Periodic Cointegration and Seasonal Cointegration. *Economics Letters* 41, 7–10.
- Franses, P. H., Koehler, A. B., 1998. A Model Selection Strategy for Time Series with Increasing Seasonality Variation. *International Journal of Forecasting* 14, 405–414.
- Gardner, J. E. S., 2006. Exponential smoothing: the state of the art—Part II. *International Journal of Forecasting* 22, 637–666.
- Gersovitz, M., MacKinnon, J. G., 1978. Seasonality in Regression: An Application of Smoothness Priors. *Journal of the American Statistical Association* 73, 264–273.
- Ghysels, E., 1991. Are Business Cycle Turning Points Uniformly Distributed Throughout the Year? Cahiers de recherche 9135, Université de Montréal, Département de sciences économiques.
- URL <http://ideas.repec.org/p/mtl/montde/9135.html>
- Ghysels, E., 1994. On the Periodic Structure of the Business Cycle. *Journal of Business and Economic Statistics* 12, 289–293.
- Hanke, J. E., Reitsh, A. G., 1996. Business Forecasting (spanish edition), 5<sup>a</sup> edición. Prentice Hall.

- Hansen, L. P., Sargent, T. J., 1993. Seasonality and Approximation Errors in Rational Expectation Models. *Journal of Econometrics* 55, 21–56.
- Herwartz, H., 1997. Performance of Periodic Error Correction Models in Forecasting Consumption Data. *International Journal of Forecasting* 13, 421–431.
- Holt, C. C., 2004a. Autor's retrospective on *Forecasting seasonals and trends by exponentially weighted averages*. *International Journal of Forecasting* 20, 11–13.
- Holt, C. C., 2004b. Forecasting seasonals and trends by exponentially weighted averages. *International Journal of Forecasting* 20, 5–10.
- Hooker, R. H., 1901. The suspension of the Berlin produce exchange and its effect upon corn prices. *Journal of the Royal Statistical Society* 64, 574–603.
- Hyndman, R. J., Akram, M., 2006. Some Nonlinear Exponential Smoothing Models are Unstable. *Monash Econometrics and Business Statistics Working Papers* 3/6.
- Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, J., 2005. Prediction intervals for exponential smoothing state space models. *International Journal of Forecasting* 24, 17–37.
- Hyndman, R. J., Koehler, A. B., Ord, J. K., Snyder, R. D., 2002. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting* 18, 439–454.
- INEGI, 2000. Volumen de las ventas mensuales al mayoreo de automóviles según categoría: subcompactos. *La Industria Automotriz en México*.
- INEGI, 2001. Volumen de las ventas mensuales al mayoreo de automóviles según categoría: subcompactos. *La Industria Automotriz en México*.

- INEGI, 2005. Volumen de las ventas mensuales al mayoreo de automóviles según categoría: subcompactos. *La Industria Automotriz en México*.
- Koehler, A. B., Snyder, R. D., Ord, J. K., 2001. Forecasting models and prediction intervals for the multiplicative Holt-Winters method. *International Journal of Forecasting* 17, 269–286.
- Makridakis, S., Andersen, A., Carbone, R., 1982. The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Journal of Forecasting* 1, 111–153.
- Makridakis, S., Wheelwright, S. C., Hyndman, R. J., 1998. *Forecasting Methods and Applications*, 3ª edición. John Wiley.
- Marsden, J., Tromba, A., 1998. *Cálculo vectorial*, 4ª edición. Addison Wesley Longman, New York.
- Mcleod, A. I., 1993. Model Adequacy and Periodic Correlation In Time Series Forecasting. *International Statistical Review* 61, 387–393.
- Novalés, A., de Fruto, R. D., 1997. Forecasting with periodic models: A comparison with the time invariant coefficient models. *International Journal of Forecasting*.
- Ord, J. K., Koehler, A. B., Snyder, R. D., 1997. Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of American Statistical Association* 92, 1621–1629.
- Osborn, D. R., 1988. Seasonality and Habit Persistence in a Life-Cycle Model of Consumption. *Journal of Applied Econometrics* 3, 255–266.
- Pagano, M., 1978. On Periodic and Multiple Autoregression. *Annals of Statistics* 6, 1310–1317.

- Pegels, C. C., 1969. Exponential smoothing: some new variations. *Management Science* 12, 311–315.
- Poynting, J. H., 1884. A comparison of the fluctuations in the price of wheat and in the cotton and silk imports into Great Britain. *Journal of the Royal Statistical Society* 47, 345–364.
- Priestly, G. T., 1973. Discussion of paper by Chatfield and Prothero. *Journal of the Royal Statistical Society A* 136 Part 3, 331–332.
- Rasmussen, R., 2004. On time series data and optimal parameters. *International Journal of Management Science* 32, 111–120.
- Shiskin, J., 1957. Electronic computers and business indicators. *National Bureau of Economic Research Occasional Paper*.
- Spencer, J., 1904. On the graduation of the rates of sickness and mortality. *Journal of the Institute of Actuaries* 38, 334.
- Todd, R., 1990. Periodic Linear-Quadratic Methods for Modeling Seasonality. *Journal of Economic Dynamics and Control* 14, 763–795.
- Troutman, B. M., 1979. Some Results in Periodic Autoregression. *Biometrika* 66, 219–228.
- Vencchia, R. L., Ballerini, R., 1991. Testing for Periodic Autocorrelations in Seasonal Time Series Data. *Biometrika* 78, 53–63.
- Wells, J. M., 1997. Modelling seasonal patterns and long-run trends in u.s. time series. *International Journal of Forecasting*.
- Wilson, G. T., 1973. Discussion of paper by Chatfield and Prothero. *Journal of the Royal Statistical Society A* 136 Part 3, 315–319.

---

Winters, P. R., 1960. Forecasting sales by exponentially weighted moving averages. *Management Science* 6, 324–342.

Wooldridge, J. M., 2001. *Introducción a la Econometría*. Thomson Learning.

---

# Biografía



Nací en el municipio de Guasave en el estado de Sinaloa, siendo mis padres el Profesor Roberto Madrigal Larrondo y la Profesora María del Refugio Espinoza Chávez. Estudié Ingeniería Industrial con Especialidad en Calidad y Productividad en el Instituto Tecnológico de Culiacán y me gradué bajo la opción de promedio. Obtuve el grado de Maestro en Ciencias en Ingeniería de Sistemas con la tesis «Modelos de Espacios de Estados Subyacentes al Método Multiplicativo de Holt-Winters con Múltiple Estacionalidad». Como profesor en la Licenciatura en Actuaría de la Facultad de Ciencias Físico-Matemáticas, he impartido los cursos de Probabilidad, Estadística No Paramétrica, Regresión Lineal, Análisis de Series Temporales y Análisis de Supervivencia.