



UNIVERSIDAD DE SANTIAGO DE CHILE  
FACULTAD DE CIENCIA  
DEPARTAMENTO DE MATEMÁTICA Y CIENCIA DE LA COMPUTACIÓN

---

SEMINARIO DE TESIS

# Un modelo temporal para anticipar congestión en el HUAP

---

ANGÉLICA GARRIDO AGURTO

PROFESOR GUÍA:  
FELIPE ELORRIETA L.

COMISIÓN:  
CLAUDIO VARGAS R.  
CLAUDIA MATUS C.

SANTIAGO, CHILE  
2018

# Un modelo temporal para anticipar congestión en el HUAP.

*Trabajo de titulación presentado a la Facultad de Ciencia en cumplimiento parcial de los requisitos exigidos para optar al título de Ingeniera Estadística*

# Resumen

---

La congestión en los Servicios de Urgencia, también conocida en inglés como *overcrowding*, es un problema que se presenta por lo menos hace dos décadas en nuestro país. Este fenómeno también se presenta en países desarrollados y ha sido ampliamente investigado en la literatura en el último tiempo. En Chile esta problemática no ha sido estudiada en profundidad. Es por esto que en este trabajo se estudia la congestión en el Hospital de Urgencia Asistencia Pública (HUAP) y se buscan formas de poder anticiparla a través de series temporales.

Para medir la congestión se calculó el índice internacional EDWIN para cada hora, durante dos años, junto con diferentes variables que pueden explicar los momentos de congestión y ser útiles para anticipar dichos momentos.

La selección de variables independientes fue a partir de un análisis de correlaciones respecto del índice, el cuál indica que las variables cantidad de pacientes en espera de atención, hospitalizados y que abandonan el Servicio de Urgencia, son variables que pueden explicar la congestión.

Para poder anticipar periodos de congestión, se utilizaron modelos de series temporales, desde el punto de vista bayesiano, como los modelos lineales dinámicos (MLD) y también desde el ámbito más clásico a través de modelos que permiten la intervención de variables exógenas, como los ARIMAX.

Debido a que la estimación de parámetros en los MLD exige implementar diferentes algoritmos complejos como el *Filtro de Kalman* o el *Algoritmo Gibbs Sampling*, sumado a la gran cantidad de datos en la serie de tiempo y las exigencias computacionales que la estimación implica, estos modelos no presentaron resultados que pudieran ser confiables.

Es por lo anterior que se decidió ajustar 8 modelos ARIMAX, de los cuáles 4 incluyen una componente estacional, con periodo de 24 horas. De estos 8, los modelos ARIMAX(5,1,1), tanto el que incluye a la variable *Cantidad de pacientes en espera* rezagada en una hora (modelo 3), como el que tiene las 3 variables exógenas rezagas (modelo 4), fueron los que resultaron ser mejores. Por su parte el modelo 3 presentó un menor AIC y se obtuvo un mejor ajuste que con el modelo 4, sin embargo este último entregó un menor MAPE y por ende mejores predicciones .

**Palabras claves:** Modelos lineales dinámicos, ARIMAX, Series temporales, Componente estacional, EDWIN, Rezago, Servicio de Urgencia.

*A mi papá, por ser mi pilar;  
lo logramos viejo!*

# Índice general

<b>1. Introducción</b>	<b>10</b>
1.1. Motivación	10
1.1.1. Servicios de Urgencia	10
1.1.2. Sobre el Hospital de Urgencia Asistencia Pública (HUAP)	10
1.1.3. Planteamiento del Problema	11
1.2. Objetivos	13
1.2.1. Objetivo General	13
1.2.2. Objetivos Específicos	13
1.3. Metodología	14
1.3.1. Características generales de la base de datos	14
1.3.2. Etapas de Desarrollo	14
<b>2. Marco Teórico</b>	<b>15</b>
2.1. Elementos de la Inferencia Bayesiana	15
2.2. Modelos Lineales Dinámicos	16
2.2.1. Estructura General	16
2.2.2. Análisis Inferencial	18
2.2.2.1. Filtro de Kalman	18
2.2.3. Suavizado	19
2.2.4. Algoritmo FFBS	20
2.2.5. Inferencia de los parámetros	20
2.2.6. Gibbs Sampling	22
2.2.7. Distribución de Pronóstico	23
2.2.8. Comparación de modelos	24
2.2.8.1. Criterio de Información de la Devianza (DIC)	24
2.3. Modelos clásicos de series de tiempo	25
2.3.1. Introducción	25
2.3.2. Componentes de una serie de tiempo	25
2.3.3. Conceptos básicos	26
2.3.4. Modelos Estacionarios	26
2.3.4.1. Procesos lineales	26
2.3.4.2. Modelo Auto-regresivo (AR)	27
2.3.4.3. Modelos de medias móviles (MA)	27
2.3.4.4. Modelo auto-regresivo de medias móviles (ARMA)	27
2.3.5. Metodología de Box-Jenkins	27
2.3.6. Modelos para series con tendencia	28
2.3.6.1. Modelo ARIMA	28
2.4. Modelos para series estacionales	29
2.4.0.1. Modelo SARIMA	29
2.5. Modelos con Variables Exógenas	29
2.5.1. Modelos ARIMAX	29

2.5.2.	Modelos ARIMAX Estacionales . . . . .	29
2.5.3.	Evaluación de Modelos . . . . .	30
2.6.	Escalas internacionales para medir Congestión . . . . .	30
2.6.1.	National Emergency Department Overcrowding Study (NEDOCS) . . . . .	30
2.6.2.	Emergency Department Work Index (EDWIN) . . . . .	31
<b>3.</b>	<b>Análisis Descriptivo</b>	<b>32</b>
3.1.	Descripción de la base de datos . . . . .	32
3.1.1.	Descripción de Variables . . . . .	32
3.1.2.	Filtros en la Base de Datos . . . . .	33
3.1.3.	Filtros por anomalías y valores atípicos . . . . .	34
3.2.	Análisis descriptivo de variables de interés . . . . .	35
3.3.	Análisis del Tiempo de Espera y Tiempo de Uso de Camilla . . . . .	36
<b>4.</b>	<b>Cálculo del Índice EDWIN</b>	<b>40</b>
4.1.	Obtención de la Serie de Tiempo . . . . .	40
4.2.	Cálculo del Índice EDWIN por hora . . . . .	41
4.2.1.	Análisis Descriptivo EDWIN . . . . .	42
<b>5.</b>	<b>Análisis de la serie de tiempo</b>	<b>46</b>
5.1.	Análisis de ACF EDWIN . . . . .	46
5.2.	Análisis de CCF para el Índice de Congestión . . . . .	47
5.2.1.	Cantidad de Pacientes en Espera de Atención . . . . .	47
5.2.2.	Tiempo Mediano Espera . . . . .	48
5.2.3.	Tiempo Mediano de Uso de Cama . . . . .	48
5.2.4.	Cantidad de Pacientes Hospitalizados . . . . .	49
5.2.5.	Cantidad de Pacientes Sin Diagnóstico . . . . .	49
<b>6.</b>	<b>Ajuste de Modelos Lineales Dinámicos</b>	<b>51</b>
6.1.	Estructura de los modelos propuestos . . . . .	51
6.2.	Estimación de parámetros . . . . .	54
<b>7.</b>	<b>Estimación de Modelos ARIMAX</b>	<b>61</b>
7.1.	Ajuste de Modelos ARIMAX . . . . .	61
7.2.	Ajuste de Modelos ARIMAX Estacionales . . . . .	65
7.3.	Predicciones a través del modelo ARIMAX 3 . . . . .	70
7.4.	Predicciones a través del modelo ARIMAX 4 . . . . .	73
<b>8.</b>	<b>Conclusiones</b>	<b>77</b>
<b>9.</b>	<b>Anexos</b>	<b>79</b>

# Índice de figuras

2.1. Esquema Box-Jenkins . . . . .	28
3.1. Boxplot tiempo de espera según Triage . . . . .	37
3.2. Boxplot tiempo de espera según diagnóstico . . . . .	37
3.3. Boxplot tiempo de atención según triage . . . . .	38
3.4. Boxplot tiempo de atención según destino . . . . .	38
3.5. Tiempos de Espera Total y Hospitalizados . . . . .	39
4.1. Base de datos inicial . . . . .	41
4.2. Base Serie de Tiempo por hora . . . . .	41
4.3. EDWIN por año . . . . .	43
4.4. EDWIN por Mes . . . . .	43
4.5. Boxplot EDWIN según día de la semana . . . . .	44
4.6. EDWIN por Hora . . . . .	45
5.1. Gráfica de la Serie de tiempo EDWIN . . . . .	46
5.2. Gráfica de la Serie de tiempo EDWIN . . . . .	47
5.3. Periodograma EDWIN . . . . .	47
5.4. Zoom periodograma EDWIN . . . . .	47
5.5. CCF Pacientes Espera Atención . . . . .	48
5.6. CCF Tiempo Mediano Espera . . . . .	48
5.7. CCF Tiempo Mediano de Uso de Cama . . . . .	49
5.8. CCF Pacientes hospitalizados . . . . .	49
5.9. CCF Pacientes sin diagnóstico . . . . .	50
6.1. $\beta_0$ Modelo 1 . . . . .	55
6.2. $\beta_1$ Modelo 1 . . . . .	55
6.3. $\beta_2$ Modelo 1 . . . . .	56
6.4. $\beta_3$ Modelo 1 . . . . .	56
6.5. $\phi_1$ Modelo 3 . . . . .	57
6.6. $\beta_0$ Modelo 3 . . . . .	57
6.7. $\beta_1$ Modelo 3 . . . . .	57
6.8. $\beta_2$ Modelo 3 . . . . .	57
6.9. $\beta_3$ Modelo 3 . . . . .	57
6.10. $\phi_1$ Modelo 4 . . . . .	58
6.11. $\phi_2$ Modelo 4 . . . . .	58
6.12. $\beta_0$ Modelo 4 . . . . .	58
6.13. $\beta_1$ Modelo 4 . . . . .	58
6.14. $\beta_2$ Modelo 4 . . . . .	58
6.15. $\beta_3$ Modelo 4 . . . . .	58
6.16. Ajuste Modelo 3 . . . . .	60
6.17. Zoom ajuste Modelo 3 . . . . .	60

7.1.	Diagnóstico residuos modelo ARIMAX 1 . . . . .	62
7.2.	Diagnóstico residuos modelo ARIMAX 2 . . . . .	63
7.3.	Diagnóstico residuos modelo ARIMAX 3 . . . . .	64
7.4.	Diagnóstico residuos modelo ARIMAX 4 . . . . .	65
7.5.	Diagnóstico residuos modelo ARIMAX 5 . . . . .	66
7.6.	Diagnóstico residuos modelo ARIMAX 6 . . . . .	67
7.7.	Diagnóstico residuos modelo ARIMAX 7 . . . . .	68
7.8.	Diagnóstico residuos modelo ARIMAX 8 . . . . .	69
7.9.	Ajuste modelo ARIMAX 3 . . . . .	70
7.10.	Ajuste modelo ARIMAX 3 inf . . . . .	70
7.11.	Ajuste modelo ARIMAX 3 sup . . . . .	70
7.12.	Predicción a 72 horas . . . . .	72
7.13.	Predicción a 1754 horas . . . . .	73
7.14.	Ajuste modelo ARIMAX 4 . . . . .	73
7.15.	Ajuste modelo ARIMAX 4 inf . . . . .	74
7.16.	Ajuste modelo ARIMAX 4 sup . . . . .	74
7.17.	Predicciones a 72 horas . . . . .	75
7.18.	Predicción a 1754 horas . . . . .	76
9.1.	Ajuste MLD 1 . . . . .	79
9.2.	Zoom Ajuste MLD 1 . . . . .	79
9.3.	$\beta_0$ MLD 1 4000 estimaciones eliminadas . . . . .	79
9.4.	$\beta_1$ MLD 1 4000 estimaciones eliminadas . . . . .	79
9.5.	$\beta_2$ MLD 1 4000 estimaciones eliminadas . . . . .	80
9.6.	$\beta_3$ MLD 1 4000 estimaciones eliminadas . . . . .	80
9.7.	Ajuste MLD 2 . . . . .	80
9.8.	Zoom Ajuste MLD 2 . . . . .	80
9.9.	$\beta_0$ MLD 2 . . . . .	81
9.10.	$\beta_1$ MLD 2 . . . . .	81
9.11.	$\beta_2$ MLD 2 . . . . .	81
9.12.	$\beta_3$ MLD 2 . . . . .	81
9.13.	Ajuste MLD 3 . . . . .	81
9.14.	Zoom Ajuste MLD 3 . . . . .	81
9.15.	Ajuste MLD 4 . . . . .	82
9.16.	Zoom Ajuste MLD 4 . . . . .	82

# Índice de cuadros

3.1. Consultas en cada área por año . . . . .	33
3.2. Resumen de tiempos de espera en horas por Triage . . . . .	34
3.3. Tiempos de espera en horas de pacientes sin diagnóstico . . . . .	34
3.4. Frecuencias de consultas por Día de la semana . . . . .	36
3.5. Frecuencias consultas por Triage . . . . .	36
3.6. Frecuencia destino de los pacientes después del alta . . . . .	36
3.7. Frecuencias consultas sin diagnóstico . . . . .	37
4.1. Estado Urgencia según EDWIN por día . . . . .	42
6.1. Estimación parámetros modelo 2 . . . . .	56
6.2. Estimación parámetros modelo 5 . . . . .	59
6.3. Información sobre la Devianza . . . . .	59
7.1. Estimación ARIMA(5,1,0) . . . . .	61
7.2. Estimación Modelo ARIMAX 1 . . . . .	62
7.3. Estimación Modelo ARIMAX 2 . . . . .	63
7.4. Estimación Modelo ARIMAX 3 . . . . .	63
7.5. Estimación Modelo ARIMAX 4 . . . . .	64
7.6. Estimación $SARIMA(5, 1, 0)(0, 0, 2)_{24}$ . . . . .	65
7.7. Estimación Modelo ARIMAX 5 . . . . .	66
7.8. Estimación Modelo ARIMAX 6 . . . . .	66
7.9. Estimación Modelo ARIMAX 7 . . . . .	67
7.10. Estimación Modelo ARIMAX 8 . . . . .	68
7.11. Resumen Modelos ARIMAX . . . . .	69
7.12. Comparación EDWIN Real y EDWIN estimado . . . . .	71
7.13. Predichos Mínimos y Máximos de congestión . . . . .	71
7.14. Predicción primeras 24 horas . . . . .	72
7.15. Comparación EDWIN Real y EDWIN estimado . . . . .	74
7.16. Predichos Mínimos y Máximos de congestión . . . . .	74
7.17. Predicción primeras 24 horas . . . . .	75
9.1. Estimación parámetros MLD 1 . . . . .	80
9.2. Estimación parámetros MLD 3 . . . . .	82
9.3. Estimación parámetros MLD 4 . . . . .	82
9.4. Predicciones 72 horas ARIMAX 3 . . . . .	83
9.5. Predicciones 72 horas ARIMAX 4 . . . . .	84

# Capítulo 1

## Introducción

### 1.1. Motivación

El aumento desmesurado en la demanda de atención en las urgencias tiene consecuencias negativas para el servicio, en forma de falta de espacio físico, sobrecarga tanto en los servicios que facilitan los diagnósticos, por ejemplo laboratorio y radiología, como de los profesionales del hospital, alargamiento de la lista de espera de ingresos programados y pérdidas en productividad por las dificultades para programar el trabajo y sobre todo, se pueden producir demoras en la asistencia a pacientes que sufren situaciones de riesgo vital. [2]

La congestión en los servicios de urgencia produce problemas, tanto a los pacientes, implicándoles tiempos de espera más largos, como al personal, quienes muchas veces se ven con una carga laboral elevada, debido a la alta demanda y la falta de profesionales disponibles en dichos momentos de congestión. [15]

#### 1.1.1. Servicios de Urgencia

Los servicios de urgencia componen una parte esencial del sistema de salud de cualquier país, este es el lugar donde se atiende a los usuarios en uno de los momentos de mayor fragilidad en su historia vital. [16]

Estos servicios funcionan las 24 horas del día y durante los 365 días del año, con el objetivo de realizar atenciones ante problemas de salud que afecten de manera urgente a las personas.

En un principio los pacientes son evaluados de tal manera de determinar su nivel de gravedad, el que está protocolizado en una escala de categorización que va desde el C1 que representa una urgencia vital, al C5 que indica que se trata consulta general. De esta manera, se prioriza a quienes necesitan de atención más urgente. [9]

Una consulta de urgencia puede definirse como una situación clínica que precisa de atención inmediata, por lo que el resultado final para la salud del paciente depende, no sólo de la naturaleza del servicio ofrecido, sino también, de la rapidez con que se preste tal servicio.

#### 1.1.2. Sobre el Hospital de Urgencia Asistencia Pública (HUAP)

El Hospital de Urgencia Asistencia Pública (HUAP) o también conocido como ex Posta Central, fue inaugurado por el Doctor Alejandro del Río Soto-Aguilar el 7 de Agosto de 1911 y

fue el primer servicio de urgencia permanente del país.

Ubicada en la comuna de Santiago, la ex Posta Central es el establecimiento más importante en atención de urgencias médicas de adultos de la Región Metropolitana, además de ser el punto de derivación nacional del sistema público de salud para pacientes politraumatizados y quemados.

La cantidad de pacientes atendidos en este servicio de urgencia ha aumentado gradualmente, pasando a atender desde alrededor de 59 mil pacientes al año, en el 2014 [10], a atender cerca de 70 mil usuarios durante el 2017 [4] .

### 1.1.3. Planteamiento del Problema

La congestión de los Servicios de Urgencia es un problema ampliamente investigado y analizado en la literatura en el último tiempo. La principal razón de este fenómeno se debe a lo universal de este problema, que afecta tanto a los países desarrollados como a los países en vías de desarrollo, como Chile y a las graves consecuencias que este puede provocar. Internacionalmente el problema de la congestión de los Servicios de Urgencia es conocido como *Overcrowding*.

La investigación se ha centrado en estudiar profundamente las causas de este problema, las principales consecuencias y las posibles medidas que se pueden tomar para mejorar este escenario que enfrentan muchos Servicios de Urgencia del mundo. [7]

Algunos estudios argumentan que la congestión podría reducirse considerablemente anticipando aumentos en la demanda de los servicios de urgencia e implementar mejoras operacionales que maximicen la capacidad del servicio antes de que dichos aumentos de demanda se produzcan. Sin embargo, los métodos para predecir la demanda (es decir, llegadas de pacientes al servicio de urgencias) no han sido bien estudiados. Varios investigadores han propuesto métodos como la regresión lineal o métodos de series de tiempo para pronosticar las llegadas a los servicios de urgencia. [12]

Así como es un problema de alto interés a resolver en los servicios de urgencia alrededor del mundo, las autoridades del HUAP también se ven preocupadas por buscar una solución que les permita tanto medir la congestión, como adelantarse a estos periodos. Es por esto que el propósito de este trabajo es principalmente dar una posible respuesta a esta problemática en la ex Posta Central, de tal forma que puedan implementar medidas de prevención y mejoras operacionales que les permita adelantarse a los periodos de congestión.

Así como se ha propuesto la implementación de métodos estadísticos para predecir el aumento en la demanda, en cuanto a las llegadas a los servicios de urgencia, también es posible plantear métodos de series temporales para predecir momentos de congestión.

De esta forma se pretende *implementar, a través de series temporales, un método que le permita a los encargados de la urgencia del HUAP medir la congestión y así mismo, que les permita anticiparse a dichos periodos de alta demanda.*

Particularmente, para esto se utilizarán *Modelos Lineales Dinámicos* [21] , desde el punto de vista de la teoría bayesiana, debido a que estos permiten la inclusión de variables (las que

pueden influir en la congestión). Variables que serán definidas una vez analizados los datos de forma preliminar.

De la misma forma se probará la implementación de modelos clásicos que permiten la influencia de variables exógenas en la congestión, como lo son los modelos ARIMAX.

## 1.2. Objetivos

### 1.2.1. Objetivo General

Proponer un modelo de series temporales que pueda anticipar periodos de congestión en el HUAP.

### 1.2.2. Objetivos Específicos

- Recopilar, analizar, verificar y preparar la base de datos de la urgencia del HUAP, de tal forma de obtener una serie de tiempo.
- Estudiar y evaluar metodologías para medir congestión en urgencias.
- Construir un índice que permita medir la congestión en el tiempo.
- Realizar un análisis descriptivo de la serie de tiempo.
- Evaluar y seleccionar variables de interés.
- Modelar la congestión a partir de Modelos Lineales Dinámicos.
- Modelar la congestión a partir de Modelos ARIMAX.
- Seleccionar y evaluar supuestos de los modelos ajustados.
- Predecir futuros periodos de congestión en el HUAP a partir de los MLD y ARIMAX.
- Evaluar predicciones obtenidas.

## 1.3. Metodología

### 1.3.1. Características generales de la base de datos

Se cuenta con información de todas las consultas realizadas al servicio de urgencias de la ex Posta Central en el periodo del 1 de Enero de 2016 al 31 de Diciembre de 2017. En total 244483 pacientes acudieron a este servicio.

### 1.3.2. Etapas de Desarrollo

1. Investigar en la literatura cómo medir congestión en los servicios de urgencia, evaluar si los métodos utilizados internacionalmente pueden ser aplicados a la realidad nacional y en particular al comportamiento en la ex Posta Central.
2. Así mismo será necesario investigar en la literatura qué métodos pueden ser utilizados para anticipar los periodos de congestión en urgencias y estudiar sobre los modelos de series temporales dinámicos y los modelos ARIMAX y cómo podrían ser útiles para resolver este problema.
3. Luego, será necesario analizar y preparar la base de datos de tal forma de obtener una serie de tiempo, es decir, es necesario calcular diferentes variables a partir de la base de datos original, en cada hora, para los 731 días contemplados en los dos años antes mencionados.
4. Una vez obtenida la base de datos como serie de tiempo y con ella las variables que serán útiles para la creación del índice EDWIN, se debe preceder al cálculo de dicho índice e interpretar según la escala de este, el estado de congestión del HUAP.
5. A partir del cálculo del índice EDWIN por hora, se debe realizar un análisis descriptivo de este, como serie de tiempo y también, obtener las correlaciones de las posibles variables que puedan influir en los valores del índice, de tal forma de determinar cuáles serán las variables que finalmente serán parte del modelo.
6. Una vez obtenido el análisis descriptivo, definidas de forma preliminar las variables que serán utilizadas y estimado el nivel de congestión según la escala EDWIN, se procederá a modelar la congestión a partir de Modelos Lineales Dinámicos, seguido del análisis de los supuestos de dichos modelos.
7. El análisis de los supuestos de los modelos obtenidos y los criterios estadísticos permitirán determinar cuál es el modelo más adecuado para que, a partir de este, sea posible realizar predicciones de periodos futuros de congestión en el HUAP.
8. De la misma forma que con los modelos dinámicos, se ajustará un modelo ARIMAX del índice EDWIN, a partir de las mismas variables exógenas definidas a través del análisis de correlaciones, para así predecir periodos de congestión en la ex Posta Central a partir del mejor modelo, determinado mediante el análisis de los supuestos y criterios estadísticos.

# Capítulo 2

## Marco Teórico

A continuación, se mostrarán algunos aspectos de la Estadística Bayesiana, cuya teoría es muy importante para este trabajo, pues se utilizarán los Modelos Lineales Dinámicos (MLD) desde el enfoque Bayesiano, para realizar las predicciones de congestión. Algunos de los conceptos acerca de los MLD son abordados por Harrison & Stevens (1976) [8], West & Harrison (1997) [21] y Gamerman (1998) [6]

### 2.1. Elementos de la Inferencia Bayesiana

En la inferencia Bayesiana a diferencia del análisis clásico, los parámetros se definen como variables aleatorias, por lo cuál se pueden describir mediante una distribución de probabilidad. Este análisis contempla conocimientos previos del investigador acerca del parámetro y estos se combinan con las información que se tiene sobre una muestra, de esta manera se puede actualizar la información que existe del parámetro.

Es posible derivar la base de la inferencia Bayesiana a partir de la teoría de probabilidad simple. De esta forma, el teorema probabilidad condicional de que ocurra A dado que ya ocurrió B es el siguiente:

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} \quad (2.1)$$

En este caso, al reemplazar B por las observaciones, A por el conjunto de parámetros  $\theta$  y las probabilidades por funciones de densidad, entonces se obtiene una relación entre la densidad a priori  $g_1(\theta)$  y posteriori  $h(\theta|x)$

#### **Teorema 2.1.** (*Bayes*)

Sean  $g_1(\theta)$ , la densidad a priori de  $\theta$  y  $f(x|\theta)$ , la densidad condicional de  $X$  dado  $\theta$ , entonces la densidad de  $\theta$  dado  $X = x$ , está dada por:

$$h(\theta|x) = \frac{f(x|\theta)g_1(\theta)}{\int_{-\infty}^{+\infty} f(x|\theta)g_1(\theta)d\theta} \quad (2.2)$$

*Demostración:* Sean  $z$  y  $g_2$ , las funciones de densidad de  $(X, \theta)$  y de  $X$  respectivamente. Por definición de densidad condicional se tiene:

$$f(x|\theta) = \frac{z(x, \theta)}{g_1(x)} \quad (2.3)$$

$$f(\theta|x) = \frac{z(x, \theta)}{g_2(x)} \quad (2.4)$$

Despejando  $z(x, \theta)$  de 2.3 y reemplazando en 2.4, se obtiene:

$$h(\theta|x) = \frac{f(x|\theta)g_1(\theta)}{g_2(\theta)} \quad (2.5)$$

por otra parte se tiene que:

$$g_2(x) = \int_{-\infty}^{+\infty} z(x, \theta)d\theta = \int_{-\infty}^{+\infty} f(x|\theta)g_1(\theta)d\theta \quad (2.6)$$

Luego, reemplazando en 2.5, se obtiene:

$$h(x|\theta) = \frac{f(x|\theta)g_1(\theta)}{\int_{-\infty}^{+\infty} f(x|\theta)g_1(\theta)d\theta} \quad (2.7)$$

El denominador de  $h(x|\theta)$  se define como la verosimilitud marginal y puede ser considerado como una constante desconocida  $c$ , que se utiliza para que la distribución a posteriori cumpla las condiciones necesarias para que se pueda definir como función de densidad. Entonces la inferencia sobre  $\theta$  puede ser realizada por la siguiente expresión equivalente:

$$h(x|\theta) \propto f(x|\theta)g_1(\theta) \quad (2.8)$$

## 2.2. Modelos Lineales Dinámicos

Los modelos lineales dinámicos son una herramienta útil para modelar series temporales debido al dinamismo y flexibilidad de sus parámetros, lo que implica tener predicciones mucho más certeras que las que se obtienen a partir de los modelos clásicos. El dinamismo de los parámetros es aportado por la información que puede aportar el pasado de la variable respuesta y de las respectivas covariables, produciendo estimaciones de parámetros en cada tiempo. A su vez, la flexibilidad de los parámetros se produce gracias a las nuevas tecnologías que permiten la implementación eficiente de los métodos Bayesianos para la estimación de los parámetros.

Los modelos lineales son capaces de explicar y estudiar el comportamiento del vector de observaciones en el tiempo  $Y_t$ , para luego formar la serie temporal predictora a  $k$  pasos  $Y_{t+k}$ ,  $k > 0$ .

### 2.2.1. Estructura General

Los modelos lineales dinámicos son considerados una generalización de los modelos de regresión lineal que permiten la evolución y cambios en los parámetros a través del tiempo, a diferencia de los parámetros fijos del modelo de regresión lineal. Es por esto que sus estructuras son bastantes parecidas.

La estructura de los MLD para la variable respuesta  $Y$ , en el tiempo  $t$ , se basa en las siguientes ecuaciones:

$$Y_t = F_t' \theta_t, \epsilon_t \sim N(0, V_t) \quad (2.9)$$

$$\theta_t = G_t \theta_{t-1} + w_t, w_t \sim N(0, W_t) \quad (2.10)$$

Donde,

- $F_t$ : Vector de variables explicativas de orden  $p \times 1$ , es decir, es el vector que contiene a  $X_1, X_2, \dots, X_p$ .
- $\theta_t$ : Vector de parámetros del modelo de dimensión  $p \times 1$ .
- $G_t$ : Matriz de dimensión  $p \times p$ , que describe la evolución de los parámetros  $\theta_t$  en el tiempo.

- $V_t$ : Vector de varianza de los errores asociados a las observaciones.
- $W_t$ : Matriz de varianza-covarianza de los errores asociados a los parámetros.

Las ecuaciones 2.9 y 2.10 son llamadas *ecuación de observaciones* y *ecuación de evolución*, respectivamente.

En términos generales entonces, los Modelos Lineales Dinámicos se definen como una estructura cuádruple  $\{F_t, G_t, V_t, W_t\}$ , para cada tiempo  $t$ .

Volviendo a la analogía con los modelos de regresión lineal, si se considera que no existe varianza a nivel de la ecuación de sistema, es decir  $W_t = 0, \forall t$  y que la matriz de transición  $G_t$  corresponde a una matriz identidad de orden  $p$ , entonces se está en presencia de un modelo de regresión lineal clásico, pues no existen cambios a través del tiempo. Esto se considera un caso particular de los MDL.

Un modelo lineal dinámico queda totalmente determinado a partir de la estructura general de las ecuaciones 2.9 y 2.10 y de la definición del punto inicial, es decir, la estimación para el tiempo  $t = 0$

Ecuación de Observación:

$$Y_t = F_t' \theta_t, \epsilon_t \sim N(0, V_t) \quad (2.11)$$

Ecuación de Evolución:

$$\theta_t = G_t \theta_{t-1} + w_t, w_t \sim N(0, W_t) \quad (2.12)$$

Información Inicial:

$$(\theta_0 | D_0) \sim N(m_0, C_0)$$

Donde la condicional  $D_0$  es la información que se conoce al momento del estudio,  $m_0$  es el vector de medias inicial y  $C_0$  es la matriz de varianzas-covarianzas inicial, las cuales se asumen conocidas.

La sucesión de errores  $\epsilon_t$  y  $w_t$  son independientes. También son independientes de  $(\theta_0 | D_0)$ .

Es posible definir las ecuaciones de observación y de evolución a través de su distribución a lo largo del tiempo  $t$ , de la siguiente forma:

$$Y_t | \theta_t, V_t, W_t \sim N(F_t \theta_t, V_t) \quad (2.13)$$

$$\theta_t | \theta_{t-1} \sim N_p(G_t \theta_{t-1}, W_t) \quad (2.14)$$

Es posible notar en las ecuaciones que implícitamente se considera el conjunto de información a priori al tiempo  $(t - 1)$ , es decir, se condiciona al conjunto  $D_{t-1}$ . En particular este almacena la información de los valores pasados de la serie  $Y_{t-1}, Y_{t-2}, \dots, Y_1$ , los valores de  $V_t$ , de la matriz  $W_t$  y de la información inicial  $D_0$ .

## 2.2.2. Análisis Inferencial

Una vez presentada la estructura de los modelos lineales dinámicos, es natural cuestionarse cuál será la próxima etapa que permita obtener resultados, es decir, qué se debe realizar para obtener predicciones para el tiempo  $t + k$ ,  $k > 0$ , a partir del conjunto de información antes mencionado. Para esto es necesario recurrir a la inferencia estadística, la cuál mediante una serie de pasos, definidos especialmente para este tipo de modelos permitirá responder este cuestionamiento.

### 2.2.2.1. Filtro de Kalman

El *Filtro de Kalman* es el primer ciclo que se debe realizar ante cualquier tipo de estructura de modelo, ya que, este consiste en conocer el comportamiento desde el principio de la serie hasta su fin y encontrar las distribuciones del vector de parámetros asociados y de la serie en estudio.

Para desarrollar este filtro, es necesario recordar al conjunto  $D_0$ , definido anteriormente, el cuál contiene toda la información inicial existente, vale decir, en  $t = 0$ . Análogo a esto, ahora es posible denominar  $D_t$  al conjunto que posee la información hasta un determinado tiempo  $t$ .

El Filtro de Kalman es un esquema de inferencia secuencial a través del tiempo, el cual consta de tres opciones, las que se denominan *evolución*, *predicción* y *actualización*, las que son de gran utilidad para lograr el objetivo de conocer la serie y su distribución.

El primer paso corresponde a *evolución*, este consiste en el seguimiento de pasar de un tiempo a otro, pero sin conocer la información al tiempo anterior. Esto, proviene de la distribución a posteriori del vector de estado con la información disponible a  $t - 1$ , es decir,

$$\theta_{t-1}|D_{t-1} \rightarrow \theta_t|D_{t-1}$$

El segundo paso corresponde a *predicción*, este consiste en obtener la cadena de la variable respuesta, a partir de la evolución del vector  $\theta$  y del conjunto de información al tiempo anterior, es decir,

$$\theta_t|D_{t-1} \rightarrow Y_t|D_{t-1}$$

El tercer paso corresponde a *actualización*, proviene de la evolución del vector de estado, el cual sirve para actualizar el vector dada la información ya actualizada, es decir, se obtiene la estimación de  $\theta$  actualizado,

$$\theta_t|D_{t-1} \rightarrow \theta_t|D_t$$

Las distribuciones asociadas a cada uno de los pasos del esquema secuencial, antes descrito, se presentan a continuación. Además es posible visualizar la forma en que cada componente del modelo se actualiza a través del tiempo, o equivalentemente, a medida que aumenta la información.

Recordando la estructura general del modelo dinámico,

$$Y_t = F_t' \theta_t + \epsilon_t, \epsilon_t \sim N(0, V_t)$$

$$\theta_t = G_t' \theta_{t-1} + w_t, w_t \sim N(0, W_t)$$

El esquema de inferencia queda determinado por,

a. Considere que la distribución a posteriori al tiempo  $t - 1$  es:

$$\theta_{t-1} | D_{t-1} \sim N_p(m_{t-1}, C_{t-1})$$

, Para algún vector de medias  $m_{t-1}$  y matriz de varianzas-covarianzas  $C_{t-1}$  conocidas. Estos valores son considerados los iniciales para el esquema.

b. Ahora, considere la distribución de evolución del vector de estado dada por:

$$\theta_t | D_{t-1} \sim N_p(a_t, R_t) \quad (2.15)$$

Donde,

$$\begin{aligned} a_t &= G_t m_{t-1} \\ R_t &= G_t C_{t-1} G_t' + W_t \end{aligned}$$

c. La distribución de la predicción a un paso, está dada por:

$$Y_t | D_{t-1} \sim N(f_t, Q_t) \quad (2.16)$$

Donde  $f_t = F_t' a_t$  y  $Q_t = F_t' R_t F_t + V_t$

d. Finalmente, la actualización de la distribución a posteriori de los parámetros será:

$$\theta_t | D_t \sim N_p(m_t, C_t) \quad (2.17)$$

con

$$m_t = a_t + A_t e_t \quad \text{y} \quad C_t = R_t + A_t A_t' Q_t$$

donde

$$A_t = R_t F_t Q_t^{-1} \quad \text{y} \quad e_t = Y_t - f_t$$

Con este filtro se conoce la serie y se infieren los parámetros de su distribución, para cada instante en el tiempo.

### 2.2.3. Suavizado

A través del suavizado, es posible estudiar el comportamiento de la serie desde otro enfoque, pues se puede *reconstruir* el vector de parámetros desde el principio hasta su fin  $(\theta_t, \theta_{t-1}, \dots, \theta_1)$ , habiendo observado la serie en su totalidad, a través del filtro de Kalman. Este proceso es retrospectivo y de aquí es posible encontrar la distribución suavizada asociada al vector de estado.

Considerando las matrices del modelo general dado por las ecuaciones 2.9 y 2.10, se definen las distribuciones suavizadas para algún  $t < T$  como,

$$\theta_{t+1} | D_T \sim N(s_t, S_t)$$

donde,

$$\begin{aligned} s_t &= m_t + C_t G_{t+1}' R_{t+1}^{-1} (s_{t+1} - a_{t+1}) \\ S_t &= C_t - C_t G_{t+1}' R_{t+1}^{-1} (R_{t+1} - S_{t+1}) R_{t+1}^{-1} G_{t+1} C_t \end{aligned}$$

### 2.2.4. Algoritmo FFBS

El algoritmo Forward Filtering Backward Sample (FFBS) presentado por Carter & Kohn (1994) y Frühwirth-Schanatter (1994) [5], es utilizado para obtener la distribución a posteriori de los parámetros. Este consiste en dos etapas, realizar un estudio de la serie hacia adelante, a través de la distribución filtrada, es decir,  $\theta_t | \theta_{t-1}$  y luego de haber observado la serie completa, realizar un estudio retrospectivo ( $\theta_{t-1} | \theta_t$ ), el cual también es conocido como muestreo de la distribución suavizada.

Es importante notar, que este algoritmo logra capturar información importante entre la dependencia implícita que existe entre las varianzas y el vector de parámetros de estado, pues, al obtener los valores de las varianzas, se logra actualizar las estimaciones de  $\theta_t$ , formando poco a poco una cadena de información mucho más certera y precisa.

Las etapas del algoritmo son:

1. Muestrear  $\theta_n$  desde la distribución actualizada obtenida en  $\theta_t | D_t \sim N(m_t, C_t)$  y hacer  $t = n - 1$ .
2. Muestrear  $\theta_t$  desde su distribución condicional  $\theta_t | \theta_{t+1}, V, W \sim N(\theta_t, \Sigma_t)$
3. Hacer  $t = t - 1$  y volver al paso 2, hasta llegar a  $t = 1$

Para obtener el primer paso se debe realizar el filtro de Kalman dados los valores de  $V$  y  $W$ , donde se obtiene la estimación para el último tiempo, es decir, en el tiempo  $n$ , pues, es este el vector de parámetros que se tomará como inicio para el segundo paso. Además, se deben guardar los valores de  $m_t$  y  $C_t$ , para utilizarlos en el segundo paso.

### 2.2.5. Inferencia de los parámetros

Hemos visto técnicas solo para conocer el comportamiento del vector de estado ( $\theta$ ), pero no nos podemos olvidar que existen más parámetros asociados a este modelo, como la varianza de la ecuación de estado ( $V_t$ ) y la matriz de varianza covarianza de los parámetros ( $W_t$ ), las cuales son de vital importancia para el funcionamiento del modelo.

Al definir el conjunto de matrices que definen el modelo ( $F_t, G_t, V_t, W_t$ ), se pueden conocer al menos dos de estas matrices, la matriz de regresión y de transición, dejando para la estimación la varianza de la ecuación de estado y la matriz varianza-covarianza de la ecuación de regresión. Por lo tanto se mostrará la forma de estimar estos valores considerando el comportamiento de la serie y viceversa.

Para obtener las estimaciones de las varianzas, se debe considerar el caso de  $F_t$  y  $G_t$  que son conocidas, y varianzas constantes para cada tiempo, esto es,  $V_t = V$  y  $W_t = W$ ,  $\forall t$ . Esta información sirve para obtener un modelo más parsimonioso. Por lo cual es necesario encontrar las distribuciones de cada una de las varianzas, pero claro está que no es posible encontrar una expresión analítica de las distribuciones marginales, por lo que se recurrirá a las distribuciones conjuntas condicionales al vector de parámetros y a la varianza. Estas distribuciones son las que se utilizan dentro de la metodología MCMC, que se presentará más adelante.

Se encontrarán las distribuciones de las siguientes densidades,

$$\pi(\phi | \theta, \Phi) \propto \prod_{t=1}^n f(y_t | \theta_t, \phi) p(\phi | \theta, \Phi) \quad (2.18)$$

$$\pi(\Phi|\theta, \phi) \propto \prod_{t=2}^n f(\theta_t|\theta_{t-1}, \Phi)p(\Phi|\theta, \phi) \quad (2.19)$$

donde,

$$\phi = \frac{1}{V} = V^{-1}$$

y

$$\Phi = \frac{1}{W} = W^{-1}$$

Para poder llevar a cabo este trabajo, es necesario considerar prioris independientes para cada una de las varianzas, esto es:

$$\phi \sim G\left(\frac{nv}{2}, \frac{nvSv}{2}\right)$$

$$\Phi \sim Wishart\left(\frac{nw}{2}, \frac{nwSw}{2}\right),$$

Conociendo las distribuciones a priori de 2.18 y 2.19, se tiene,

$$\begin{aligned} \pi(\phi|\theta, \Phi) &\propto \prod_{t=1}^n f(y_t|\theta_t, \phi)p(\phi|\theta, \Phi) \\ &\propto \prod_{t=1}^n f_N(F_t'\theta_t, \phi^{-1})f_G\left(\frac{nv}{2}, \frac{nvSv}{2}\right) \\ &\propto \frac{1}{\phi^{\frac{-n}{2}}} e^{\frac{-1}{2\phi^{-1}} \sum_{t=1}^n (y_t - F_t'\theta_t)^2} \phi^{\frac{nv}{2}-1} e^{\frac{nvSv}{2}\phi} \\ &\propto \underbrace{\phi^{\frac{nv+n}{2}-1} e^{\frac{-\phi}{2}(nvSv + \sum_{t=1}^n (y_t - F_t'\theta_t)^2)}}_{\text{núcleo } G(a/2, b/2)} \end{aligned}$$

Donde,

$$a = nv + n \quad y \quad b = nvSv + \sum_{t=1}^n (y_t - F_t'\theta_t)^2$$

Por lo tanto, la distribución a posteriori será

$$\phi|\theta, \Phi \sim G\left(\frac{a}{2}, \frac{b}{2}\right) \Rightarrow V|\theta, \Phi \sim G.I\left(\frac{a}{2}, \frac{b}{2}\right),$$

donde  $G.I$  denota una distribución Gamma Inversa de parámetros  $\frac{a}{2}$  y  $\frac{b}{2}$  respectivamente.

De la misma forma, es posible encontrar la distribución a posteriori para la matriz de varianza-covarianza de los parámetros, como sigue,

$$\begin{aligned} \pi(\Phi|\theta, \phi) &\propto \prod_{t=2}^n f(\theta_t|\theta_{t-1}, \phi)p(\Phi|\theta, \phi) \\ &\propto \prod_{t=2}^n f_N(G_t\theta_t - 1, \Phi^{-1})f_W\left(\frac{nw}{2}, \frac{nwSw}{2}\right) \\ &\propto \underbrace{|\Phi|^{\frac{n}{2}} e^{\frac{-1}{2\Phi^{-1}} \sum_{t=2}^n (\theta_t - G_t\theta_{t-1})(\theta_t - G_t\theta_{t-1})' |\Phi|} \frac{nw}{2}^{-p-1} e^{\frac{-1}{2} tr(\frac{nwSw\Phi}{2})}}_{\text{núcleo } W(c/2, d/2)} \end{aligned}$$

Donde,

$$c = nw + n - p \quad y \quad d = nwSw + \sum_{t=2}^n (\theta_t - G_t\theta_{t-1})(\theta_t - G_t\theta_{t-1})'$$

Por lo tanto, la distribución a posteriori será,

$$\Phi|\theta, \phi \sim W\left(\frac{c}{2}, \frac{d}{2}\right) \Rightarrow W|\theta, \phi \sim W.I\left(\frac{c}{2}, \frac{d}{2}\right),$$

donde  $W.I$  denota la distribución Wishart Inversa de parámetros  $\frac{c}{2}$  y  $\frac{d}{2}$ .

Finalmente es indispensable obtener la distribución a posteriori condicional al valor de las varianzas y de los parámetros de estado ya observados, para el vector de parámetros de estado, es decir,  $\theta_t|\theta_{t-1}, V, W$ .

En base a Carter & Kohn (1994) [3] y Frühwirth-Schnatter (1994) [5], donde señalan que para  $t = 1, \dots, n - 1$ , la distribución presenta el siguiente comportamiento,

$$\theta_t|\theta_{t+1}, V, W \sim N(\mu_t, \Sigma_t) \quad (2.20)$$

Con

$$\begin{aligned} \mu_t &= (G'_t W^{-1} G_t + C_t^{-1})(G'_{t-1} W^{-1} \theta_{t+1} + C_t^{-1} m_t) \\ \Sigma_t &= (G'_t W^{-1} G_t + C_t^{-1})^{-1} \end{aligned}$$

Es necesario enlazar todas las distribuciones ya estudiadas, para poder obtener resultados a partir de ellas. Para esto se recurre al algoritmo de la teoría Markov Chain Monte Carlo, conocido como “Gibbs Sampling”.

### 2.2.6. Gibbs Sampling

El algoritmo Gibbs sampling puede verse como un caso particular del algoritmo M-H. A pesar de esto, sus fundamentos metodológicos y motivación histórica son completamente distintos.

Este consiste en muestrear cada una de las partes en forma independiente, a partir de sus distribuciones a posterioris conjuntas, produciendo que en cada ciclo se actualicen los parámetros en función de los valores obtenidos en el ciclo anterior.

Para nuestro caso, en particular, el algoritmo estará compuesto de tres pasos, pues, deseamos las estimaciones de  $\theta_t$ ,  $V$  y  $W$ .

1. Muestrear la varianza de las observaciones a partir de su distribución a posteriori condicional, encontrada en 2.18.
2. Muestrear la matriz de varianzas-covarianzas de los parámetros, a partir de su distribución a posteriori condicional, encontrada en 2.19
3. Muestrear los parámetros a partir de su distribución condicional a los valores de las varianzas, es decir  $\pi(\theta|y_n, V, W)$ , la cual se obtiene a través del algoritmo FFBS.

Una de las ventajas más importantes de este algoritmo es que todas las simulaciones son aceptadas, y en cada transición se obtiene un punto diferente de la cadena. Esto se debe a que la probabilidad de aceptación M-H es 1 en todo momento.

Mostrado el análisis inferencial para conocer la serie y construir los parámetros necesarios asociados, es posible presentar la forma de obtener las predicciones que entregará el modelo dinámico.

### 2.2.7. Distribución de Pronóstico

Una vez que se conoce el comportamiento de nuestra serie en estudio, podemos proceder a obtener un pronóstico de la variable en estudio.

Por ello, es necesario conocer las leyes de probabilidad que contienen la relación que existirá entre lo observado hasta un determinado tiempo  $t$  y la estimación de la futura observación a  $k - \text{pasos}$ , expresado de otra forma,  $Y_{t+k}|D_t$ .

Se define la función de pronóstico  $f_t(k)$ , para algún tiempo  $t$  y para valores de  $k > 0$  como:

$$f_t(k) = E(Y_{t+k}|D_t) \quad (2.21)$$

Esta función es equivalente al cálculo del nivel esperado de la serie en el tiempo  $t + k$ , es decir,

$$f_t(k) = E(\mu_{t+k}|D_t) = E(F'_{t+k}\theta_{t+k}|D_t), \quad k \geq 1,$$

donde  $F_{t+k}$  es conocida y la generalización de  $\theta_{t+k}$  está dada por,

$$\theta_{t+k} = \prod_{j=1}^k G_{t+j}\theta_t + \sum_{i=1}^k \left[ \prod_{j=1}^k G_{t+i-j} \right] \omega_{t+i}$$

Se debe notar que cuando 2.21 es evaluada en  $k = 0$ , el pronóstico será un estimador puntual a posteriori del nivel actual de la serie.

De forma similar al esquema inferencial secuencial, se obtiene la metodología de la distribución pronóstico, pues en primer lugar se realiza la actualización del vector de estado y luego se obtienen los parámetros de dicha distribución. Es importante destacar que este proceso es recursivo.

El esquema queda de la siguiente forma:

1. *Distribución de estado:*

$$\theta_{t+k}|D_t \sim N(a_t(k), R_t(k)),$$

con,

$$a_t(k) = G_{t+k}a_t(k-1) \quad R_t = G_{t+k}R_t(k-1)G'_{t+k} + W_{t+l}$$

2. *Distribución del Pronóstico:*

$$Y_{t+k}|D_t \sim N(f_t(k), Q_t(k)),$$

con

$$f_t(k) = F'_{t+k}a_t(k) \quad Q_t = F'_{t+k}R_t(k-1)F_{t+k} + V_{t+l},$$

con valores iniciales dados por:

$$a_t(0) = m_t \quad R_t(0) = C_t$$

Finalmente, al realizar cada uno de los pasos descritos en esta sección es posible conocer la serie en su totalidad, realizar un análisis retrospectivo y realizar un análisis de pronósticos, a partir de lo estudiado hasta un determinado tiempo.

## 2.2.8. Comparación de modelos

La comparación de modelos, ya sea fijos o dinámicos, es muy útil en situaciones donde existen diversos tipos de modelos aplicados a los mismos datos, pues, se produce la incertidumbre de qué modelo será el mejor al momento de obtener resultados. Para solucionar este problema, se procede a la utilización de métodos Bayesianos, que miden tanto la dificultad del modelo, como la capacidad predictiva.

Existen diversas metodología para obtener los valores asociados, una de estas es el Criterio de información sobre la Devianza.

### 2.2.8.1. Criterio de Información de la Devianza (DIC)

El progreso en la computación bayesiana debido a los métodos MCMC ha hecho posible el ajustar modelos estadísticos complejos lo que implica el determinar el modelo que mejor se ajuste de una gran cantidad de posibles candidatos. Por lo tanto, se ha hecho cada vez más necesario e importante el desarrollar criterios que ayuden en la selección del modelo más eficiente.

Una propuesta hecha por Spiegelhalter, Best y Carlin (2002) [17] es el Deviance Information Criterion (DIC), una versión o generalización bayesiana del conocido Criterio de Información de Akaike (AIC), el cual se relaciona además con el Criterio de Información Bayesiana (BIC).

El DIC se basa en la distribución a posteriori de la log-verosimilitud. Identifica el número de parámetros en los modelos, lo que es requerido para el cálculo del Criterio de Información Bayesiana y el Criterio de Información de Akaike.

Asumamos en general, que la distribución condicional de los datos  $y = (y_1, \dots, y_n)'$ , depende de un vector de parámetros  $\theta$  con dimensión  $p$ . Desde un punto de vista frecuentista, la valoración del modelo se basa en la devianza, la diferencia de la log-verosimilitud entre los modelos ajustados y el saturado. El modelo saturado se refiere al modelo con tantos parámetros como observaciones, lo que conlleva a un ajuste perfecto de los datos. La distribución posteriori de la devianza clásica para un modelo bayesiano se define por,

$$D(\theta_t) = -2\log f(Y_t|\theta_t) + 2\log h(Y_t)$$

donde  $f(Y_t|\theta_t)$  es la función de verosimilitud,  $Y_t$  es el vector de datos temporales,  $\theta_t$  corresponde a todos los parámetros del modelo y  $h(\cdot)$  es una función que depende solo de los datos.

El DIC consiste en dos componentes: un término que determina la bondad de ajuste y el otro término que penaliza el incremento de la complejidad del modelo. Entre mayor ajuste existe del modelo a los datos, mayores son los valores de la verosimilitud.  $\bar{D}$  se define como dos veces la log-verosimilitud. El segundo componente cuantifica la complejidad del modelo mediante el número efectivo de parámetros,  $p_D$ , definido como la diferencia entre la media posteriori de la desviación y la desviación evaluada en la media posteriori de los parámetros:

$$\begin{aligned} p_D &= E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) \\ &= \bar{D} - D(\bar{\theta}) \\ &= E_{\theta|y} \left[ -2\log f(y|\theta) + 2\log f(y|\bar{\theta}(y)) \right] \end{aligned}$$

De la expresión anterior puede notarse que  $p_D$  es el término que penaliza para verosimilitudes aproximadamente normales y para información a priori difusa. En esta situación, la media posteriori iguala el estimador de máxima verosimilitud (MLE)  $\hat{\theta}_t$ , y la distribución a posteriori

de  $\theta_t$  es bien aproximada por una distribución normal multivariada con media igual a el estimador de máxima verosimilitud,

$$\theta|y \approx N \left( \hat{\theta}_t \frac{\partial^2 \theta_t}{\partial \theta_t^2} | \theta_t = \hat{\theta}_t \right)$$

Representando la desviación a través de una expansión de Taylor de segundo orden alrededor del estimador de máxima verosimilitud, el término de primer orden desaparece y la desviación es aproximada por la suma de la desviación en el estimador de máxima verosimilitud y un término cuadrático que sigue una distribución  $\chi^2$  con  $p$  grados de libertad. Al obtener a esperanza, el número efectivo de parámetros iguala a  $p$ , el verdadero número de parámetros.

El Criterio de Información de la Devianza (DIC) se define entonces como la suma de ambos componentes antes mencionados

$$DIC = \bar{D} + p_D = 2\bar{D} - D(\bar{\theta}) = D(\bar{\theta}) + 2p_D$$

En una situación clásica, el DIC puede ser visto como una generalización del conocido Criterio de Información de Akaike (AIC), el cual se define por  $AIC = D(\hat{\theta}) + 2p$  donde  $\hat{\theta}$  es el estimador de máxima verosimilitud (MLE) del vector de parámetros y siendo  $p$  el número de parámetros. Para modelos no jerárquicos,  $p \approx p_D$ ,  $\hat{\theta} \approx \theta$  y  $DIC \approx AIC$ .

## 2.3. Modelos clásicos de series de tiempo

### 2.3.1. Introducción

Los modelos clásicos de series de tiempo usan el pasado de las variables en orden para predecir sus valores futuros. De esta forma, en términos de predicción, la confiabilidad de la estimación debe basarse en el rendimiento de la muestra, según Stock and Watson (2012) [18].

Los modelos de series de tiempo en su mayoría pueden producir predicciones bastante precisas, especialmente en casos donde hay relación multidimensional entre las variables.

Los modelos de series de tiempo usando Box-Jenkins cuyo enfoque ha sido propuesto por Box and Jenkins en 1970. [1] Este enfoque ha sido ampliamente utilizado en la literatura debido a su rendimiento y simplicidad.

### 2.3.2. Componentes de una serie de tiempo

Una serie de tiempo puede ser escrita como:

$$\begin{aligned} \text{Serie Aditiva} \quad x_t &= T_t + S_t + \varepsilon_t \\ \text{Serie Multiplicativa} \quad x_t &= T_t S_t + \varepsilon_t \end{aligned}$$

donde,

- $T_t$ : Componente de tendencia.
- $S_t$ : Componente Estacional.
- $\varepsilon_t$ : Componente aleatorio.

### 2.3.3. Conceptos básicos

#### Estacionaridad

- Un proceso estocástico  $\{Z(t), t \in T\}$  se dice estrictamente estacionario si todas las distribuciones de dimensión finita permanecen iguales en el tiempo, es decir,  $F(x_1, \dots, x_n) = F(x_{k+1}, \dots, x_{k+n})$
- Un proceso estocástico  $Z = \{Z(t), t \in T\}$  se dice débilmente estacionario (o estacionario de 2° orden) si,
  - i)  $\mathbb{E}(Z_t) = \mu = cte$
  - ii)  $\mathbb{E}(Z_t^2) < \infty$
  - iii)  $cov(Z_t, Z_s) = \gamma(t - s), \forall t, s \in \mathbb{Z}$

#### Función de autocovarianza

Para un proceso débilmente estacionario se llama función de autocovarianza a:

$$\begin{aligned}\gamma(k) &= \mathbb{E}[(X_t - \mu)(X_{t+k} - \mu)] \\ &= cov(X_t, X_{t+k}) \\ &= cov(X_{t+k}, X_t)\end{aligned}$$

De la misma forma, podemos definir la función de autocorrelación (FAC, ACF):

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}, \rho(k) \in [-1, 1]$$

#### Ruido Blanco (White Noise)

Un Ruido Blanco es un proceso estocástico que se caracteriza por el hecho de que sus valores de señal en dos tiempos diferentes no guardan correlación estadística. Un proceso débilmente estacionario se dice ruido blanco si:

- i)  $\mathbb{E}(Y_t) = cte$
- ii)  $V(Y_t) = \sigma^2$
- iii)  $cov(Y_t, Y_{t+k}) = \gamma(k) = 0, \forall t \in T, t \neq k$

Usualmente se considera el caso de  $\mathbb{E}(Y_t) = 0$ , que se conoce como ruido blanco centrado, Notación:  $Y_t \sim RB(0, \sigma^2)$ , ( $Y_t \sim WN(0, \sigma^2)$ ) [13]

### 2.3.4. Modelos Estacionarios

#### 2.3.4.1. Procesos lineales

**Definición:** Se define el operador de rezago como,  $BX_t = X_{t-1}$

Además, se cumple que  $B^j X_t = X_{t-j}$

**Definición:** Se define el operador de diferencia con rezago 1, denotado por  $\nabla$ , como

$$\nabla X_t = X_t - X_{t-1} = X_t - BX_t = (1 - B)X_t$$

por lo tanto

$$\nabla = (1 - B)$$

**Observación :** No se puede aplicar el operador de rezago a dos series que no sean comparables, es decir, que tengan frecuencias distintas (meses, semanas, días).

### 2.3.4.2. Modelo Auto-regresivo (AR)

Un proceso estocástico  $\{X_t\}, t \in T, T \in \mathbb{Z}$ , se dice auto-regresivo de orden  $p$  ( $AR(p)$ ) si,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t, t \in T, p \geq 1$$

donde,

$$\begin{aligned}\varepsilon_t &\sim RB(0, \sigma^2) \\ cov(\varepsilon_t, X_{t-j}) &= \mathbb{E}(\varepsilon_t X_{t-j}) = 0, j > 0 \\ cov(\varepsilon_t, \varepsilon_t) &= \sigma^2\end{aligned}$$

$\phi_1, \dots, \phi_p$  son coeficientes fijo (a determinar). Entonces se define el polinomio,

$$\begin{aligned}\Phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \Phi(B)X_t &= \varepsilon_t\end{aligned}$$

**Notación:**  $X_t \sim AR(p), \{X_t\} \sim AR(p)$

### 2.3.4.3. Modelos de medias móviles (MA)

Sea  $\{\varepsilon_t\}$  una sucesión de ruido blanco. Consideremos el proceso  $\{X_t\}$  definido por,

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

Este modelo se conoce como modelo de medias móviles, o en inglés *moving average*, desde ahora  $MA(q)$

Otra forma de escribir el proceso,

$$\begin{aligned}X_t &= \varepsilon_t - \theta_1 B \varepsilon_t - \dots - \theta_q B^q \varepsilon_t \\ &= (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \\ &= \Theta(B) \varepsilon_t\end{aligned}$$

**Notación:**  $X_t \sim MA(q), \{X_t\} \sim MA(q)$

### 2.3.4.4. Modelo auto-regresivo de medias móviles (ARMA)

Sea  $\{X_t\}$  un proceso definido por,

$$X_t = \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

El cual es una combinación de los procesos  $AR(p)$  y  $MA(q)$ .

Donde,  $\varepsilon_t \sim RB(0, \sigma^2)$  y  $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ , son coeficientes fijos.

Otra forma de escribir el proceso es

$$\Phi(B)X_t = \Theta(B)\varepsilon_t$$

**Notación:**  $X_t \sim ARMA(p, q), \{X_t\} \sim ARMA(p, q)$

### 2.3.5. Metodología de Box-Jenkins

En el análisis de los modelos paramétricos, el método de Box & Jenkins [1] ha recibido mucha atención en la última década. El método incluye el ajuste ARMA y ARIMA para un conjunto de Datos. La estrategia para la construcción del modelo se basa en un ciclo iterativo, en el que la elección de la estructura del modelo se basa en sus propios datos. Las etapas del ciclo iterativo son

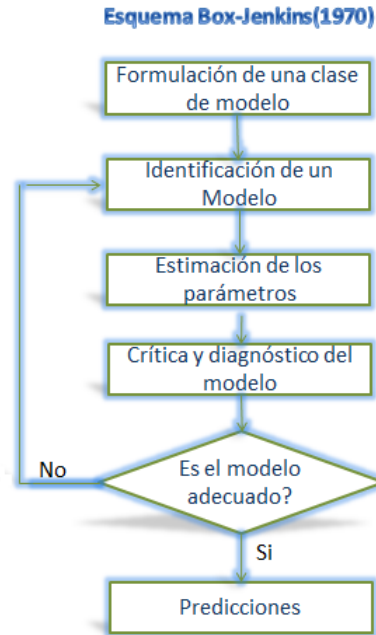


Figura 2.1: Esquema Box-Jenkins

### 2.3.6. Modelos para series con tendencia

#### Diferenciación

Cuando estamos frente a series no estacionarias, esto se puede resolver usando el operador de diferencia  $\nabla^d = (1 - B)^d$ , tal que

$$Y_t = (1 - B)^d X_t$$

Al aplicar el operador de rezago, se debe tener cuidado, debido a que este puede aumentar la varianza del proceso. Es por esto, que si se tiene duda sobre aplicar el operador de diferencia o no, se debe estudiar la variabilidad de la serie original y diferenciada.

#### 2.3.6.1. Modelo ARIMA

Sea  $Y_t = (1 - B)^d X_t$  y suponemos  $Y_t \sim ARMA(p, q)$ , entonces

$$\Phi(B)Y_t = \Theta(B)\varepsilon_t$$

Luego podemos reescribir el proceso como

$$\Phi(B)(1 - B)^d X_t = \Theta(B)\varepsilon_t$$

**Notación:**  $X_t \sim ARIMA(p, d, q)$ ,

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

## 2.4. Modelos para series estacionales

### 2.4.0.1. Modelo SARIMA

Teniendo en cuenta la descomposición clásica  $X_t = T_t + S_t + N_t + \varepsilon_t$ . Se proponen los modelos SARIMA, para permitir la aleatoriedad en la forma estacional de ciclo a ciclo.

Supongamos una serie con periodo anual, diremos que puede ser escrito como,

$$\Phi(B^{12})X_t = \Theta(B^{12})\varepsilon_t$$

Es posible notar que  $cov(\varepsilon_t, \varepsilon_{t+k})$  es cero solo en los casos en que  $k$  es múltiplo de 12. Para incorporar esta diferencia entre series, se tiene que  $\varepsilon_t \sim ARMA(p, q)$ . Luego combinando esto se tiene

$$\Phi(B^{12})\Phi(B)X_t = \Theta(B^{12})\Theta(B)\varepsilon_t$$

Finalmente se define un modelo SARIMA como:

$$\Phi(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D X_t = \Theta(B)\Theta_Q(B^s)\varepsilon_t$$

**Notación:**  $\{X_t\} \sim SARIMA(p, d, q)x(P, D, Q)_s$

## 2.5. Modelos con Variables Exógenas

### 2.5.1. Modelos ARIMAX

Los modelos ARIMAX son una extensión del ARIMA, adaptándole al modelo, una o más variables exógenas representadas por  $X_t$ . Del mismo modo, se requiere que las series que se incluyan como variables exógenas sean estacionarias o sean transformadas.

En pocas palabras, un modelo ARIMAX se puede ver como un modelo de regresión múltiple con uno o más términos autorregresivos (AR) y/o uno o más términos de media móvil (MA). Los términos autorregresivos para una variable dependiente son simplemente valores rezagados de esa variable que tiene una relación estadísticamente significativa con su valor más reciente. Los términos de medias móviles no son más que residuos (errores rezagados) resultantes de estimaciones hechas previamente. [19]

Luego, estos modelos quedan expresados como:

$$\Phi(B)(1-B)^d Y_t = \beta_i(B)X_{i,t} + \Theta(B)\varepsilon_t \quad (2.22)$$

Donde  $\beta_{it}$  corresponde al parámetro que acompaña a cada una de las variables independientes, entonces  $i$  puede ir desde 1 hasta el valor de la cantidad de variables independientes.

### 2.5.2. Modelos ARIMAX Estacionales

Estos modelos son similares a los modelos ARIMAX, con la diferencia que estos, al igual que los modelos SARIMA, se pueden aplicar a series de tiempo estacionales y tienen una estructura similar.

Estos modelos quedan expresados como:

$$\Phi(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \beta_i(B)X_{i,t} + \Theta(B)\Theta_Q(B^s)\varepsilon_t$$

Donde al igual que en los modelos ARIMAX, la cantidad de parámetros  $\beta$  depende de la cantidad de variables exógenas que se incluya en el modelo.

### 2.5.3. Evaluación de Modelos

Es importante recordar que los modelos de series de tiempo deben cumplir con el supuesto de que los errores deben ser ruido blanco, es decir, deben tener media igual a cero, varianza constante y no deben estar correlacionados.

Para verificar que los errores no están correlacionados se pueden utilizar los siguientes test de hipótesis:

- **Test de Box-Pierce**

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_l = 0 \quad v/s \quad \exists \rho_k \neq 0, \quad \forall k > 0$$

El estadístico es

$$Q = \sum_{h=1}^l N \hat{\rho}_h^2 \sim \chi_l^2$$

- **Test de Ljung-Box**

Bajo la misma hipótesis del test anterior, para muestras pequeñas, se tiene el estadístico

$$Q^* = N(N+2) \sum_{h=1}^l \frac{\hat{\rho}_h^2}{N-h} \sim \chi_l^2$$

## 2.6. Escalas internacionales para medir Congestión

El problema de congestión en los servicios de urgencia(SU), es un problema que aqueja los sistemas de salud en todo el mundo. Se han diseñado varias escalas para medir la congestión en urgencias a partir del diseño de métodos de cuantificación de los factores que se han relacionado con dicho problema, entre las que se han propuesto se encuentran: Emergency Department Work Index (EDWIN) y el National Emergency Department Overcrowding Study Scale (NEDOCS) [14].

### 2.6.1. National Emergency Department Overcrowding Study (NEDOCS)

El National Emergency Department Overcrowding Study score (NEDOCS) es un instrumento calculado en tiempo real a partir de un modelo de regresión lineal que considera siete variables operacionales para determinar el grado de congestión en base a lo que perciben los médicos y enfermeras de la urgencia, la expresión para el cálculo de la escala NEDOCS está dada por: [20]:

$$NEDOCS = -20 + 85,8 \frac{L_{ED}}{b_{ED}} + 600 \frac{L_h}{b_h} + 5,64L_{rp} + 0,93W_{ED} + 13,4W_{AT} \quad (2.23)$$

Donde,

- $L_{ED}$ : Número total de pacientes registrados en urgencias.
- $b_{ED}$ : Número de camas en el servicio.
- $L_h$ : Número de pacientes hospitalizados en el servicio de urgencias.
- $b_h$ : Número de camas destinadas para hospitalización en el servicio.

- $L_{rp}$ : Número de pacientes en urgencias usando ventilación mecánica.
- $W_{ED}$ : El mayor tiempo de espera para los pacientes de urgencias (en horas).
- $W_{AT}$ : Tiempo en la sala de espera del último paciente atendido.

A partir de esta expresión se obtiene un *score* que está asociado a un nivel de congestión, que va desde normal a un nivel de congestión considerado peligroso. Según este índice se definen 5 niveles de congestión según el valor del puntaje obtenido:

- **Normal:** Entre 0 y 50
- **Ocupado:** Entre 51 y 100
- **Congestionado:** Entre 101 y 140
- **Congestión severa:** Entre 141 y 180
- **Colapsado:** Sobre 180

### 2.6.2. Emergency Department Work Index (EDWIN)

Este modelo, al igual que la escala NEDOCs es utilizado en tiempo real, se consideran 5 variables asociadas principalmente al número de pacientes y a la complejidad de estos. Se calcula como: [11]

$$\frac{\sum_{i=1}^5 n_i t_i}{N_a (B_T - B_A)}, i = 1, \dots, 5 \quad (2.24)$$

Donde,

- $n_i$ : Número de pacientes con categoría triage  $i$ .
- $t_i$ : Categoría triage  $i$ .
- $N_a$ : Número de médicos atendiendo.
- $B_T$ : Número total de camas de atención en el servicio.
- $B_A$ : Número de pacientes ingresados.

Se definen tres niveles de congestión, definidos en los siguientes intervalos:

- $EDWIN < 1,5$ : Bajo Control
- $1,5 \leq EDWIN < 2$ : Ocupado
- $EDWIN > 2$ : Colapsado

## Capítulo 3

# Análisis Descriptivo

### 3.1. Descripción de la base de datos

La base de datos cuenta con información de todos los pacientes que realizaron alguna consulta al servicio de urgencias de la ex Posta Central en el periodo del 1 de Enero de 2016 al 31 de Diciembre de 2017. En total se cuenta con 244483 consultas al servicio en el periodo de tiempo antes señalado.

La base de datos cuenta inicialmente con 185 variables, entre estas se encuentran variables demográficas, fechas y horas que van desde el ingreso de cada paciente hasta el egreso de estos del servicio, pasando por cada instancia, variables que caracterizan a los pacientes, como sexo y edad, variables que contemplan diversos tiempos de espera, entre otras.

#### 3.1.1. Descripción de Variables

Algunas variables de interés son:

- **Edad:** Edad del paciente en años.
- **Sexo:** Género del paciente. En los registros se encuentra Femenino, Masculino, Desconocido e Indefinido.
- **Área:** Área de atención a la que es derivada el paciente según su diagnóstico.
- **Fecha de Ingreso:** Fecha en que el paciente ingresa al servicio.
- **Hora de Ingreso:** Hora en que el paciente ingresa al servicio.
- **Fecha de Categorización:** Fecha en que el paciente es categorizado.
- **Hora de Categorización:** Hora en que el paciente es categorizado.
- **Fecha de Atención:** Fecha en que el paciente es atendido en el box.
- **Hora de Atención:** Hora en que el paciente es atendido en el box.
- **Fecha de Egreso:** Fecha en que el paciente es dado de alta.
- **Hora de Egreso:** Hora en que el paciente es dado de alta.
- **Tiempo de Espera:** Tiempo de espera del paciente desde la categorización hasta que es llamado a box.

- **Tiempo en Urgencia:** Tiempo que el paciente permanece en el servicio de urgencia, desde la Hora de Ingreso hasta la de Egreso.
- **Triage:** Clasificación del paciente según su gravedad. Las categorías van de C1 a C5, siendo los pacientes C1 los de mayor gravedad y C5 los menos graves y fuera de riesgo vital.
- **Código de diagnóstico:** Código para identificar el diagnóstico del paciente.
- **Descripción del diagnóstico:** Descripción del diagnóstico del paciente.
- **Previsión:** Previsión de salud del paciente.
- **Nacionalidad:** Nacionalidad del paciente.
- **Sin Diagnóstico:** Indica si el paciente abandona la urgencia sin un diagnóstico.
- **Día de la Semana:** Día de la semana que ocurre la consulta.
- **Destino Alta:** Destino del paciente luego del alta del Servicio de Urgencia.

### 3.1.2. Filtros en la Base de Datos

La urgencia del HUAP se divide en 3 áreas de atención, estas son, *Traumatología*, *Adultos* y *Dental*. El área de traumatología recibe, como su nombre lo indica, todo tipo de consultas de origen traumológica, ya sea por esquinces, fracturas y todas aquellas consultas que contemplan lesiones en el aparato locomotor. Por otra parte el área de Adultos recibe todo tipo de consultas a medicina general y es el área que más consultas recibe al año. Finalmente, el área de Dental recibe consultas dentales de urgencia. Como se aprecian en el Cuadro 3.1, el área

	Adultos	Dental	Trauma
2016	68969	38032	9965
2017	75779	41623	10115
Total	144748	79655	20080

Cuadro 3.1: Consultas en cada área por año

con más consultas, es la de Adultos, con cerca del 59% de las consultas, seguida por Dental con alrededor del 33% del total de consultas y finalmente el área de traumatología sólo recibe cerca del 8% de las consultas.

Al ser el área Adultos la que recibe mayor cantidad de pacientes, es también la que presenta más problemas en cuanto a congestión, es por eso que este trabajo sólo se enfocará en esta área del servicio.

Por otra parte, es importante analizar los tiempos de espera, y el tiempo en que un paciente está haciendo uso de una camilla, pues estos pueden influir en la congestión o pueden ser consecuencia de esta.

Al realizar el análisis de los tiempos de espera es posible observar datos anómalos, tanto por tiempos negativos, como tiempos excesivos de espera. En el Cuadro 3.2 se puede observar por ejemplo, que el tiempo medio de espera en pacientes C1 es de cerca de 2 horas, siendo el tiempo máximo aproximadamente de 1671 horas, lo que no tiene sentido, pues estos pacientes, que son los de mayor gravedad no debieran esperar por tanto tiempo. Por otra parte, también es

posible observar que hay tiempos mínimos negativos, lo que tampoco tiene sentido.

Una de las posibles explicaciones a estos tiempos anómalos puede ser que hay pacientes que abandonan la urgencia sin un diagnóstico y no dan aviso de su retiro, esto provoca en algunos casos que su ficha quede abierta por tiempos indefinidos hasta que alguien del personal se percata de la deserción e ingresa una hora que no es la correcta.

A lo anterior se puede sumar que muchos de los registros no presentan hora, tanto de categorización, atención o egreso, esto también dificulta el cálculo de los tiempos de espera y atención. Es importante destacar que aquellos pacientes que no tiene hora de categorización, tampoco tienen un triage asociado.

En el Cuadro 3.3 se puede observar que efectivamente aquellos pacientes que abandonan la urgencia sin un diagnóstico, presentan un tiempo medio de espera mucho mayor que aquellos con diagnóstico, sin embargo de igual forma hay tiempos atípicos en estos pacientes.

Triage	Media	SD	Min	Med	Max	n
C1	1,76	41,77	-321,90	0,10	1671,30	1961
C2	0,17	0,43	-4,50	0,10	26,10	9131
C3	1,25	2,93	-489,00	0,80	181,70	52282
C4	3,08	9,49	-4,60	1,70	1491,40	74660
C5	5,74	24,89	-1,80	2,40	1178,60	4569

Cuadro 3.2: Resumen de tiempos de espera en horas por Triage

Diagnóstico	Media	SD	Min	Med	Max	n
Con Diag	1,29522	8,36139	-7,7	0,8	1671,3	118398
Sin Diag	7,11573	13,86964	-489	5,6	1178,6	24278

Cuadro 3.3: Tiempos de espera en horas de pacientes sin diagnóstico

### 3.1.3. Filtros por anomalías y valores atípicos

Luego de haber visto las tablas 3.2 y 3.3 con los tiempos de espera, queda claro que hay anomalías en la base de datos en cuanto a tiempos de espera de atención negativos y esto también sucede con los tiempos en que los pacientes están usando camilla. Debido a que la ocurrencia de estos sucesos no parece tener una razón lógica, más que error en el ingreso de los datos, se procede a eliminar a estos pacientes de la base, en total 522 observaciones.

Por otra parte, como se mencionó anteriormente, se piensa que aquellos pacientes que abandonan la urgencia sin ser atendidos son los que influyen en gran parte en los tiempos extensos de espera, es por esto que se estudió la posibilidad de imputar tiempos de espera más razonables a aquellos con valores altos, según la hora de llegada de los pacientes con la misma categoría de aquel que no tiene diagnóstico, justo antes y después que él, asignando el tiempo medio entre estos, como su tiempo de espera de atención. Si bien esta idea suena bastante lógica, en la práctica, por la naturaleza de la base de datos no fue posible de realizar, ya que ocurre que no es seguro que dos pacientes que llegan consecutivos, con el mismo triage, sean llamados en el mismo orden.

Aparte de la dificultad de imputar tiempos de espera a los pacientes sin diagnóstico, también se encontraron anomalías en los tiempos de espera según triage, viéndose esperas elevadas en pacientes con mayor gravedad, lo que en la práctica no ocurre, pues está en riesgo la vida de los pacientes. Es por lo anterior, que finalmente se decidió eliminar a aquellos pacientes con tiempos atípicos según triage.

Se dejó fuera a aquellos pacientes cuyos tiempos de espera superaban el valor de  $RIC + 3Q_3$  correspondiente a cada triage, es decir para cada triage hay un valor diferente de corte.

También se eliminó a aquellos pacientes sin categoría triage, ya que este es importante para el cálculo del índice EDWIN. No fue posible asignar un triage a los pacientes según diagnóstico, ya que para un diagnóstico no hay una única categoría.

Se encontró, por otra parte, pacientes sin fecha ni hora de atención, a algunos de estos fue posible asignar un tiempo a través de otras variables y los que no pudieron ser asignados también se quitaron de la base de datos.

Una vez solucionado el problema con el tiempo de espera por atención, es necesario también hacer filtros a partir de los tiempos de uso de camilla, esto es, el tiempo que transcurre entre la llamada a box y el abandono de la urgencia.

Para filtrar por tiempo uso de camilla, es necesario saber que hay pacientes que quedan hospitalizados en la misma urgencia, lo que hace que el tiempo que transcurre entre que el paciente es llamado y que este abandone la urgencia, pueda superar 1 día. Sin embargo, también se observan valores atípicos en estos tiempos, por lo que, para remover a estos pacientes se opta por un criterio similar al utilizado para remover los valores atípicos en los tiempos de espera.

También es posible encontrar observaciones sin fecha y hora de egreso, las que son removidas de igual forma de la base de datos, pues esta servirá para saber cuándo el paciente deja de utilizar una cama.

Finalmente fueron removidas en total 12641 observaciones, además de las 522 correspondientes a tiempos negativos. La base de datos final cuenta con 131585 observaciones, removiéndose aproximadamente el 9% de los datos.

## **3.2. Análisis descriptivo de variables de interés**

Una vez filtrada la base de datos, es posible construir una nueva base de tal forma que se pueda representar como una serie de tiempo.

Del Cuadro 3.4 es posible observar que los días lunes hay más consultas y por el contrario, los fines de semana hay menor flujo de pacientes en el servicio de urgencias, mientras que el resto de la semana la cantidad de consultas es similar.

De la Cuadro 3.5 se puede desprender que en la mayoría de las consultas, los pacientes son clasificados como C4, es decir que la mayoría de los pacientes que acuden al HUAP no son de gran gravedad, por otra parte, los pacientes con el nivel más alto de gravedad sólo alcanzan el

Día	Frecuencia	Porcentaje
Lunes	20523	15,60
Martes	19405	14,75
Miércoles	18848	14,32
Jueves	19253	14,63
Viernes	19058	14,48
Sábado	17596	13,37
Domingo	16902	12,84

Cuadro 3.4: Frecuencias de consultas por Día de la semana

0,91 %.

Triage	Frecuencia	Porcentaje
C1	1528	1,16
C2	8529	6,48
C3	49977	37,98
C4	67558	51,34
C5	3993	3,03

Cuadro 3.5: Frecuencias consultas por Triage

En cuanto a los pacientes hospitalizados, estos sólo alcanzan el 14,44 %, sin embargo estos son los que permanecen por más tiempo ocupando camas, lo que podría influir en la congestión.

Destino Alta	Frecuencia	Porcentaje
Hospitalización	19007	14,44
Otro destino	112578	85,56

Cuadro 3.6: Frecuencia destino de los pacientes después del alta

Y como se observa en la tabla 3.7, sólo un 13,4 % abandona la urgencia sin diagnóstico, sin embargo estos también podrían afectar en la congestión de la posta central.

### 3.3. Análisis del Tiempo de Espera y Tiempo de Uso de Camilla

Para poder hacer un estudio de la congestión en el HUAP, es necesario hacer un análisis de los tiempos de espera de atención y también del tiempo que tarda un paciente en ser atendido.

#### Tiempo Espera

Diagnóstico	Frecuencia	Porcentaje
Con Diag	113959	86,6
Sin Diag	17626	13,4

Cuadro 3.7: Frecuencias consultas sin diagnóstico

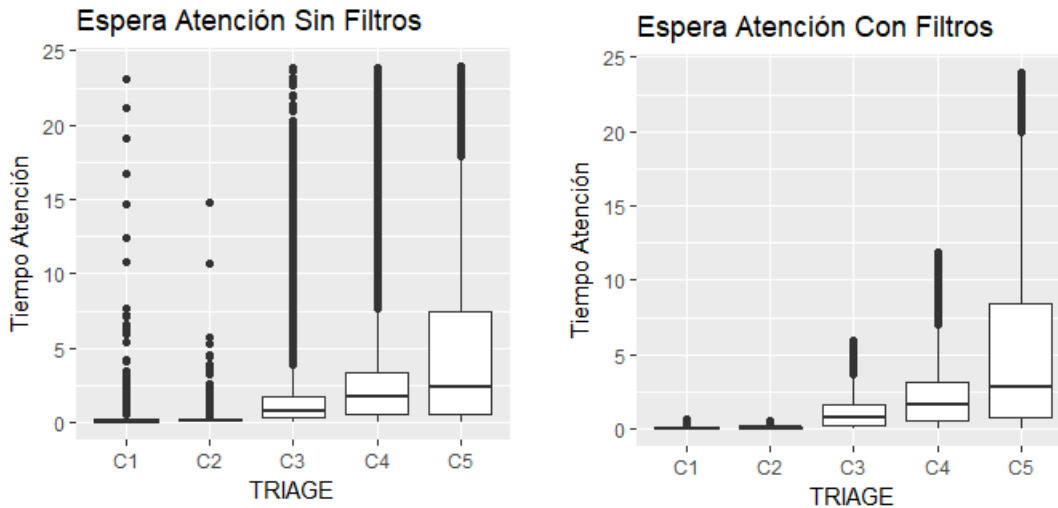


Figura 3.1: Boxplot tiempo de espera según Triage

En la figura 3.1 es posible observar los tiempos de espera antes y después de aplicar los filtros mencionados anteriormente, es posible verificar que el comportamiento de la espera después de eliminar los valores atípicos se ajusta mucho más a lo que se espera que suceda, en donde los pacientes más graves esperan a ser atendidos menos tiempo y los menos graves, esperan más.

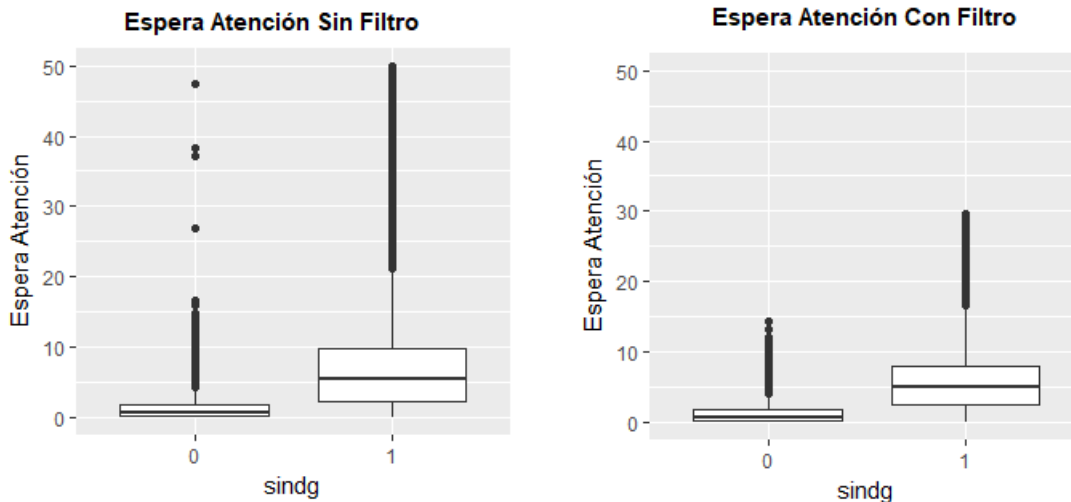


Figura 3.2: Boxplot tiempo de espera según diagnóstico

Por otra parte, aún es posible observar que los tiempos de espera en aquellas personas que abandonan la urgencia sin diagnóstico es notoriamente mayor que el de aquellos que reciben atención, sin embargo, luego de aplicados los filtros los tiempos ya no presentan los valores outliers, que en un comienzo eran tiempos muy elevados.

## Tiempo de Uso Camilla Urgencia

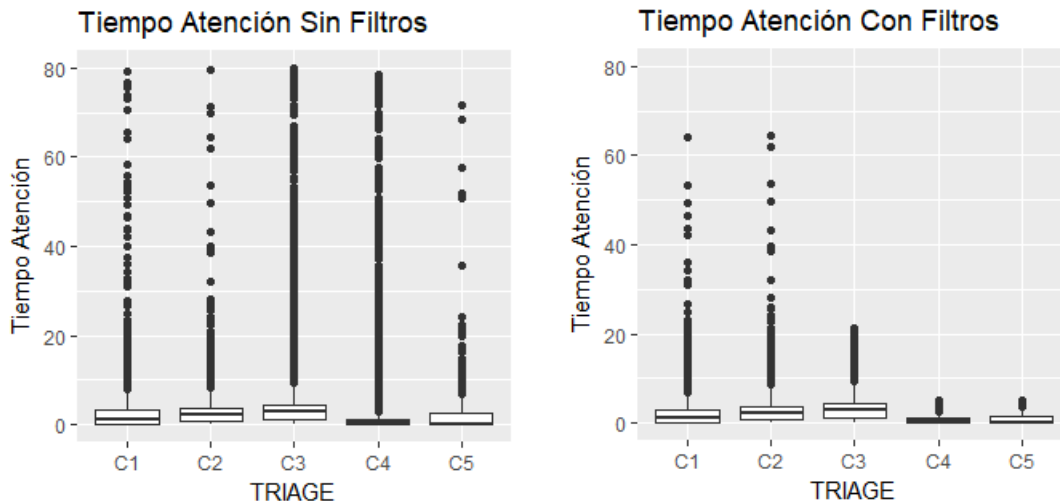


Figura 3.3: Boxplot tiempo de atención según triage

En 3.3 es posible observar que los tiempos de uso de camilla antes de aplicar los filtros, presenta valores outliers muy elevados en los pacientes C4 y C5, los cuales no deberían presentar tiempos muy elevados, pues debido a su leve gravedad no requieren de una atención muy extensa, sin embargo algún porcentaje de estos pacientes también es hospitalizado. Lo esperable en la distribución de los tiempo de uso de cama es que a mayor gravedad, permanecen más tiempo en la urgencia, lo que es posible observar una vez aplicados los filtros.

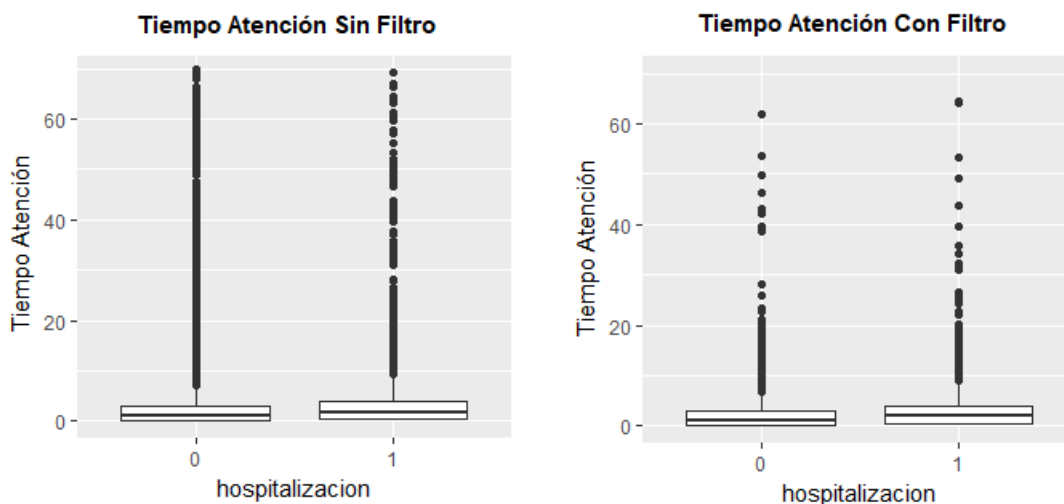


Figura 3.4: Boxplot tiempo de atención según destino

Como es de suponer, los tiempos de uso de cama en los pacientes que quedan hospitalizados son mayores que los tiempos de aquellos pacientes que no, sin embargo antes de aplicar los filtros es posible observar que hay tiempos muy elevados también en aquellos pacientes que abandonan la urgencia luego de ser atendidos, tiempos que se pueden comparar en la figura 3.5.

Finalmente se puede observar que los filtros aplicados a la base de datos cumplieron con el

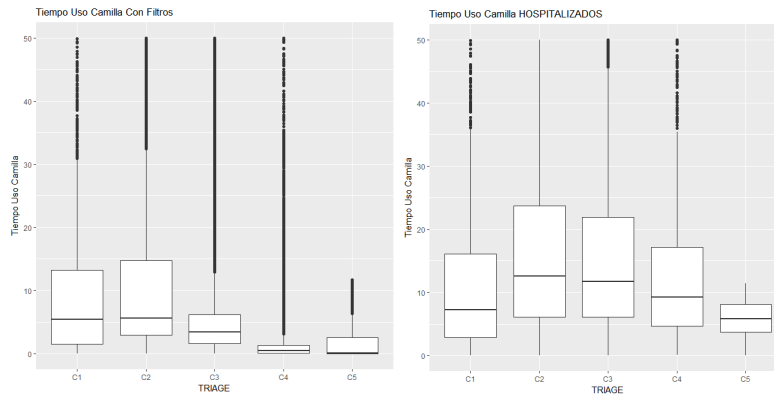


Figura 3.5: Tiempos de Espera Total y Hospitalizados

propósito de obtener tiempos, tanto de espera como de uso de cama mucho más razonables.

Una vez realizada la limpieza en la base de datos es posible proceder a construir una nueva base de tal forma que se pueda representar como una serie de tiempo y así poder calcular el índice EDWIN que permitirá finalmente medir la congestión en el HUAP.

## Capítulo 4

# Cálculo del Índice EDWIN

### 4.1. Obtención de la Serie de Tiempo

Para obtener la serie de tiempo se calculó diversas variables por hora, a partir de la base de datos formada por las observaciones que representan a cada paciente (base bruta), de tal forma de obtener las variables necesarias para calcular el índice EDWIN por hora.

Las variables calculadas fueron las siguientes:

- **Cantidad de personas esperando ser llamadas para la atención:** se cuenta a cada persona de tal forma que su tiempo de categorización es menor que una hora determinada y su tiempo de atención es mayor. De esta forma se asegura que la persona se encuentra esperando por atención en dicha hora.
- **Cantidad de personas usando una cama:** se cuenta a cada persona de tal forma que su tiempo de atención es menor que una hora determinada y su tiempo de egreso de la urgencia o de la hospitalización (en caso de ser hospitalizado) es mayor que esta. De esta forma se asegura que la persona se encuentra haciendo uso de una cama de la urgencia en una hora determinada.
- **Tiempo mediano de espera de atención:** Si la hora de categorización es menor que una hora determinada y la hora de atención es mayor que esta hora, entonces se calcula la mediana del tiempo de espera de los sujetos que cumplen la condición
- **Tiempo mediano de uso de cama:** Si la hora de atención es menor que una hora determinada y la hora de egreso de la urgencia o de hospitalización es mayor que esta hora, entonces se calcula la mediana del tiempo de uso de cama de los sujetos que cumplen la condición.
- **Cantidad de pacientes usando una cama con triage C1:** se cuenta a cada persona de tal forma que su tiempo de atención es menor que una hora determinada y su tiempo de egreso de la urgencia o de la hospitalización es mayor que esta y su categoría triage es C1. De esta forma se asegura que la persona se encuentra haciendo uso de una cama de la urgencia en una hora determinada.
- **Cantidad de pacientes usando una cama con triage C2:** se cuenta a cada persona de tal forma que su tiempo de atención es menor que una hora determinada y su tiempo de egreso de la urgencia o de la hospitalización es mayor que esta y su categoría triage es C2. De esta forma se asegura que la persona se encuentra haciendo uso de una cama de la urgencia en una hora determinada.

- **Cantidad de pacientes usando una cama con triage C3:** se cuenta a cada persona de tal forma que su tiempo de atención es menor que una hora determinada y su tiempo de egreso de la urgencia o de la hospitalización es mayor que esta y su categoría triage es C3. De esta forma se asegura que la persona se encuentra haciendo uso de una cama de la urgencia en una hora determinada.
- **Cantidad de pacientes usando una cama con triage C4:** se cuenta a cada persona de tal forma que su tiempo de atención es menor que una hora determinada y su tiempo de egreso de la urgencia o de la hospitalización es mayor que esta y su categoría triage es C4. De esta forma se asegura que la persona se encuentra haciendo uso de una cama de la urgencia en una hora determinada.
- **Cantidad de pacientes usando una cama con triage C5:** se cuenta a cada persona de tal forma que su tiempo de atención es menor que una hora determinada y su tiempo de egreso de la urgencia o de la hospitalización es mayor que esta y su categoría triage es C5. De esta forma se asegura que la persona se encuentra haciendo uso de una cama de la urgencia en una hora determinada.
- **Cantidad de pacientes hospitalizados:** se cuenta a cada persona de tal forma que su tiempo de atención es menor que una hora determinada y su tiempo de egreso de la urgencia o de la hospitalización es mayor que esta y que cumpla la condición de ser hospitalizado. De esta forma se asegura que la persona se encuentra haciendo uso de una cama de la urgencia en una hora determinada.
- **Cantidad de pacientes sin diagnóstico:** se cuenta a cada persona de tal forma que su tiempo de categorización es menor que una hora determinada y su tiempo de atención es mayor y además que haya abandonado la urgencia sin diagnóstico.

Es importante señalar que estas variables se calcularon para cada hora de los días contenidos en el intervalo de tiempo desde el 1 de enero de 2016 al el 31 de diciembre de 2017, es decir, la serie comienza el 1 de enero de 2016 a las 00:00hrs y finaliza el 31 de diciembre de 2017 a las 23:00hrs con un total de 17544 observaciones (pertenecientes a cada hora).

De forma visual, se puede observar un extracto de la base bruta en la imagen 4.1 y de la serie de tiempo en 4.2, para poder apreciar las diferencias.

I.ESPERA	USO_CAMA	FHCATEGORIZACION	FHATENCION	FHEGRESO	FHEGRESO_HOSPI	TRIAGE
0.219166667	168.77750000	26-12-2015 11:45	26-12-2015 11:58	02-01-2016 12:45	02-01-2016 12:45	C3
0.048888889	134.48555560	27-12-2015 4:16	27-12-2015 4:19	01-01-2016 18:49	01-01-2016 18:49	C1
0.334166667	121.64750000	29-12-2015 10:21	29-12-2015 10:41	03-01-2016 11:02	03-01-2016 12:20	C3

Figura 4.1: Base de datos inicial

I.DIA.HORA	ESPERA	USO_CAMA	MED_T_ESPE	MED_T_USO	C1	C2	C3
01-01-2016 0:00	0	10	NA	5.630833	0	1	5
01-01-2016 1:00	2	9	0.27680556	6.890556	0	1	5
01-01-2016 2:00	0	10	NA	10.477778	0	2	6

Figura 4.2: Base Serie de Tiempo por hora

## 4.2. Cálculo del Índice EDWIN por hora

Una vez obtenida la serie de tiempo es posible calcular el índice EDWIN para analizar la congestión en el HUAP por hora.

Este índice se calculó mediante la siguiente fórmula:

$$\frac{\sum_{i=1}^5 n_i t_i}{N_a(B_T - B_A)}$$

Donde,

- $n_i$ : Cantidad de pacientes admitidos en la urgencia por categoría triage  $i$ ,  $i = 1, \dots, 5$ .
- $t_i$ : Categoría triage.
- $N_a$ : Cantidad de médicos de turno en el Servicio de Urgencia
- $B_T$ : Número total de camas de tratamiento en la Urgencia.
- $B_A$ : Número de pacientes admitidos en el servicio que están utilizando una cama.

Como se mostró en la sección anterior, ya se cuenta con  $n_i$ ,  $t_i$  y  $B_A$  para el cálculo del índice. En cuanto a la cantidad de médicos de turno  $N_a$ , se asume que estos son constantes en el tiempo, siendo un total de 5 médicos en el servicio.

Por otra parte el número total de camas  $B_T$  en la urgencia es 57, sin embargo, cuando el servicio se encuentra congestionado, se puede habilitar camas supernumerarias, llegando como máximo a un total de 61 camas.

Se definen tres niveles de congestión, definidos en los siguientes intervalos:

- $0 < EDWIN < 1,5$ : Bajo Control
- $1,5 \leq EDWIN \leq 2$ : Ocupado
- $EDWIN > 2$ : Congestionado

#### 4.2.1. Análisis Descriptivo EDWIN

Una vez calculado el índice que determina el estado de congestión en la urgencia, es posible observar que en general son pocos los momentos en que la urgencia está congestionada, más aun, en la tabla 4.1 se puede observar que son más los periodos en que el área de adultos del HUAP se encuentra bajo control.

Estado Urgencia	Frecuencia
Bajo Control	16340
Ocupado	777
Congestión	426

Cuadro 4.1: Estado Urgencia según EDWIN por día

#### Análisis EDWIN por Año

En cuanto al índice de congestión por año, se aprecia en la figura 4.3 que el 2017 presenta valores notoriamente más altos del EDWIN que el 2016.

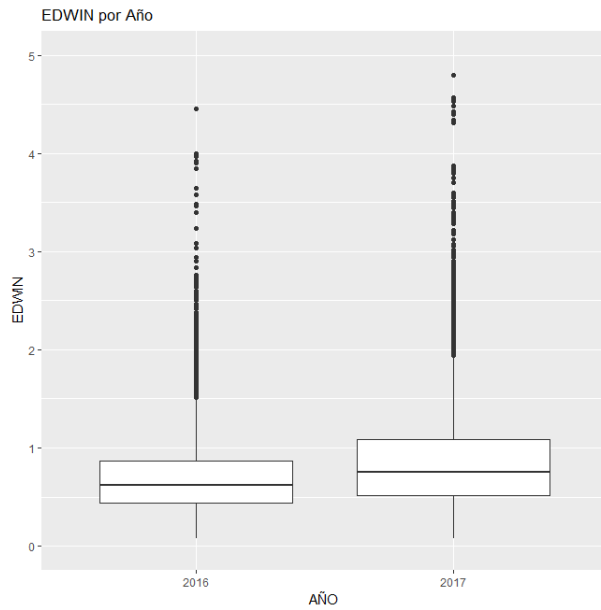


Figura 4.3: EDWIN por año

### Análisis EDWIN por Mes

Al hacer el análisis por mes, se observa un notorio aumento del índice en el mes de junio, seguido por julio, meses que concuerdan con los periodos de bajas temperaturas y en que aumentan las enfermedades respiratorias.

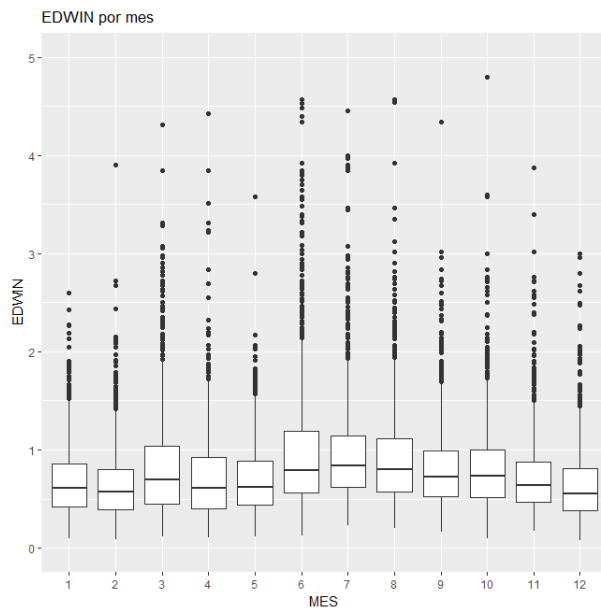


Figura 4.4: EDWIN por Mes

### Estado de la Urgencia por día

Como se aprecia en 4.5, el día de la semana que presenta mayores valores del índice es el martes, aunque no se observa una gran diferencia a lo largo de los días de la semana, sin embargo también se puede observar que el fin de semana presenta menor congestión.

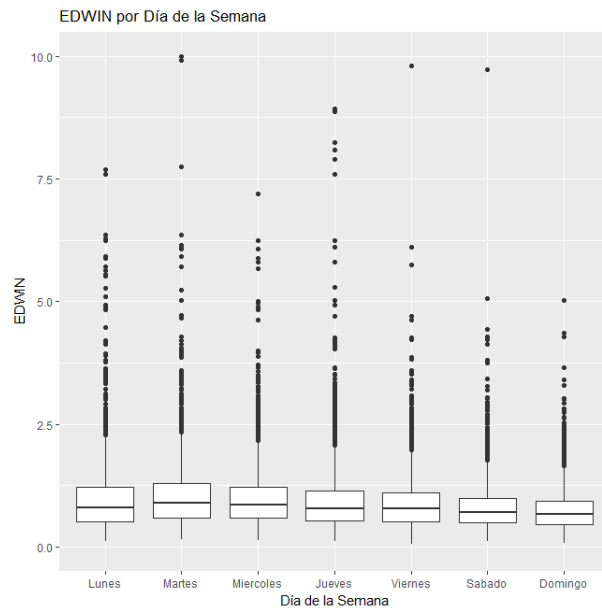


Figura 4.5: Boxplot EDWIN según día de la semana

### Análisis EDWIN por Tramo horario

Finalmente, como es de interés predecir la congestión para algunas horas futuras, es importante analizar el EDWIN por tramo horario, pues con esto podemos ver de forma preliminar cómo es el comportamiento del índice a lo largo del día.

En la figura 4.6 es posible observar que a partir de las 12 del día, los valores del EDWIN presentan un aumento considerable y por el contrario, la mañana (de 6 a 11 am) presenta los niveles de congestión más bajos.

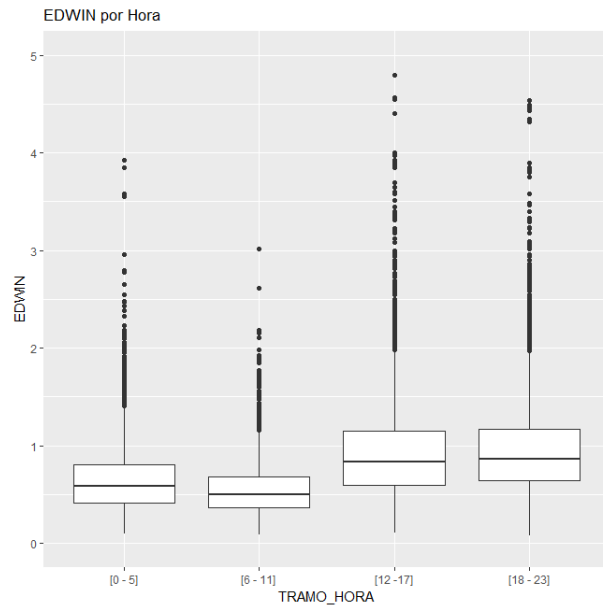


Figura 4.6: EDWIN por Hora

# Capítulo 5

## Analisis de la serie de tiempo

A partir de la gráfica de la serie de tiempo del índice de congestión, es posible confirmar los análisis anteriores, es decir, se puede observar que los meses con niveles de mas altos comprenden junio de 2016 y junio-julio de 2017, también es posible observar que hacia finales del año 2017 también se ven aumentados los momentos de congestión.

También se observan periodos en que el índice alcanza valores cercanos a 35, los que se producen al tener todas las camillas utilizadas y gran cantidad de pacientes con mayor gravedad admitidos en la urgencia

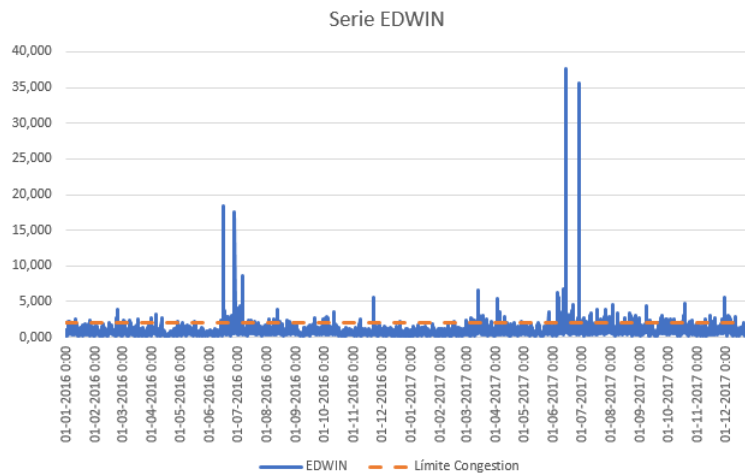


Figura 5.1: Gráfica de la Serie de tiempo EDWIN

### 5.1. Análisis de ACF EDWIN

A partir de la función de autocorrelación del índice EDWIN, se puede observar que la serie podría presentar una componente estacional, la que a simple vista parece tener periodo de 24 horas.

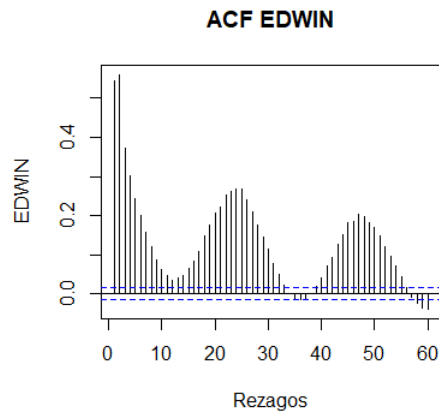


Figura 5.2: Gráfica de la Serie de tiempo EDWIN

En la figura 5.4 se puede observar que la frecuencia en que se alcanza el *peak*, está cercana a 0,04, lo que se puede constatar al calcular este valor a partir del espectro, el que arroja 0,0041667, entonces la hacer  $1/0,0041667$ , se obtiene que el periodo es 24, tal como sugiere la ACF.

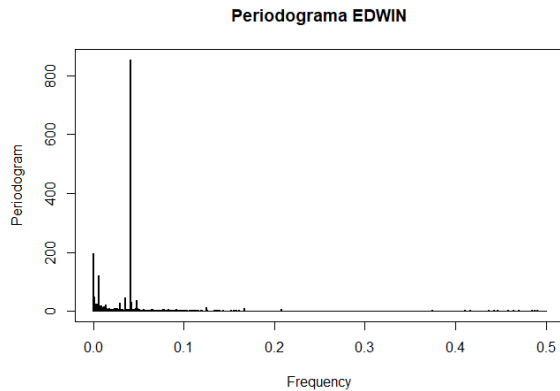


Figura 5.3: Periodograma EDWIN

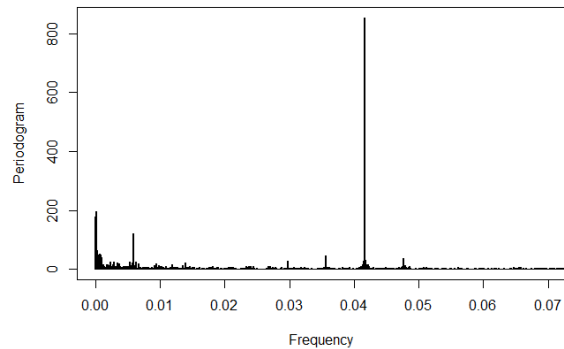


Figura 5.4: Zoom periodograma EDWIN

## 5.2. Análisis de CCF para el Índice de Congestión

Una vez hecho el análisis descriptivo del índice, es necesario analizar cuáles son las posibles covariables del EDWIN, es por esto que a continuación se hará un análisis de de la función de correlación cruzada o *CCF* por sus siglas en inglés.

Estas gráficas ayudarán a comprender cómo afectan las variables en el índice de congestión según los tiempos.

### 5.2.1. Cantidad de Pacientes en Espera de Atención

La correlación del índice EDWIN con la cantidad de pacientes en espera de atención es de 0,28 y al analizar la ccf es posible observar que alcanza el valor máximo en el lag 1, es decir, el índice se ve afectado una hora después por esta variable.

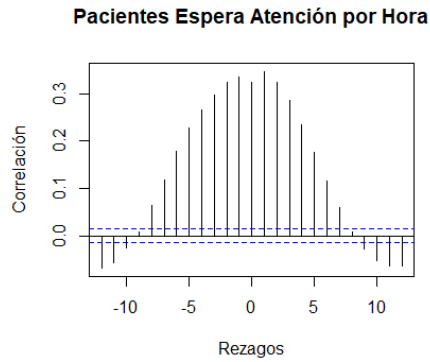


Figura 5.5: CCF Pacientes Espera Atención

### 5.2.2. Tiempo Mediano Espera

La correlación del índice EDWIN con el tiempo mediano de espera de atención es de 0,069, por lo que la correlación entre el índice y el tiempo mediano de espera es muy bajo y al observar la CCF se hace difícil hacer una interpretación.

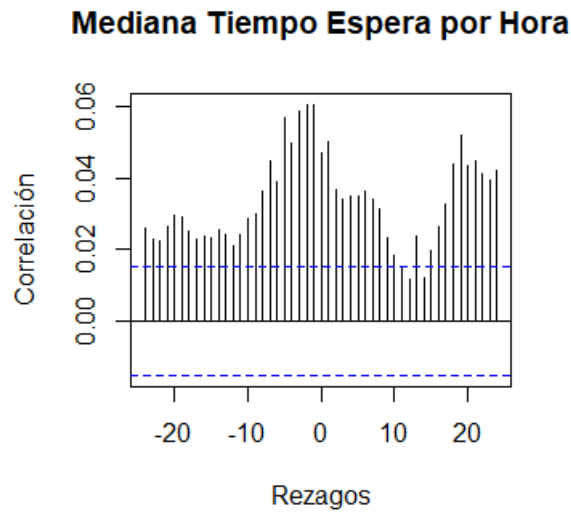


Figura 5.6: CCF Tiempo Mediano Espera

### 5.2.3. Tiempo Mediano de Uso de Cama

La correlación del índice EDWIN con el tiempo mediano de uso de camilla es de 0,036, valor que es muy bajo, por lo que implica que hay muy poca correlación entre el índice y el tiempo mediano del uso de cama y mirar la CCF, al igual con el tiempo mediano de espera, se hace difícil hacer una interpretación.

### Tiempo Mediano Atención por Hora

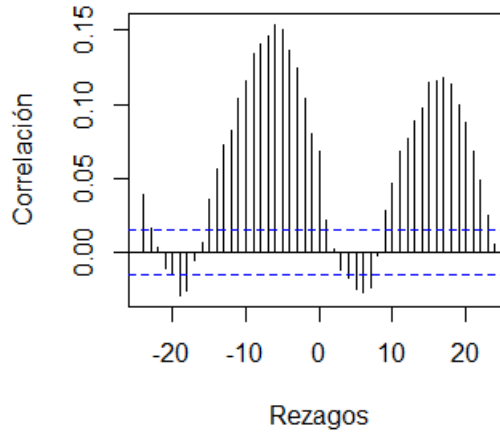


Figura 5.7: CCF Tiempo Mediano de Uso de Cama

#### 5.2.4. Cantidad de Pacientes Hospitalizados

La correlación del índice y la cantidad de pacientes hospitalizados es de 0,38, esta es la correlación más alta entre todas las variables en análisis y al observar la CCF en 5.8 es posible ver que la correlación máxima se alcanza en el lag 0, por lo que esta variable influye en el mismo tiempo al EDWIN.

### Pacientes Hospitalizados por Hora

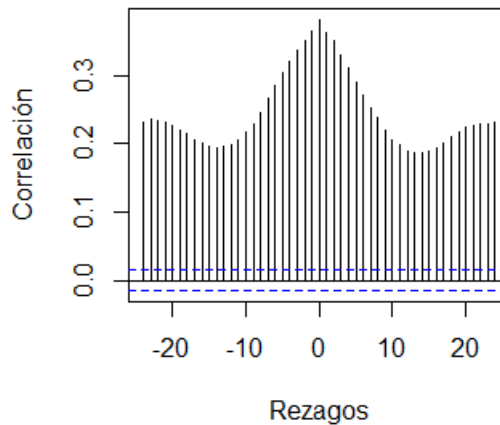


Figura 5.8: CCF Pacientes hospitalizados

#### 5.2.5. Cantidad de Pacientes Sin Diagnóstico

La correlación del índice y la cantidad de pacientes sin diagnóstico es de 0,22, valor que refleja una baja correlación con el índice. Por otra parte, al analizar la CCF en 5.9 es posible ver que la correlación máxima se alcanza en el lag 0, por lo que esta variable influye en el mismo tiempo al EDWIN.

### Pacientes Sin Diagnóstico por Hora

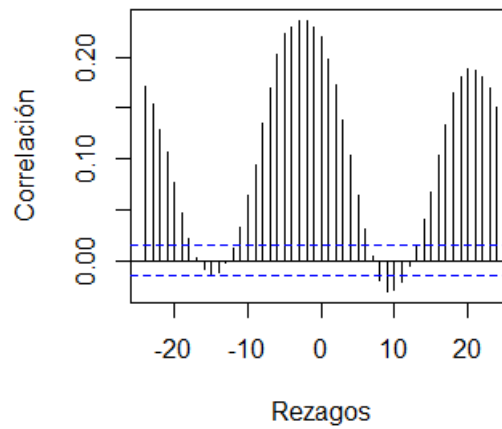


Figura 5.9: CCF Pacientes sin diagnóstico

A partir de lo observado en el análisis de correlaciones, se puede concluir que las variables con mayor relación son:

- *Cantidad de pacientes en espera*, la cual alcanza el valor máximo en el lag 1.
- *Cantidad de pacientes hospitalizados*, la cual alcanza el valor máximo de correlación en el lag 0.
- *Cantidad de pacientes sin diagnóstico*, que alcanza el máximo de correlación en el lag 0.

# Capítulo 6

## Ajuste de Modelos Lineales Dinámicos

A continuación se presentan los modelos dinámicos a ajustar, los cuáles tendrán como covariables:

- $X_1$  : Cantidad de pacientes en espera.
- $X_2$  : Cantidad de pacientes hospitalizados.
- $X_3$  : Cantidad de pacientes sin diagnóstico.

### 6.1. Estructura de los modelos propuestos

Se proponen 5 modelos a ajustar.

#### Modelo 1

El primer modelo propuesto contempla las mismas variables utilizadas en el modelo 1, pero esta vez la variable  $X_1$  estará rezagada en 1.

La estructura viene dada por:

$$Y_t = \beta_{0,t} + \beta_{1,t}X_{1,t-1} + \beta_{2,t}X_{2,t} + \beta_{3,t}X_{3,t}$$

- Ecuación de Estado:

$$\beta_{0,t} = \beta_{0,t-1} + \omega_{0,t} \quad \omega_{0,t} \sim N(0, W_0)$$

$$\beta_{1,t} = \beta_{1,t-1} + \omega_{1,t} \quad \omega_{1,t} \sim N(0, W_1)$$

$$\beta_{2,t} = \beta_{2,t-1} + \omega_{2,t} \quad \omega_{2,t} \sim N(0, W_2)$$

$$\beta_{3,t} = \beta_{3,t-1} + \omega_{3,t} \quad \omega_{3,t} \sim N(0, W_3)$$

- Matrices de definición:

$$\theta_t = (\beta_{0,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t})'$$

$$F_t = (1 + X_{1,t-1}, X_{2,t}, X_{3,t})'$$

$$G_t = \text{diag}\{1, 1, 1, 1\}_{4 \times 4}$$

## Modelo 2

El modelo 2 es propuesto por la necesidad de anticiparse a los periodos de congestión, por lo que se agregarán las 3 covariables rezagadas en una hora.

Luego, el modelo queda representado por la siguiente estructura:

$$Y_t = \beta_{0,t} + \beta_{1,t}X_{1,t-1} + \beta_{2,t}X_{2,t-1} + \beta_{3,t}X_{3,t-1}$$

- Ecuación de Estado:

$$\beta_{0,t} = \beta_{0,t-1} + \omega_{0,t} \quad \omega_{0,t} \sim N(0, W_0)$$

$$\beta_{1,t} = \beta_{1,t-1} + \omega_{1,t} \quad \omega_{1,t} \sim N(0, W_1)$$

$$\beta_{2,t} = \beta_{2,t-1} + \omega_{2,t} \quad \omega_{2,t} \sim N(0, W_2)$$

$$\beta_{3,t} = \beta_{3,t-1} + \omega_{3,t} \quad \omega_{3,t} \sim N(0, W_3)$$

- Matrices de definición:

$$\theta_t = (\phi_{1,t}, \beta_{0,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t})'$$

$$F_t = (1, X_{1,t-1}, X_{2,t-1}, X_{3,t-1})'$$

$$G_t = \text{diag}\{1, 1, 1, 1, 1\}_{4 \times 4}$$

## Modelo 3

Otro modelo lineal dinámico propuesto contempla las mismas variables utilizadas en el modelo 2, pero en este caso se agregará un parámetro Auto-Regresivo, es decir, se quiere explicar la congestión mediante las covariables y además con la misma variable respuesta en un tiempo anterior, para este caso.

Lo anterior queda representado por la siguiente estructura:

$$Y_t = \phi_{1,t}Y_{t-1} + \beta_{0,t} + \beta_{1,t}X_{1,t-1} + \beta_{2,t}X_{2,t} + \beta_{3,t}X_{3,t}$$

- Ecuación de Estado:

$$\phi_{1,t} = \phi_{1,t-1} + \omega_{\phi_{1,t}} \quad \omega_{\phi_{1,t}} \sim N(0, W_{\phi_1})$$

$$\beta_{0,t} = \beta_{0,t-1} + \omega_{0,t} \quad \omega_{0,t} \sim N(0, W_0)$$

$$\beta_{1,t} = \beta_{1,t-1} + \omega_{1,t} \quad \omega_{1,t} \sim N(0, W_1)$$

$$\beta_{2,t} = \beta_{2,t-1} + \omega_{2,t} \quad \omega_{2,t} \sim N(0, W_2)$$

$$\beta_{3,t} = \beta_{3,t-1} + \omega_{3,t} \quad \omega_{3,t} \sim N(0, W_3)$$

- Matrices de definición:

$$\theta_t = (\phi_{1,t}, \beta_{0,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t})'$$

$$F_t = (1, Y_{t-1}, X_{1,t-1}, X_{2,t}, X_{3,t})'$$

$$G_t = \text{diag}\{1, 1, 1, 1, 1\}_{5 \times 5}$$

## Modelo 4

En vista del resultado entregado por el periodograma en la sección 5.2, se agregará una componente estacional de periodo 24 al cuarto modelo, donde sólo la variable  $X_1$  estará rezagada en una hora.

El modelo lineal dinámico con la componente estacional queda representado de la siguiente forma:

$$Y_t = \phi_{1,t}Y_{t-1} + \phi_{2,t}Y_{t-24} + \beta_{0,t} + \beta_{1,t}X_{1,t-1} + \beta_{2,t}X_{2,t} + \beta_{3,t}X_{3,t}$$

- Ecuación de Estado:

$$\phi_{1,t} = \phi_{1,t-1} + \omega_{\phi_{1,t}} \quad \omega_{\phi_{1,t}} \sim N(0, W_{\phi_1})$$

$$\phi_{2,t} = \phi_{2,t-1} + \omega_{\phi_{2,t}} \quad \omega_{\phi_{2,t}} \sim N(0, W_{\phi_2})$$

$$\beta_{0,t} = \beta_{0,t-1} + \omega_{\beta_{0,t}} \quad \omega_{\beta_{0,t}} \sim N(0, W_0)$$

$$\beta_{1,t} = \beta_{1,t-1} + \omega_{\beta_{1,t}} \quad \omega_{\beta_{1,t}} \sim N(0, W_1)$$

$$\beta_{2,t} = \beta_{2,t-1} + \omega_{\beta_{2,t}} \quad \omega_{\beta_{2,t}} \sim N(0, W_2)$$

$$\beta_{3,t} = \beta_{3,t-1} + \omega_{\beta_{3,t}} \quad \omega_{\beta_{3,t}} \sim N(0, W_3)$$

- Matrices de definición:

$$\theta_t = (\phi_{1,t}, \phi_{2,t}, \beta_{0,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t})'$$

$$F_t = (1, Y_{t-1}, Y_{t-24}, X_{1,t-1}, X_{2,t}, X_{3,t})'$$

$$G_t = \text{diag}\{1, 1, 1, 1, 1, 1\}_{6 \times 6}$$

# Modelo 5

Finalmente se ajustará un modelo agregando la componente estacional de periodo 24 horas y las 3 covariables rezagadas en una hora, pues se quiere anticipar la congestión a partir de lo que se observa hasta el momento en que se obtienen las predicciones para horas futuras.

Este último modelo con la componente estacional y las covariables rezagadas queda dado por la siguiente estructura:

$$Y_t = \phi_{1,t}Y_{t-1} + \phi_{2,t}Y_{t-24} + \beta_{0,t} + \beta_{1,t}X_{1,t-1} + \beta_{2,t}X_{2,t-1} + \beta_{3,t}X_{3,t-1}$$

- Ecuación de Estado:

$$\phi_{1,t} = \phi_{1,t-1} + \omega_{\phi_{1,t}} \quad \omega_{\phi_{1,t}} \sim N(0, W_{\phi_1})$$

$$\phi_{2,t} = \phi_{2,t-1} + \omega_{\phi_{2,t}} \quad \omega_{\phi_{2,t}} \sim N(0, W_{\phi_2})$$

$$\beta_{0,t} = \beta_{0,t-1} + \omega_{\beta_{0,t}} \quad \omega_{\beta_{0,t}} \sim N(0, W_0)$$

$$\beta_{1,t} = \beta_{1,t-1} + \omega_{\beta_{1,t}} \quad \omega_{\beta_{1,t}} \sim N(0, W_1)$$

$$\beta_{2,t} = \beta_{2,t-1} + \omega_{\beta_{2,t}} \quad \omega_{\beta_{2,t}} \sim N(0, W_2)$$

$$\beta_{3,t} = \beta_{3,t-1} + \omega_{\beta_{3,t}} \quad \omega_{\beta_{3,t}} \sim N(0, W_3)$$

- Matrices de definición:

$$\theta_t = (\phi_{1,t}, \phi_{2,t}, \beta_{0,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t})'$$

$$F_t = (1, Y_{t-1}, X_{1,t-1}, X_{2,t-1}, X_{3,t-1})'$$

$$G_t = \text{diag}\{1, 1, 1, 1, 1, 1\}_{6 \times 6}$$

## 6.2. Estimación de parámetros

Conocidas las estructuras de los modelos a ajustar, se deben ejecutar los algoritmos en los que se programan las técnicas de estimación vistas en el Capítulo 2. En el caso de los 5 modelos, se ejecuta para cadenas de largo 10000.

Debido a la gran cantidad de observaciones en la serie y el tiempo computacional que esto implica, no se ejecutan más iteraciones.

Para evitar la dependencia que se puede dar entre la información inicial y los resultados, se aplica un proceso, donde se eliminan las 5000 primeras estimaciones de la cadena y se toma una muestra cada 20 dentro de esta, para cada uno de los parámetros a estimar.

Para los 2 primeros modelos, los parámetros a estimar son los siguientes:

- Varianza observacional  $V$
- Varianzas evolutivas  $W$
- Parámetros asociados a las covariables  $\beta_{i,t}, i = 0, \dots, 3$

En el caso del modelo 3, los parámetros a estimar son los siguientes:

- Varianza observacional  $V$
- Varianzas evolutivas  $W$
- Parámetros asociados a las covariables  $\beta_{i,t}, i = 0, \dots, 3$
- Parámetro de la componente Auto-Regresiva  $\phi_{1,t}$

En donde, si bien se presenta el parámetro Auto-Regresivo como  $\phi_{1,t}$ , la estimación de este se hace de igual forma que los parámetros asociados a las variables independientes, por lo que también es parte de los  $\beta$  al momento de estimar.

Finalmente para los modelos 4 y 5, los parámetros a estimar son:

- Varianza observacional  $V$
- Varianzas evolutivas  $W$
- Parámetros asociados a las covariables  $\beta_{i,t}, i = 0, \dots, 3$
- Parámetro Auto-Regresivo  $\phi_{1,t}$
- Parámetro de la componente estacional  $\phi_{2,t}$

Al igual que para el modelo 3, los parámetros  $\phi_{1,t}$  y  $\phi_{2,t}$ , se estiman en forma conjunta con los parámetros asociados a las variables independientes.

## Modelo 1

Para el modelo 1, si bien, el análisis de CCF indica que las variables seleccionadas tienen correlación significativa con el índice EDWIN, se puede observar que los parámetros que acompañan a estas variables no son significativos y presentan autocorrelación.

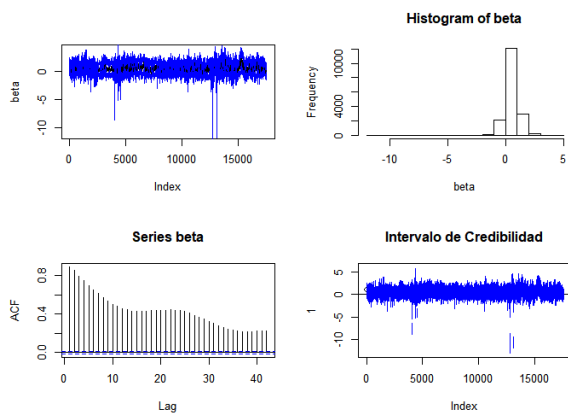


Figura 6.1:  $\beta_0$  Modelo 1

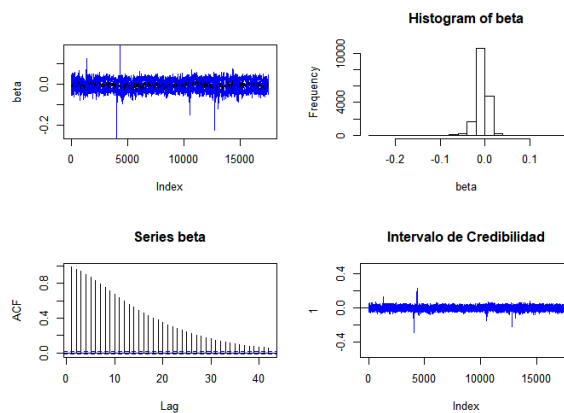


Figura 6.2:  $\beta_1$  Modelo 1

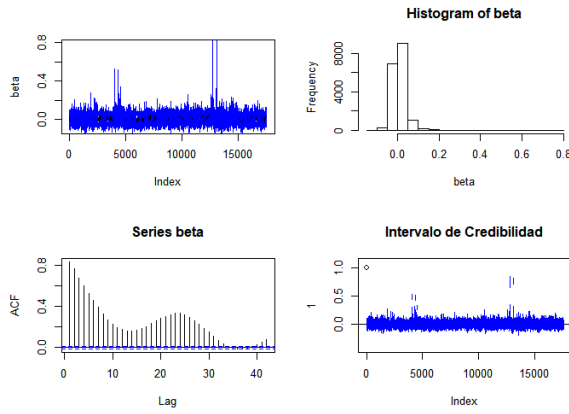


Figura 6.3:  $\beta_2$  Modelo 1

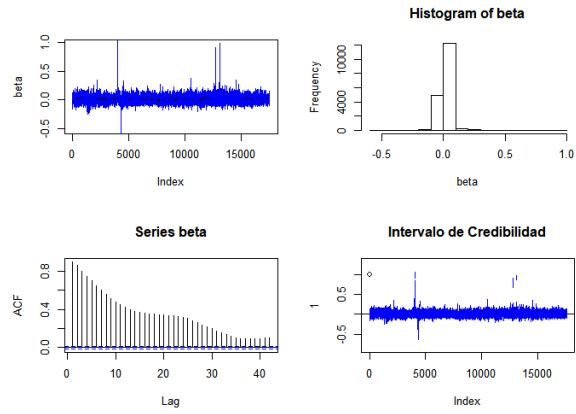


Figura 6.4:  $\beta_3$  Modelo 1

Es posible observar entonces que el modelo 1 no cumple con los supuestos, por lo que no podría ser utilizado para realizar predicciones.

## Modelo 2

Al igual que con el modelo 1, los parámetros no son significativos, y presentan autocorrelación, por lo que de la misma forma, este modelo es desechado.

De todas formas al observar las estimaciones de las varianzas, en la tabla 6.1, es posible ver que se encuentran dentro de los intervalos de confianza correspondientes.

	Estimado	Intervalo de credibilidad	
$W_{\beta_0}$	1,62E-01	1,53E-01	1,86E-01
$W_{\beta_1}$	9,43E-04	8,79E-04	1,04E-04
$W_{\beta_2}$	6,24E-04	5,84E-04	6,86E-04
$W_{\beta_3}$	9,31E-04	8,48E-04	1,03E-03
$V$	2,28E-04	2,24E-04	2,32E-04

Cuadro 6.1: Estimación parámetros modelo 2

## Modelo 3

Para el caso del modelo 3, al obtener las estimaciones de los parámetros asociados a las covariables, al igual que en los modelos anteriores, se observa que estos no son significativos, excepto por  $\phi_1$  (figura 6.7) y siguen teniendo autocorrelación. Sin embargo se puede ver que en el caso de  $\beta_1$  (figura 6.9) y  $\beta_3$  (figura 6.11) las autocorrelaciones han disminuido.

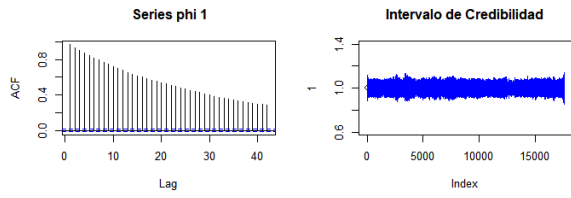


Figura 6.5:  $\phi_1$  Modelo 3

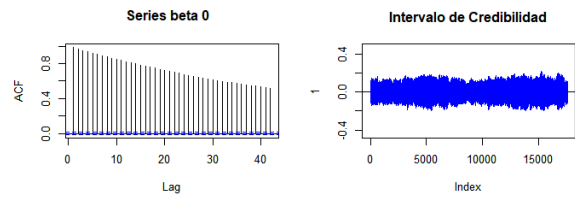


Figura 6.6:  $\beta_0$  Modelo 3

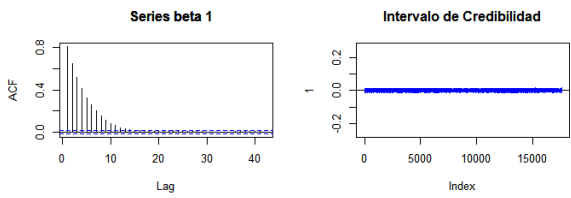


Figura 6.7:  $\beta_1$  Modelo 3

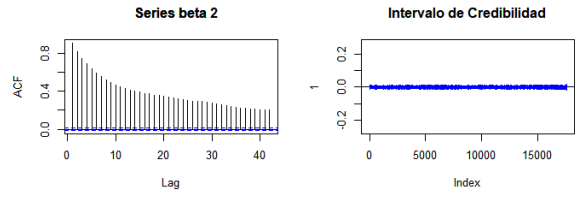


Figura 6.8:  $\beta_2$  Modelo 3

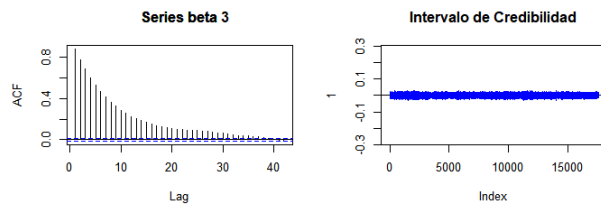


Figura 6.9:  $\beta_3$  Modelo 3

# Modelo 4

Para el caso de este modelo donde se incluye la componente estacional, al ver el diagnóstico de las estimaciones de los parámetros asociados a las covariables, el parámetro asociado a la componente auto-regresiva resulta ser significativo, sin embargo, al igual que los otros 5 parámetros sigue habiendo autocorrelación en las estimaciones.

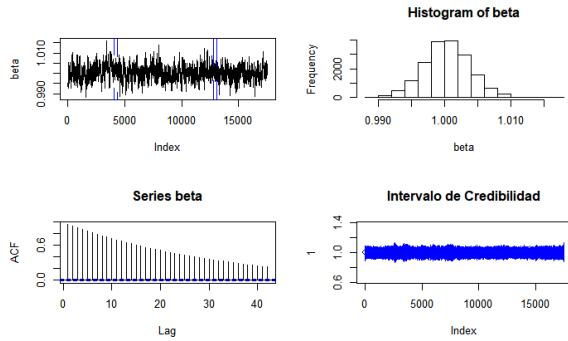


Figura 6.10:  $\phi_1$  Modelo 4

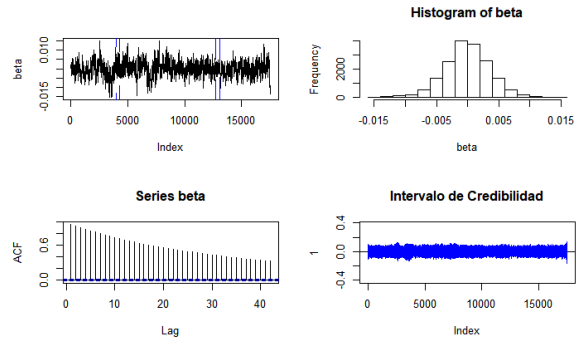


Figura 6.11:  $\phi_2$  Modelo 4

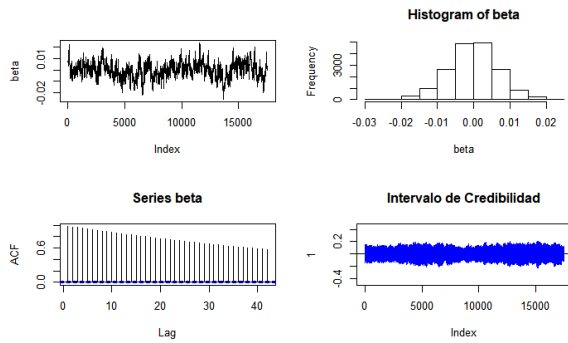


Figura 6.12:  $\beta_0$  Modelo 4

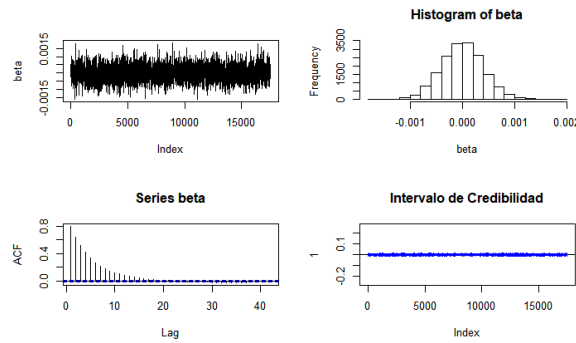


Figura 6.13:  $\beta_1$  Modelo 4

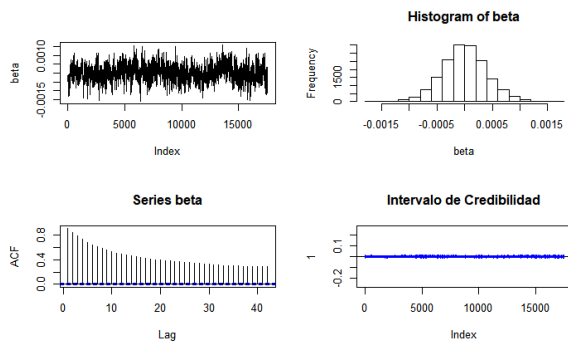


Figura 6.14:  $\beta_2$  Modelo 4

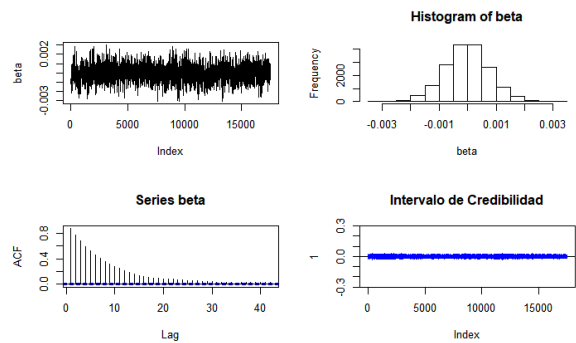


Figura 6.15:  $\beta_3$  Modelo 4

## Modelo 5

Finalmente, en el modelo 5 ocurre la misma situación que en los modelos 3 y 4, donde el parámetro correspondiente a la variable auto-regresiva resultó ser significativo, sin embargo este y los otros parámetros presentan autocorrelación.

Del mismo que para el modelo 2, se puede obtener la estimación de las varianzas, las cuáles también se muestran dentro de sus respectivos intervalos de confianza.

Parámetro	Estimación	Región de Credibilidad
$W_{\phi_1}$	1,93E-04	1,91E-04 2,46E-04
$W_{\phi_2}$	2,67E-04	2,14E-04 2,80E-04
$W_{\beta_0}$	3,73E-04	2,77E-04 3,48E-04
$W_{\beta_1}$	2,15E-05	2,10E-05 2,29E-05
$W_{\beta_2}$	1,13E-05	1,10E-05 1,18E-05
$W_{\beta_3}$	4,14E-05	4,16E-05 4,66E-05
$V$	2,27E-05	2,24E-05 2,32E-05

Cuadro 6.2: Estimación parámetros modelo 5

## DIC Modelos

A pesar de que los modelos no cumplen con los supuestos, de todas formas se puede aplicar el criterio DIC para comparar los modelos antes vistos y ver cuál de estos sería el mejor.

	$\bar{D}$	$P_d$	DIC
Modelo 1	-19082,21	109101	90018,79
Modelo 2	-42358,57	167022,2	124663,6
<b>Modelo 3</b>	-74571,48	134939	60367,48
Modelo 4	-73814,76	135343,9	61529,15
Modelo 5	-73807,3	135347,4	61540,1

Cuadro 6.3: Información sobre la Devianza

Se observa en la tabla 6.3 que el modelo 3 es el que presenta menor  $\bar{D}$  y es también el que tiene menor  $DIC$ , por lo que se podría decir que el modelo que incluye solo la componente auto-regresiva y las 3 covariables rezagadas es el mejor modelo.

En vista de que se puede decir que el tercer modelo es el mejor, se muestra el ajuste de este a los datos en la figura 6.16, donde se muestra que el ajuste es muy similar a los valores reales y el intervalo de confianza también presenta poca diferencia entre el límite inferior y superior con los datos ajustados.

Se puede observar en la figura 6.17 un *zoom* del ajuste, para poder apreciar de mejor forma las diferencias entre los valores reales, ajustados e intervalo de confianza.

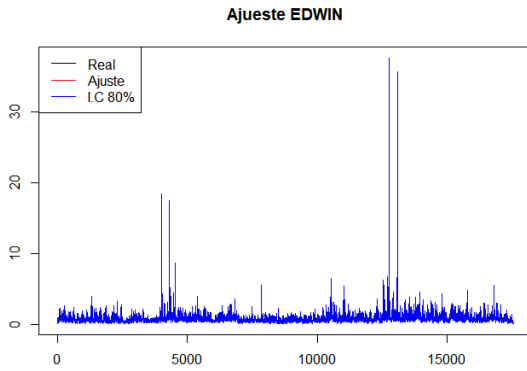


Figura 6.16: Ajuste Modelo 3

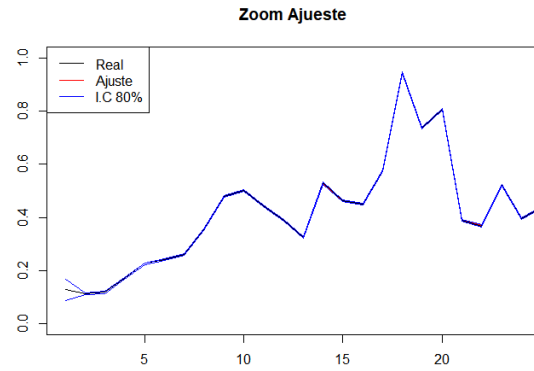


Figura 6.17: Zoom ajuste Modelo 3

Si bien se puede observar que el modelo 3 es el mejor de los 5 modelos mostrados y que el ajuste se ve similar a los valores reales, no se hará predicciones con este modelo, pues al no cumplir con los supuestos, y no presentar covariables significativas, las predicciones no serían confiables, así como tampoco lo podrían ser el ajuste visto en 6.17 ni las estimaciones de varianzas.

A pesar de no poder obtener resultados mediante los modelos antes vistos, realizando ejecuciones de cadenas de largo 10000, de todo modos se puede observar que a medida que se agregan componentes a los modelos, la correlación de algunos parámetros disminuye y se puede observar que el parámetro auto-regresivo resulta significativo en los últimos dos modelos.

Del mismo modo, al analizar la diferencia entre eliminar las primeras 4000 y 5000 estimaciones, se pudo observar que la autocorrelación de los parámetros disminuye en pequeñas cantidades cuando se eliminan los primeros 5000 valores, lo que hace pensar que al ejecutar cadenas de un largo mayor, los modelos propuestos podrían entregar mejores resultados para los parámetros estimados, lo que podría permitir realizar predicciones a través de alguno de estos modelos.

# Capítulo 7

## Estimación de Modelos ARIMAX

Los modelos ARIMAX propuestos son obtenidos a través de la función `auto.arima` presente en R, esta función proporciona una opción rápida para construir pronósticos con series temporales, debido a que evalúa entre todos los posibles modelos, al mejor modelo considerando diversos criterios: estacionariedad, estacionalidad, diferencias, entre otras.

A través de los modelos propuestos por el `auto.arima` se puede obtener luces de cuáles podrían ser los parámetros a utilizar en los modelos que incluyen las covariables.

Es por lo anterior, que se propondrán ocho modelos, en los cuáles se deberá seguir los pasos del esquema Box-Jenkins para obtener predicciones.

Además se incluirán las 3 covariables seleccionadas mediante la CCF y el análisis de las correlaciones:

- $X_1$  : Cantidad de pacientes en espera.
- $X_2$  : Cantidad de pacientes hospitalizados.
- $X_3$  : Cantidad de pacientes sin diagnóstico.

### 7.1. Ajuste de Modelos ARIMAX

A través del `auto.arima`, la estructura obtenida es  $y_t \sim ARIMA(5, 1, 0)$ , resultando las siguientes estimaciones:

	ar1	ar2	ar3	ar4	ar5
Estimación	-0.5693	-0.1136	-0.1329	-0.1819	-0.1033
s.e.	0.0075	0.0085	0.0085	0.0085	0.0075

Cuadro 7.1: Estimación ARIMA(5,1,0)

A partir de esto, se tiene a cuáles podrían ser los parámetros para ajustar los modelos agregando las variables exógenas. Entonces los modelos a estimar son los siguientes:

### Modelo 1

El primer modelo será ARIMAX(5,1,0), con la variable “Pacientes en espera” ( $X_1$ ) rezagada en una hora. Del cuál se obtienen las estimaciones observadas en la tabla 7.2.

	ar1	ar2	ar3	ar4	ar5	$X_1$	$X_2$	$X_3$
Estimación	-0.6121	-0.1717	-0.1927	-0.2357	-0.1336	0.0117	0.0720	-0.0032
s.e.	0.0075	0.0086	0.0086	0.0086	0.0075	0.0008	0.0021	0.0013

Cuadro 7.2: Estimación Modelo ARIMAX 1

En este modelo se puede ver que todas las variables son significativas, por lo que se puede realizar el análisis de los residuos para comprobar que sean ruido blanco.

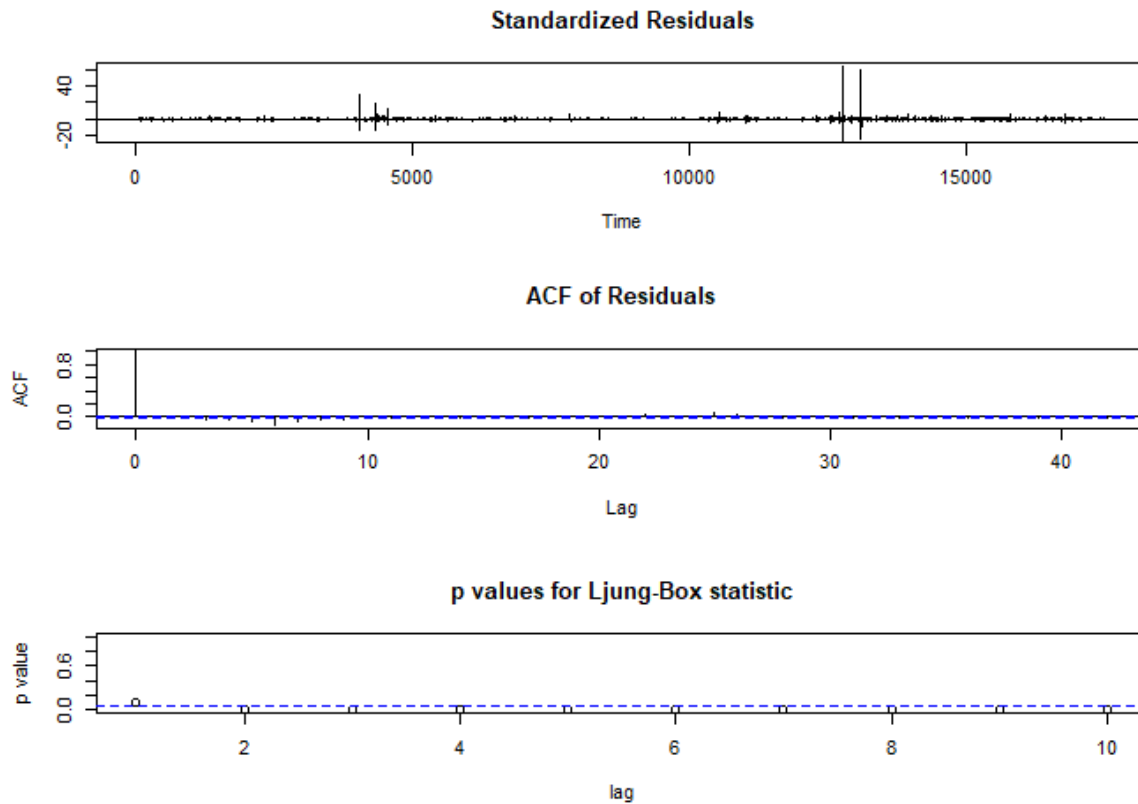


Figura 7.1: Diagnóstico residuos modelo ARIMAX 1

Se observa en la figura 7.1 que los residuos presentan correlación, pues los p-valor del test de Ljung-Box, indican que se rechaza la hipótesis nula, que dice que no existe correlación.

## Modelo 2

El segundo modelo también será ARIMAX(5,1,0), pero ahora las tres covariables serán rezagadas en una hora.

	ar1	ar2	ar3	ar4	ar5	$x_1$	$x_2$	$x_3$
Estimación	-0.6148	-0.1837	-0.1965	-0.2332	-0.1349	0.022	0.0376	-0.0203
s.e.	0.0076	0.0088	0.0087	0.0087	0.0075	0.001	0.0022	0.0016

Cuadro 7.3: Estimación Modelo ARIMAX 2

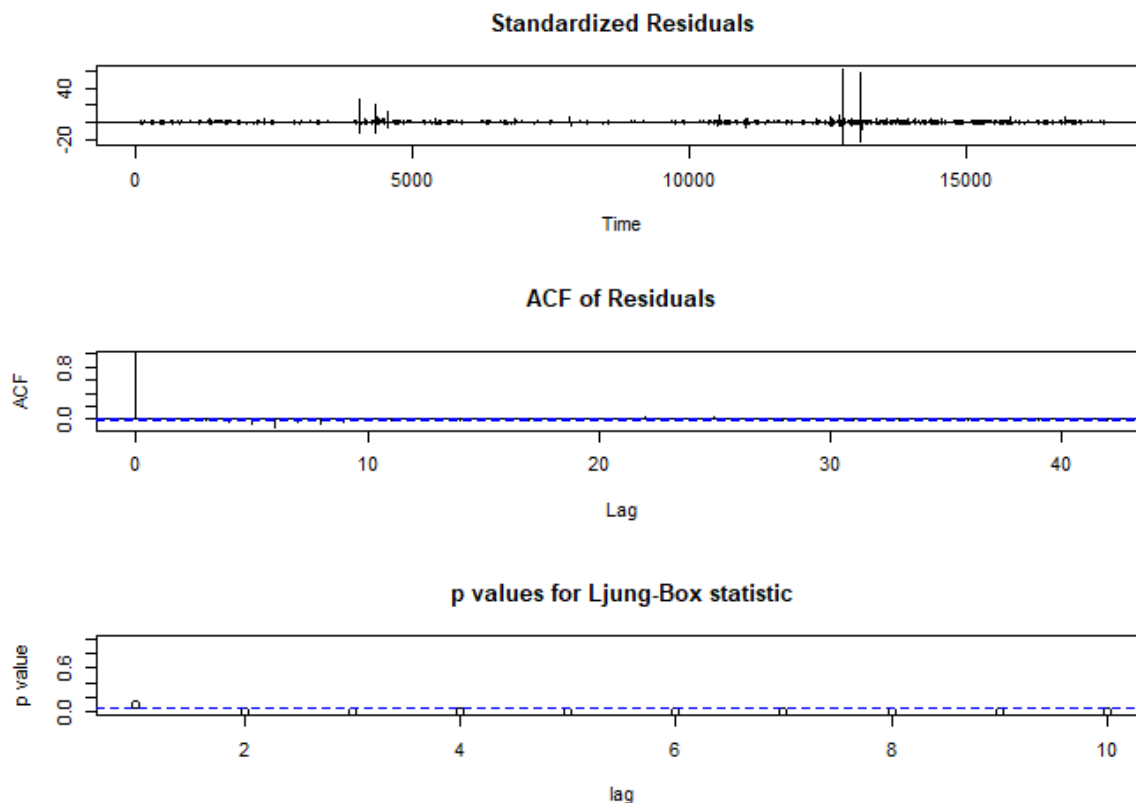


Figura 7.2: Diagnóstico residuos modelo ARIMAX 2

En este modelo, como se puede ver en la tabla 7.3, todas las variables y parámetros, son significativos, entonces se puede hacer un análisis de los residuos para verificar los supuestos.

Se puede ver en la figura 7.2, que los residuos no son ruido blanco, por lo que no sería adecuado realizar predicciones a través de este modelo.

## Modelo 3

Para el modelo 3, se agregará una componente de medias móviles, y se considerará sólo  $X_1$  con rezago en una hora, por lo que en este caso  $y_t \sim ARIMAX(5, 1, 1)$ .

	ar1	ar2	ar3	ar4	ar5	ma1	$X_1$	$X_2$	$X_3$
Estimación	0.2731	0.3535	-0.0579	-0.0735	0.0196	-0.9968	0.0125	0.0570	-0.0050
s.e.	0.0076	0.0079	0.0083	0.0078	0.0076	0.0007	0.0008	0.0014	0.0013

Cuadro 7.4: Estimación Modelo ARIMAX 3

En este modelo, se puede ver en la tabla 7.4 que todas las variables y parámetros son significativos y por otra parte, se observa en la figura 7.3 que se cumplen los supuestos.

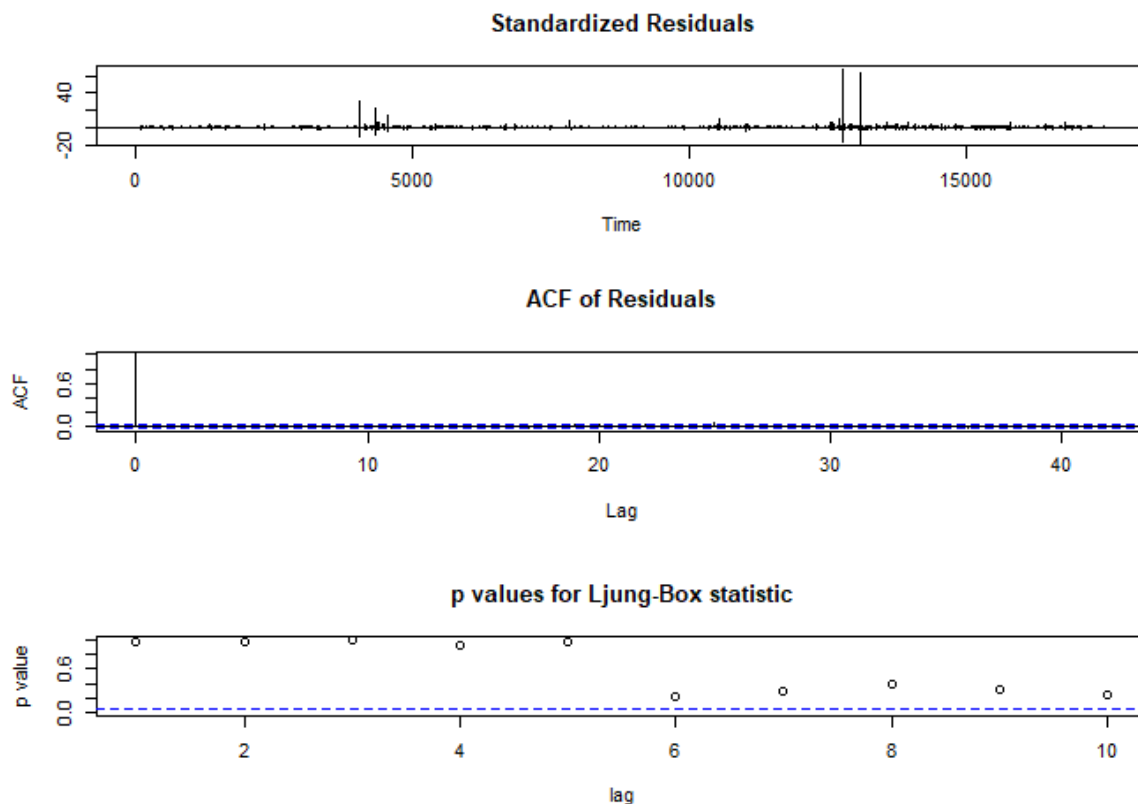


Figura 7.3: Diagnóstico residuos modelo ARIMAX 3

## Modelo 4

El cuarto modelo, igual que el modelo anterior es un ARIMAX(5,1,1), pero esta vez con las 3 covariables rezagadas en una hora.

	ar1	ar2	ar3	ar4	ar5	ma1	$X_1$	$X_2$	$X_3$
Estimación	0.2630	0.3395	-0.0530	-0.0712	0.0133	-0.9964	0.0223	0.0427	-0.0204
s.e.	0.0076	0.0079	0.0082	0.0078	0.0076	0.0008	0.0010	0.0014	0.0015

Cuadro 7.5: Estimación Modelo ARIMAX 4

Este modelo también muestra que todos los parámetros y covariables son significativos, como se aprecia en la tabla 7.6. Por otra parte, se ve en la imagen 7.4 que se cumplen los supuestos.

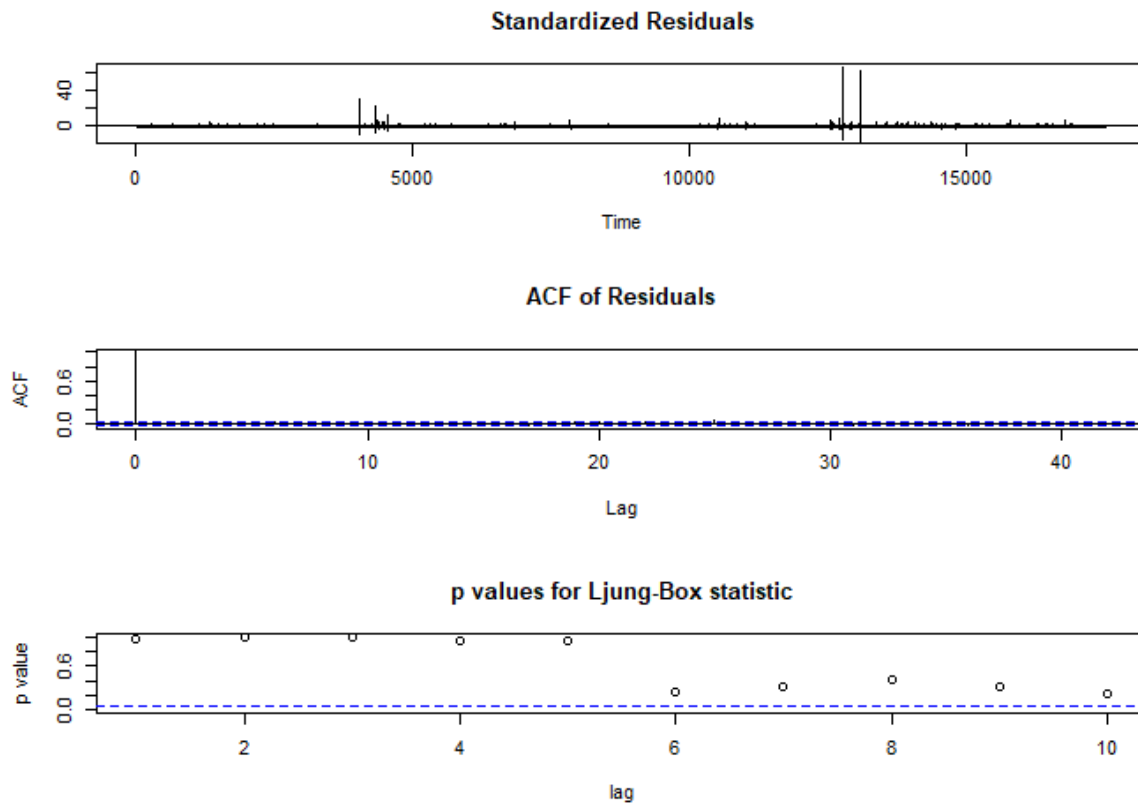


Figura 7.4: Diagnóstico residuos modelo ARIMAX 4

## 7.2. Ajuste de Modelos ARIMAX Estacionales

En vista que en el capítulo anterior se comprobó que la serie tiene una componente estacional de periodo 24, se aplica la función `auto.arima`, pero esta vez con esta componente. De esta forma se podrá obtener un modelo que pueda guiarnos en la elección de los parámetros del ARIMAX estacional.

El modelo resultante es  $SARIMA(5, 1, 0)(0, 0, 2)_{24}$  y la estimación es la que se ve en la tabla 7.6

	ar1	ar2	ar3	ar4	ar5	sma1	sma2
Estimación	-0.5780	-0.1247	-0.1419	-0.1892	-0.1079	0.0490	0.0423
s.e.	0.0076	0.0086	0.0086	0.0086	0.0075	0.0076	0.0073

Cuadro 7.6: Estimación  $SARIMA(5, 1, 0)(0, 0, 2)_{24}$

A partir del modelo  $SARIMA(5, 1, 0)(0, 0, 2)_{24}$  se tiene una noción preliminar acerca de los parámetros con que se puede ajustar modelos estacionales con variables exógenas.

## Modelo 5

Este modelo, al igual que con el modelo 1, incluirá las 3 covariables seleccionadas y se considerará un rezago en la variable “cantidad de pacientes en espera” ( $X_1$ ), como se observó en la CCF anteriormente.

El modelo a estimar es  $ARIMAX(5, 1, 0)(0, 0, 2)_{24}$ , cuya estimación se puede ver en 7.7 y donde todos los parámetros y variables resultan ser significativos.

	ar1	ar2	ar3	ar4	ar5	sma1	sma2	x1	x2	x3
Estimación	-0.6141	-0.1739	-0.1941	-0.2368	-0.1344	0.0193	0.0219	0.0114	0.0717	-0.0031
s.e.	0.0075	0.0087	0.0086	0.0086	0.0075	0.0076	0.0074	0.0008	0.0021	0.0013

Cuadro 7.7: Estimación Modelo ARIMAX 5

Por otra parte se ve en la figura 7.5 que el modelo no cumple con que los residuos son ruido blanco.

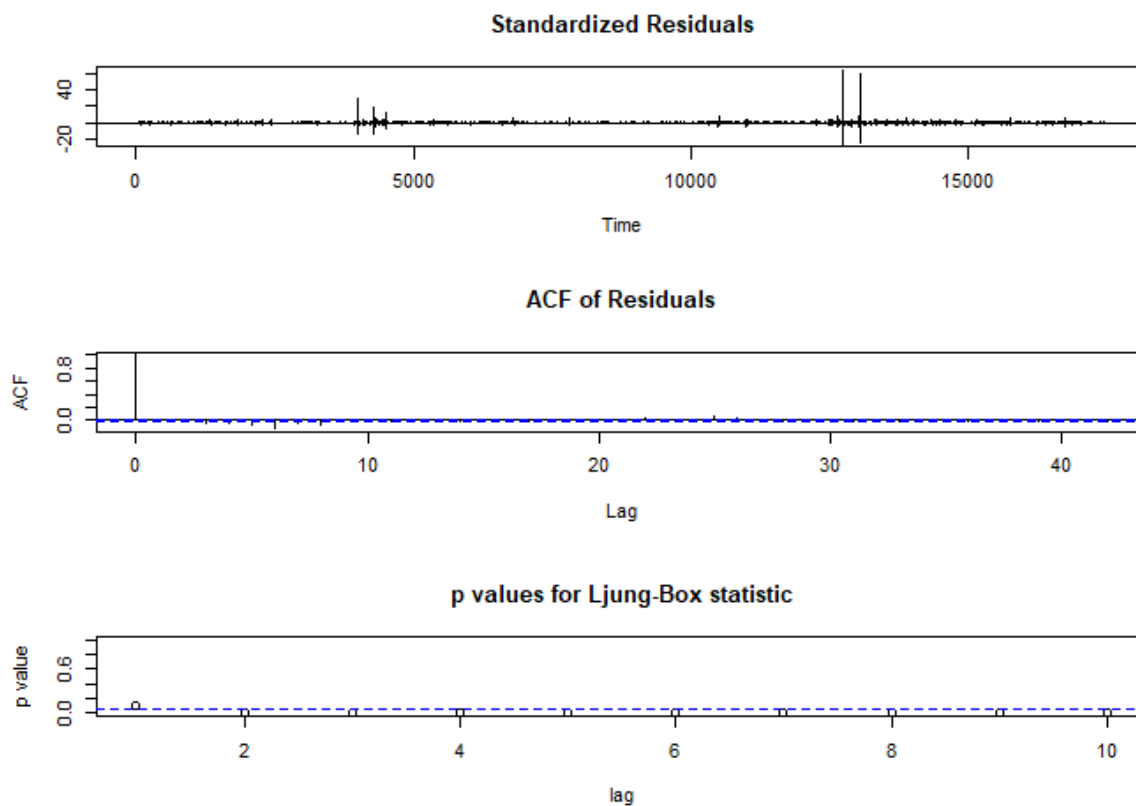


Figura 7.5: Diagnóstico residuos modelo ARIMAX 5

## Modelo 6

Para este caso el modelo será un  $ARIMAX(5, 1, 0)(0, 0, 2)_{24}$  y se tienen los mismos parámetros que en el modelo anterior, sin embargo se incluirán las 3 covariables rezagadas en una hora.

	ar1	ar2	ar3	ar4	ar5	sma1	sma2	$X_1$	$X_2$	$X_3$
Estimación	-0.5912	-0.1455	-0.1631	-0.2086	-0.1208	0.0344	0.0258	0.0134	0.0135	-0.0091
s.e.	0.0075	0.0087	0.0086	0.0086	0.0075	0.0079	0.0074	0.0010	0.0023	0.0017

Cuadro 7.8: Estimación Modelo ARIMAX 6

En la tabla 7.8 se puede ver que todos los parámetros y variables son significativos.

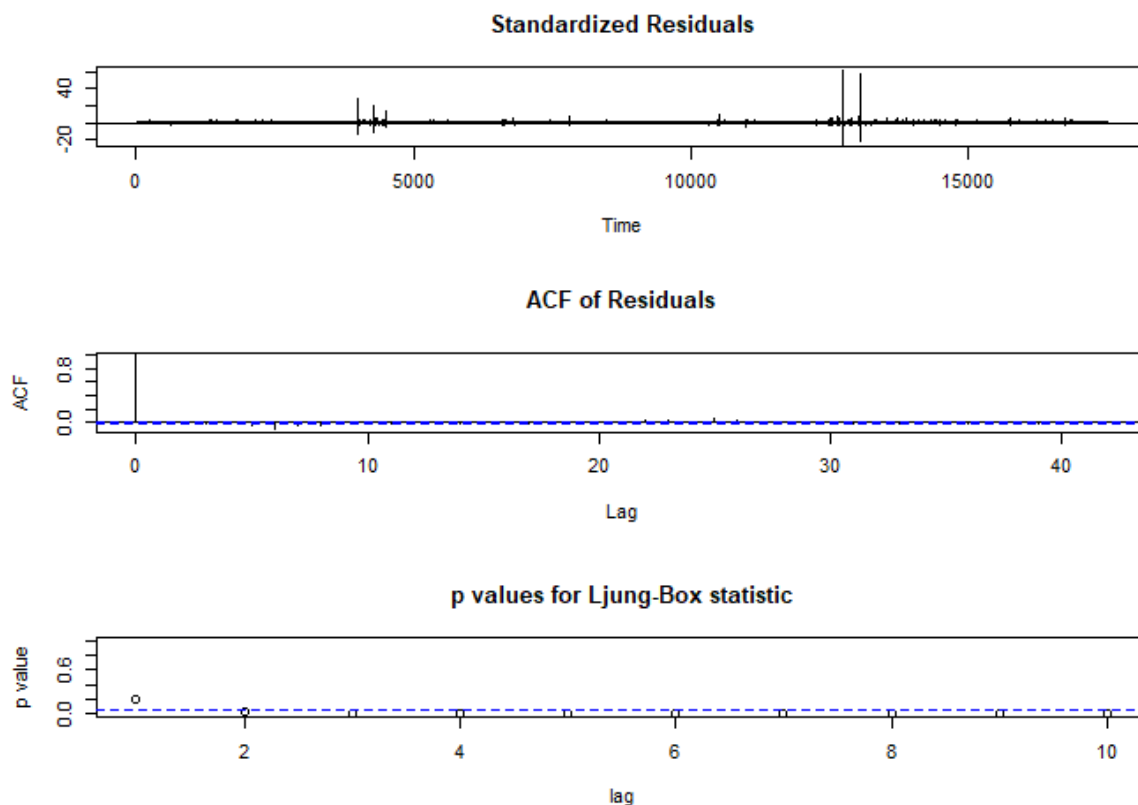


Figura 7.6: Diagnóstico residuos modelo ARIMAX 6

Entonces al observar la figura 7.6 se observa que no se cumplen los supuestos de los residuos, por lo que este modelo no sería adecuado para realizar predicciones.

## Modelo 7

De igual forma que en el modelo 3, se incluirá una componente de medias móviles al modelo ARIMAX estacional, de esta forma  $y_t \sim ARIMAX(5, 1, 1)(0, 0, 2)_{24}$ .

	ar1	ar2	ar3	ar4	ar5	ma1	sma1	sma2	$X_1$	$X_2$	$X_3$
Estimación	0.2713	0.3591	-0.0511	-0.0744	0.0181	-0.9973	0.0352	0.0254	0.0053	0.0585	0.0061
s.e.	0.0076	0.0078	0.0083	0.0078	0.0076	0.0006	0.0076	0.0074	0.0006	0.0014	0.0010

Cuadro 7.9: Estimación Modelo ARIMAX 7

De la tabla 7.9 se desprende que todos los parámetros y variables son significativos, mientras que en la figura 7.7 se puede ver que se cumplen los supuestos.

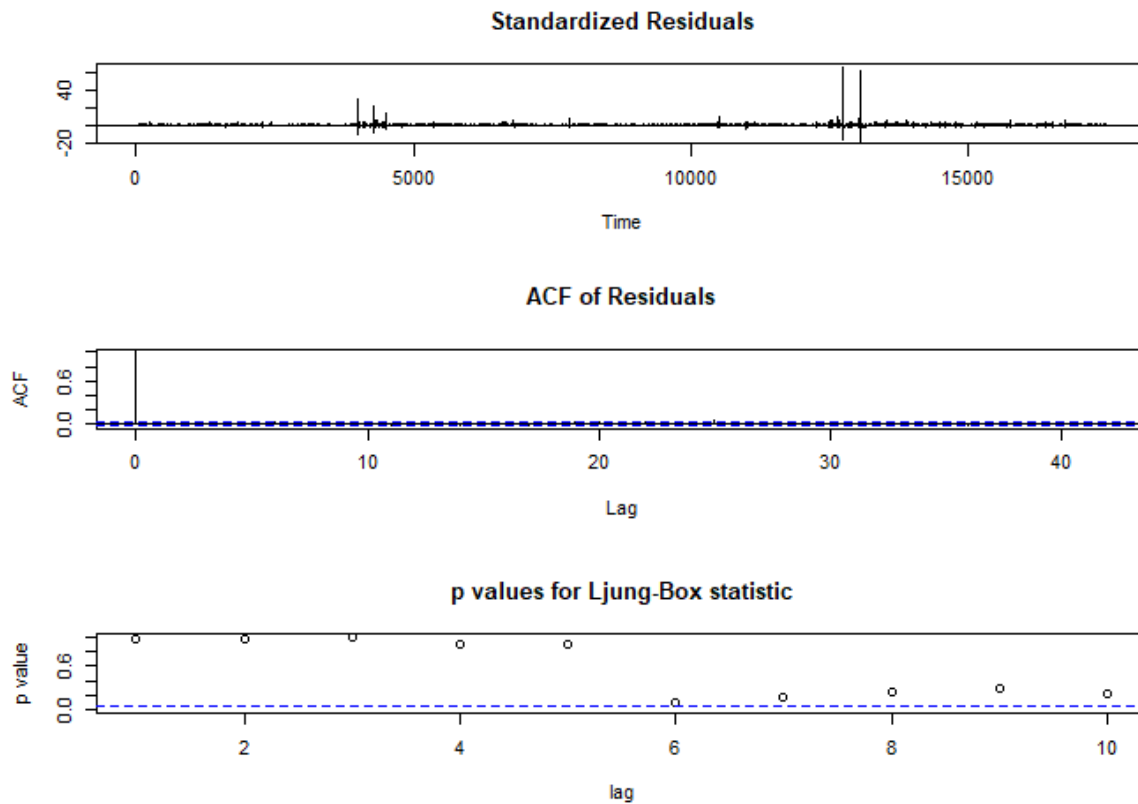


Figura 7.7: Diagnóstico residuos modelo ARIMAX 7

## Modelo 8

Finalmente se ajustará un modelo  $SARIMA(5, 1, 1)(0, 0, 2)_{24}$  con las 3 variables exógenas antes vistas, las cuáles están todas rezagadas en una hora.

En la tabla 7.10 se puede ver las estimaciones de los parámetros y es posible observar que todos los parámetros y las covariables son significativas en el modelo.

	ar1	ar2	ar3	ar4	ar5	ma1	sma1	sma2	$X_1$	$X_2$	$X_3$
Estimación	0.3202	0.3799	-0.0438	-0.0593	0.0326	-0.9961	0.0337	0.0216	0.0123	0.0154	-0.0088
s.e.	0.0076	0.0080	0.0084	0.0079	0.0076	0.0009	0.0079	0.0074	0.0011	0.0019	0.0017

Cuadro 7.10: Estimación Modelo ARIMAX 8

En la figura 7.8 es posible apreciar que se cumplen los supuestos de los residuos del modelo, por lo que se podría obtener predicciones confiables a partir de este.

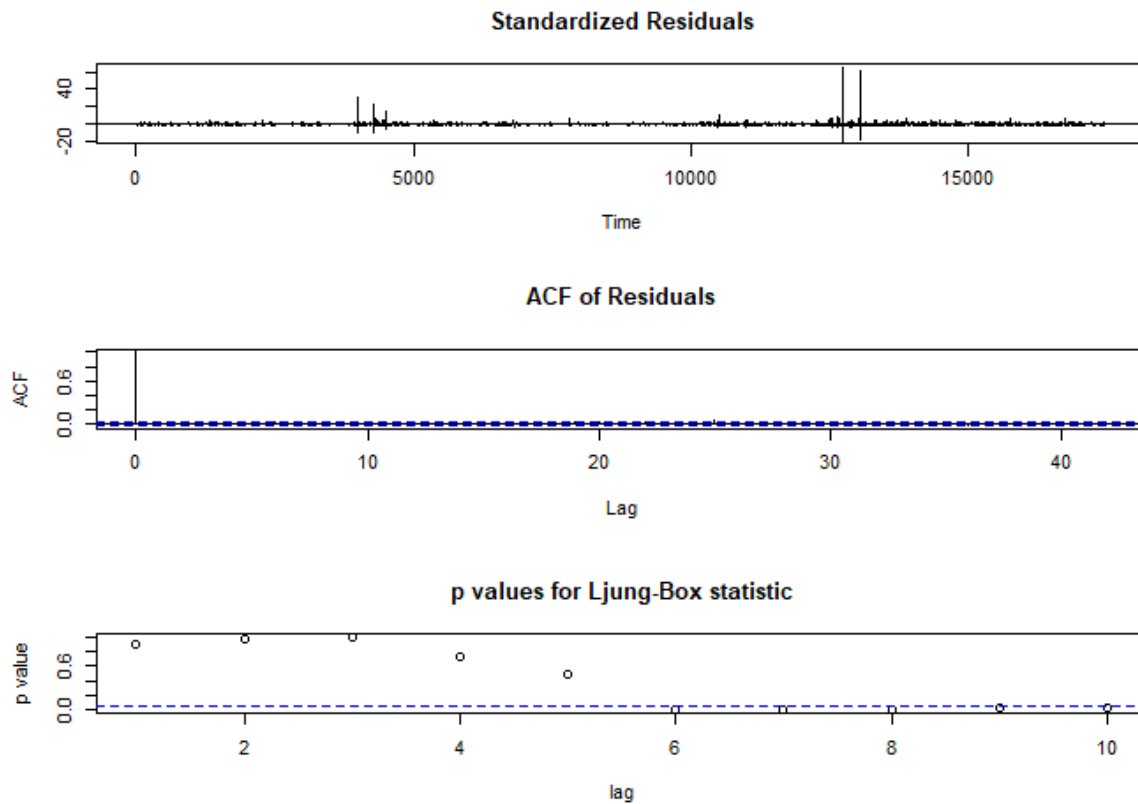


Figura 7.8: Diagnóstico residuos modelo ARIMAX 8

## Elección del mejor modelo

En vista de los resultados de los modelos propuestos anteriormente, es necesario compararlos, para así decidir cuál es el mejor.

	AIC	MAPE	Media	Homocedasticidad	Independencia
Modelo 1	28365,29	49,71 %	✓	✓	×
Modelo 2	29046,25	109,82 %	✓	✓	×
<b>Modelo 3</b>	26599,86	36,54 %	✓	✓	✓
<b>Modelo 4</b>	27171,81	31,08 %	✓	✓	✓
Modelo 5	28455,12	49,03 %	✓	✓	×
Modelo 6	29573,01	139,08 %	✓	✓	×
Modelo 7	26719,82	38,53 %	✓	✓	✓
Modelo 8	28187,1	58,06 %	✓	✓	✓

Cuadro 7.11: Resumen Modelos ARIMAX

A partir de los AIC de cada uno de los 8 modelos propuestos, el cálculo del promedio porcentual absoluto de los errores (MAPE) y el cumplimiento o no de los supuestos, se desprende que los modelos 3 y 4 son candidatos para realizar predicciones, es decir los modelo  $ARIMAX(5, 1, 1)$  con una y las 3 variables rezagadas.

### 7.3. Predicciones a través del modelo ARIMAX 3

Siendo el modelo  $ARIMAX(5, 1, 1)$  el que presenta parámetros significativos, cumple los supuestos y tiene menor AIC, se realizará el ajuste y predicciones mediante este.

La figura 7.9 muestra en rojo el ajuste del modelo sobre los valores reales, en donde se puede ver que los valores ajustados son similares a los reales y en gran parte detecta los momentos de congestión (sobre la línea azul), sin embargo en el caso de los valores *peak* del índice EDWIN, el modelo no alcanza a ajustarse por completo a estos.

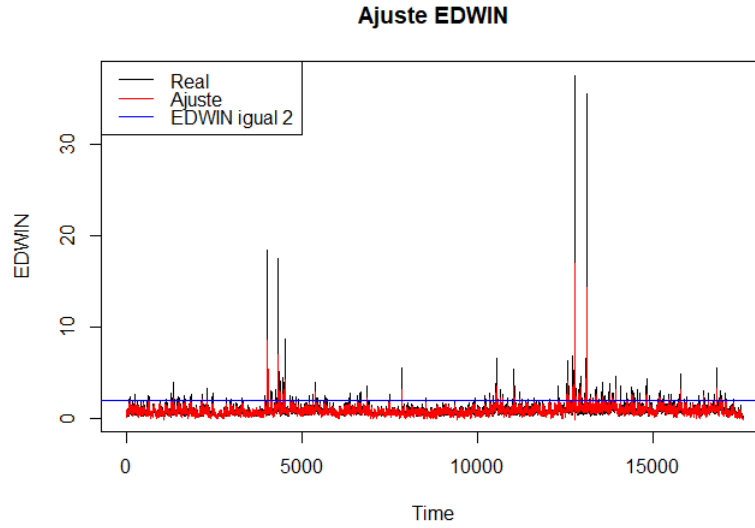


Figura 7.9: Ajuste modelo ARIMAX 3

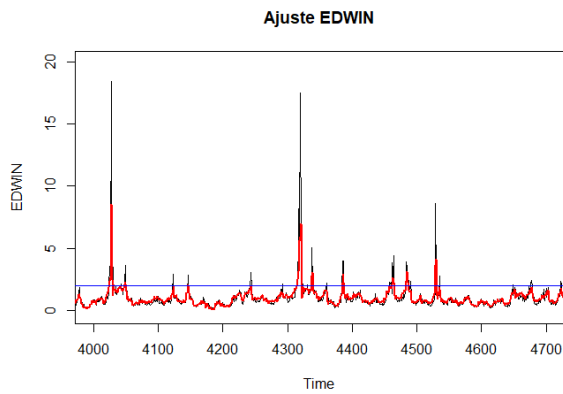


Figura 7.10: Ajuste modelo ARIMAX 3 inf

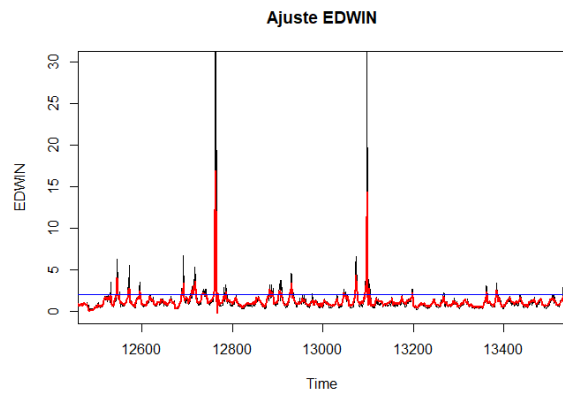


Figura 7.11: Ajuste modelo ARIMAX 3 sup

En las figuras 7.10 y 7.11 se puede ver de manera ampliada los tramos en que se muestra los valores máximos del EDWIN, en donde se verifica que los valores ajustados no son capaces de llegar a los valores más altos del índice, sin embargo se puede ver que en su mayoría se ajusta bien a los valores que indican congestión.

Por otra parte, en la tabla 7.12 se puede ver una forma de verificar si el modelo se ajusta realmente a los periodos de congestión, según los niveles vistos en la Sección 2.6.2. En esta tabla, se aprecia que el modelo sólo es capaz de ajustarse bien a momentos de congestión en el 42% de estas ocasiones, mientras que un 58% de los momentos, el ajuste muestra que el servicio no se encuentra congestionado.

AJUSTE			
REAL	BAJO CONTROL	OCUPADO	CONGESTION
BAJO CONTROL	<b>16123</b>	213	4
OCUPADO	438	<b>314</b>	25
CONGESTION	60	<b>187</b>	<b>179</b>

Cuadro 7.12: Comparación EDWIN Real y EDWIN estimado

DIA/HORA	Máx Congestión	Predicho	DIA/HORA	Mín Congestión	Predicho
15-06-2017 17:00	37,6	12,199	16-06-2016 22:00	2	2,158
29-06-2017 17:00	35,6	11,396	02-03-2017 23:00	2,014	1,224
15-06-2017 19:00	34,4	12,116	13-03-2017 19:00	2,015	1,54
16-06-2016 18:00	18,4	6,239	10-03-2016 15:00	2,029	1,683
28-06-2016 23:00	17,5	5,93	04-05-2016 19:00	2,031	1,366
28-06-2016 21:00	11,067	5,599	11-09-2017 18:00	2,04	1,161
07-07-2016 16:00	8,65	3,64	05-01-2016 17:00	2,043	1,273
12-06-2017 19:00	6,76	2,837	07-07-2016 17:00	2,046	4,165

Cuadro 7.13: Predichos Mínimos y Máximos de congestión

Como se mencionaba anteriormente, el modelo tiende a subestimar los valores del índice EDWIN, en la tabla ??, se puede apreciar que a pesar de subestimar, en los peaks del índice sigue indicando congestión, sin embargo en el caso de los valores mínimos que indican congestión (cerca de 2), el modelo no es capaz de detectar que efectivamente el sistema se encuentra en colapso.

## Predicciones a 72 horas del Modelo ARIMAX 3

Como se ve en la figura 7.12 hay predicciones que podrían no ser significativas, pues los intervalos de confianza contienen el cero.

En la tabla 7.14 se puede observar que efectivamente hay predicciones que contienen el cero, sin embargo las primeras 3 predicciones resultan ser significativas y luego, 17 horas después se vuelve a encontrar valores que resultan serlo.

Predicciones EDWIN

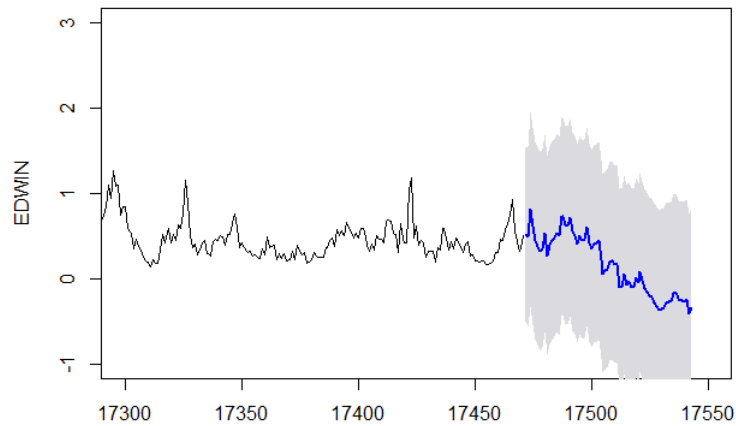


Figura 7.12: Predicción a 72 horas

DÍA/HORA	Predichos	IC 90 %	DÍA/HORA	Predichos	IC 90 %
28-12-2017 23:00	0,526	-0,488 1,539	29-12-2017 11:00	0,479	-0,680 1,637
29-12-2017 0:00	0,494	-0,558 1,545	29-12-2017 12:00	0,538	-0,620 1,697
29-12-2017 1:00	0,814	-0,325 1,953	29-12-2017 13:00	0,504	-0,655 1,662
29-12-2017 2:00	0,617	-0,533 1,768	29-12-2017 14:00	0,726	-0,433 1,884
29-12-2017 3:00	0,447	-0,708 1,603	29-12-2017 15:00	0,731	-0,427 1,889
29-12-2017 4:00	0,383	-0,774 1,541	29-12-2017 16:00	0,630	-0,528 1,789
29-12-2017 5:00	0,322	-0,836 1,480	29-12-2017 17:00	0,623	-0,535 1,782
29-12-2017 6:00	0,358	-0,800 1,516	29-12-2017 18:00	0,718	-0,441 1,876
29-12-2017 7:00	0,535	-0,624 1,693	29-12-2017 19:00	0,560	-0,598 1,719
29-12-2017 8:00	0,271	-0,888 1,429	29-12-2017 20:00	0,508	-0,650 1,667
29-12-2017 9:00	0,389	-0,769 1,548	29-12-2017 21:00	0,410	-0,749 1,568
29-12-2017 10:00	0,446	-0,712 1,605	29-12-2017 22:00	0,511	-0,647 1,670

Cuadro 7.14: Predicción primeras 24 horas

Sin considerar que hay predicciones no significativas, se puede apreciar que el servicio estaría bajo control en las próximas 24 horas, lo que concuerda con el comportamiento de las primeras 24 horas de los 2 años anteriores, donde no se observó valores elevados del índice.

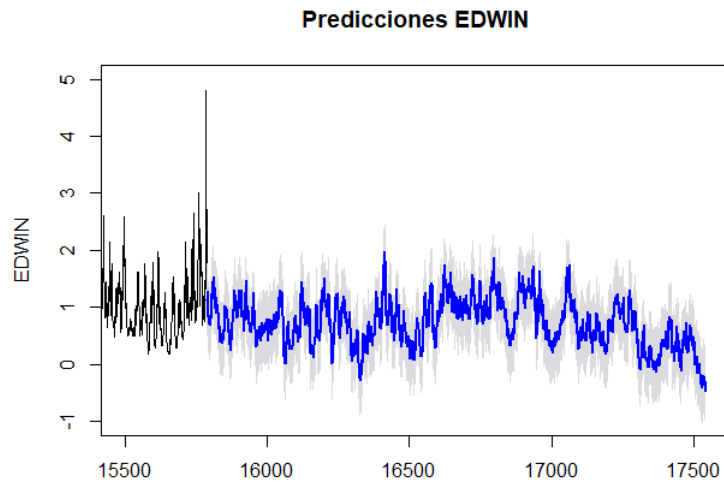


Figura 7.13: Predicción a 1754 horas

Por otra parte, al observar un intervalo de tiempo más grande de predicción, en la figura 7.13, se puede obtener predicciones más certeras y con intervalos de confianza más pequeños y con menos predicciones no significativas.

#### 7.4. Predicciones a través del modelo ARIMAX 4

Como se observó en la tabla 7.11 el modelo ARIMAX(5,1,1) con las 3 variables rezagadas, resultó tener el menos MAPE, lo que indica que el error de predicción es más pequeño que en los demás modelos.

En la figura 7.14 se muestra en rojo el ajuste del modelo sobre los valores reales, en donde, al igual que en el modelo 3, se observa que los valores ajustados son similares a los reales, excepto en los valores *peak* del índice.

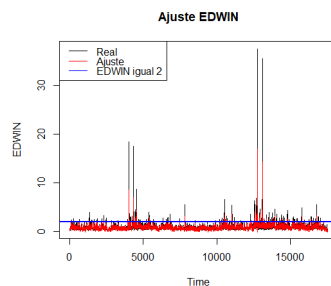


Figura 7.14: Ajuste modelo ARIMAX 4

En las figuras 7.15 y 7.16 se puede ver de manera ampliada los tramos en que se presentan los valores más altos del índice, donde se verifica que los valores ajustados no alcanzan a alcanzar a los reales más altos, sin embargo se puede ver que en la mayoría de los casos, se detecta la congestión.

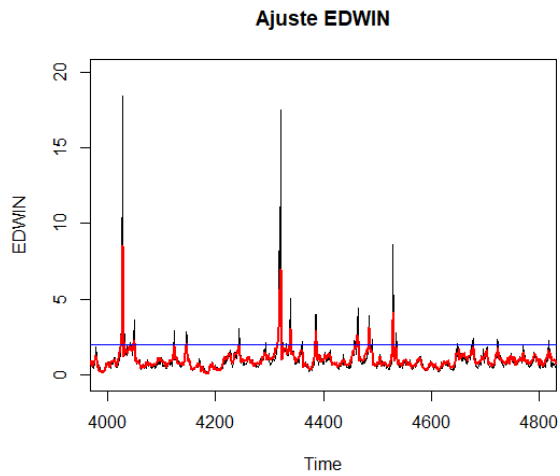


Figura 7.15: Ajuste modelo ARIMAX 4 inf

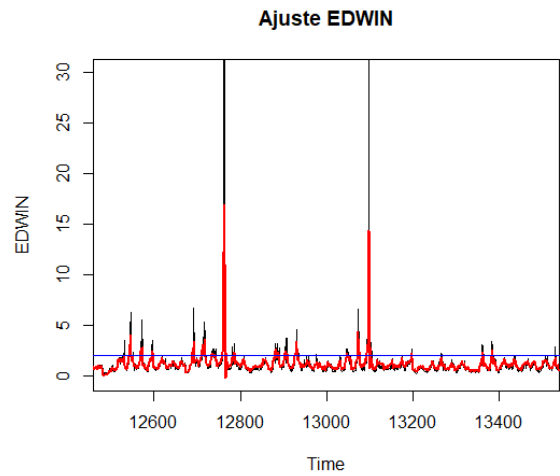


Figura 7.16: Ajuste modelo ARIMAX 4 sup

AJUSTE			
REAL	BAJO CONTROL	OCUPADO	CONGESTION
BAJO CONTROL	<b>16157</b>	181	2
OCUPADO	461	<b>292</b>	24
CONGESTION	65	<b>200</b>	<b>161</b>

Cuadro 7.15: Comparación EDWIN Real y EDWIN estimado

Al mirar la tabla 7.15, es posible darse cuenta que los casos en que el modelo detecta congestión cuando realmente la hay, se presenta en menos casos que el modelo 3, sin embargo al observar las veces en que el modelo indica que el servicio se encuentra ocupado, cuando realmente hay congestión, estas son más que en el modelo anterior, lo que logra equiparar el porcentaje en que el modelo predice servicio ocupado y congestión.

DIA/HORA	Máx Congestión	Predicho	DIA/HORA	Mín Congestión	Predicho
15-06-2017 17:00	37,6	11,808	16-06-2016 22:00	2	2,159
29-06-2017 17:00	35,6	11,159	02-03-2017 23:00	2,014	1,411
15-06-2017 19:00	34,4	11,637	13-03-2017 19:00	2,015	1,67
16-06-2016 18:00	18,4	6,043	10-03-2016 15:00	2,029	1,481
28-06-2016 23:00	17,5	5,791	04-05-2016 19:00	2,031	1,263
28-06-2016 21:00	11,067	5,263	11-09-2017 18:00	2,04	1,036
07-07-2016 16:00	8,65	3,366	16-02-2016 20:00	2,043	1,445
12-06-2017 19:00	6,76	2,858	07-07-2016 17:00	2,046	4,204

Cuadro 7.16: Predichos Mínimos y Máximos de congestión

Al analizar los valores mínimos y máximos de congestión, no se observan grandes diferencias con el modelo anterior, pues de la misma forma se aprecia que en los valores máximos sigue indicando congestión y para los valores cercanos a 2, se obtiene subestimaciones.

## Predicciones a 72 horas del Modelo ARIMAX 4

En la gráfica 7.17, se puede apreciar que a diferencia del modelo 3, se obtienen más predicciones en que el intervalo de confianza no contiene al cero, lo que a simple vista, podría indicar que efectivamente se obtiene mejores predicciones con este modelo.

Por otra parte, al observar las primeras 24 predicciones en la tabla 7.17, se puede ver que al igual que para el modelo 3, las primeras 3 predicciones resultaron ser significativas, junto con el intervalo que va desde las 13 a las 18 horas.

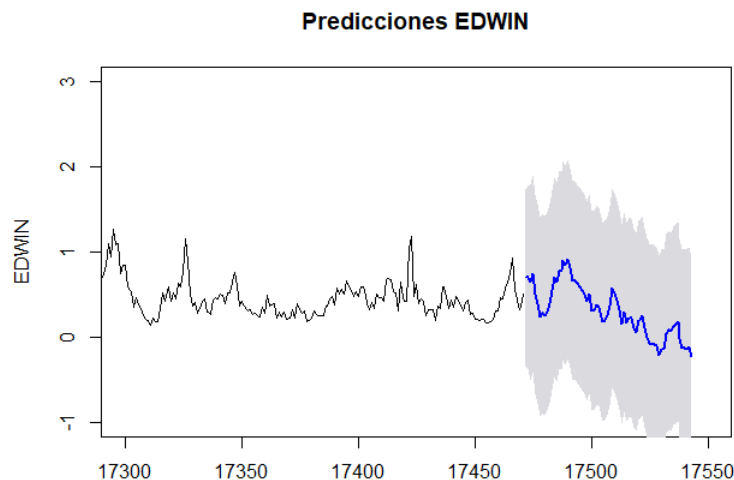


Figura 7.17: Predicciones a 72 horas

DÍA/HORA	Predichos	IC 95 %	DÍA/HORA	Predichos	IC 95 %
28-12-2017 23:00	0,698	-0,333 1,728	29-12-2017 11:00	0,706	-0,459 1,870
29-12-2017 0:00	0,706	-0,361 1,772	29-12-2017 12:00	0,645	-0,520 1,809
29-12-2017 1:00	0,642	-0,506 1,790	29-12-2017 13:00	0,781	-0,384 1,945
29-12-2017 2:00	0,747	-0,412 1,905	29-12-2017 14:00	0,782	-0,382 1,947
29-12-2017 3:00	0,504	-0,659 1,667	29-12-2017 15:00	0,905	-0,259 2,070
29-12-2017 4:00	0,394	-0,770 1,558	29-12-2017 16:00	0,842	-0,322 2,007
29-12-2017 5:00	0,239	-0,925 1,404	29-12-2017 17:00	0,919	-0,246 2,083
29-12-2017 6:00	0,283	-0,881 1,448	29-12-2017 18:00	0,813	-0,352 1,977
29-12-2017 7:00	0,257	-0,907 1,421	29-12-2017 19:00	0,661	-0,504 1,825
29-12-2017 8:00	0,302	-0,863 1,466	29-12-2017 20:00	0,668	-0,496 1,833
29-12-2017 9:00	0,393	-0,771 1,557	29-12-2017 21:00	0,627	-0,538 1,792
29-12-2017 10:00	0,548	-0,617 1,712	29-12-2017 22:00	0,590	-0,575 1,755

Cuadro 7.17: Predicción primeras 24 horas

En vista de los dos modelos que se utilizaron para obtener predicciones, se pudo observar que el modelo 3 presenta un mejor ajuste a la serie del EDWIN, sin embargo el modelo 4 resulta ser mejor, pues entrega mejores predicciones y cumple con el objetivo de poder *anticipar*, pues es el que incluye a las 3 variables rezagadas en una hora.

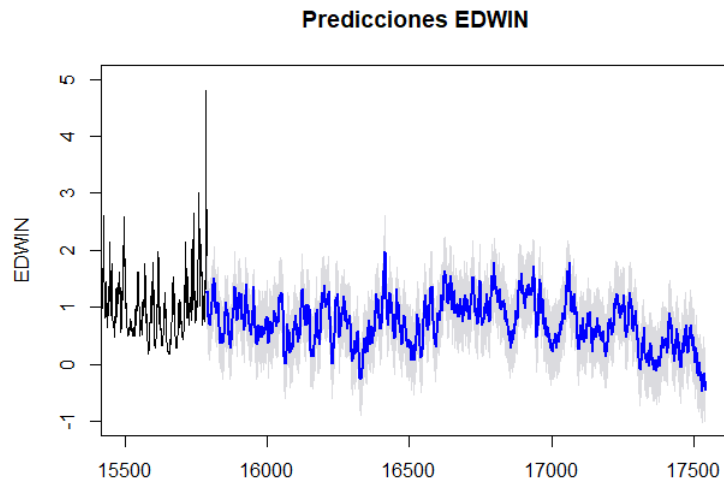


Figura 7.18: Predicción a 1754 horas

Sin embargo, al igual que para el modelo 3, se puede observar que al obtener predicciones a partir del 10% de los datos, (figura 7.18), se puede obtener predicciones más certeras y con intervalos de confianza más pequeños y con menos predicciones no significativas.

# Capítulo 8

## Conclusiones

En este trabajo fue posible abordar un tema de interés a nivel mundial para los profesionales del área de salud, más específicamente para los encargados de los servicios de urgencia, pues como se mencionó en la motivación, la congestión trae problemas, tanto a los pacientes como al personal que trabaja en estos.

Con el fin de anticipar la congestión en el tiempo, para encontrar alguna solución a este problema, se creó, a partir de la base de datos del HUAP, el índice EDWIN, que permite medir la congestión de forma numérica, para un instante determinado. Además, también se logró medir otras variables que permitieron explicar el comportamiento del índice en el tiempo.

Posteriormente se probaron diferentes tipos de modelos de series de tiempo. En primer lugar se usó el enfoque bayesiano, a partir del uso de los modelos lineales dinámicos (MLD), que en un principio se pensó que permitirían explicar el comportamiento del índice de congestión a partir de otras variables obtenidas de la urgencia del HUAP, sin embargo en este trabajo no fue posible cumplir con este objetivo.

Una de las posibles razones de que no se pueda obtener estimaciones y predicciones a través de los MLD, es que la cantidad de iteraciones realizadas no fueron suficientes, pues como se pudo observar, al eliminar 5000 en vez de las 4000 primeras estimaciones de los parámetros, se obtuvo una pequeña disminución de la autocorrelación en cada uno de estos.

Debido a la cantidad de observaciones en la serie de tiempo y el requerimiento de realizar muchas iteraciones para poder obtener una muestra que entregue resultados óptimos, se debe utilizar gran cantidad de recursos computacionales para realizar estimaciones a partir de los MLD, lo que también implica largos tiempos de implementación de cada uno de los modelos planteados.

En vista de que no se pudo obtener resultados mediante los modelos bayesianos, los modelos ARIMAX fueron de gran utilidad para realizar pronósticos de la congestión, a pesar de que las predicciones deben ser a un mediano o corto plazo, pues a medida que avanzan las predicciones en el tiempo, estas se vuelven negativas y se obtiene subestimaciones del índice.

Al realizar la comparación entre el modelo ARIMAX(5,1,1) con una variable rezagada (modelo 3) y el ARIMAX(5,1,1) con las tres variables rezagadas (modelo 4), se puede concluir que si bien el modelo 3 se ajusta mejor a los valores reales, el modelo 4 es el que entrega mejores predicciones, mostrando un menor porcentaje de error de predicción (MAPE) y es el que finalmente cumple con el objetivo de anticipar la congestión.

Si bien los tipos de modelos no son comparables en cuanto a resultados y estimaciones, es importante mencionar que los modelos lineales dinámicos, a diferencia de los ARIMAX, presentan mayor dificultad de implementación, pues debido al dinamismo de los parámetros se debe obtener la estimación de cada uno de estos, para cada observación en el tiempo.

Como se vio, las variables *Cantidad de pacientes en espera*, *Cantidad de pacientes hospitalizados* y *Cantidad de pacientes sin diagnóstico*, utilizadas para explicar y anticipar la congestión, resultaron ser significativas en los modelos clásicos, sería de utilidad estudiar la posibilidad de calcular otras variables a partir de la información disponible en el HUAP y del mismo modo estudiar si estas podrían influir en el índice, para intentar obtener mejores resultados.

Como trabajo futuro se propone analizar la eficacia del EDWIN para medir la congestión, contrastándolo con la percepción del personal acerca del estado del servicio de urgencia. Del mismo modo se podría estudiar otros índices internacionales que se puedan ajustar a la realidad del HUAP o analizar las variables disponibles en los registros, con el fin de intentar crear un nuevo índice que se ajuste al contexto de la ex Posta Central.

# Capítulo 9

## Anexos

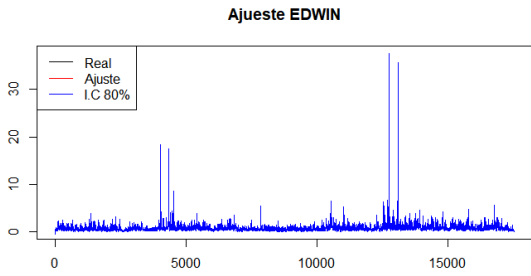


Figura 9.1: Ajuste MLD 1

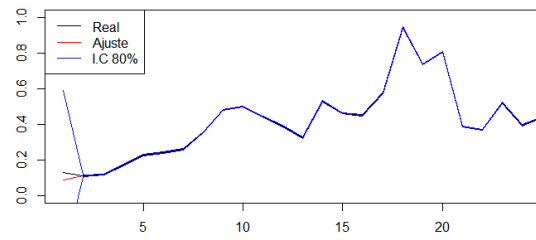


Figura 9.2: Zoom Ajuste MLD 1

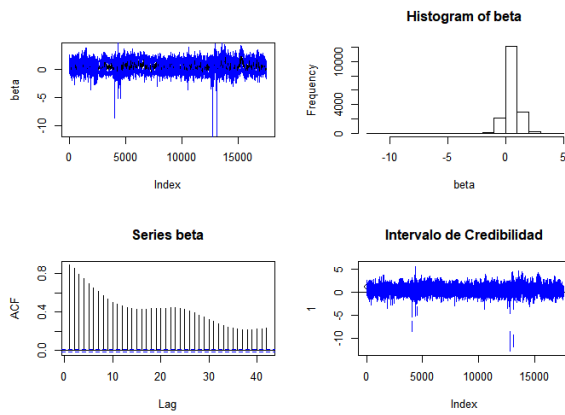


Figura 9.3:  $\beta_0$  MLD 1 4000 estimaciones eliminadas

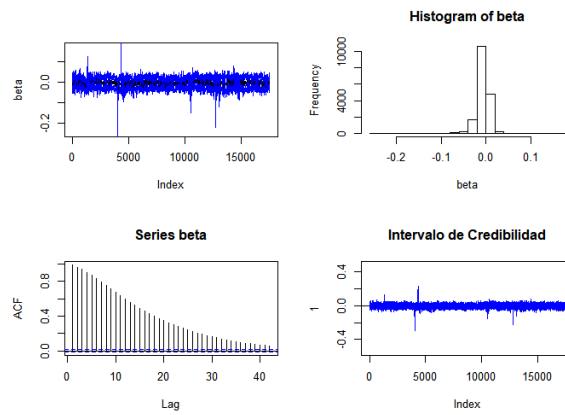


Figura 9.4:  $\beta_1$  MLD 1 4000 estimaciones eliminadas

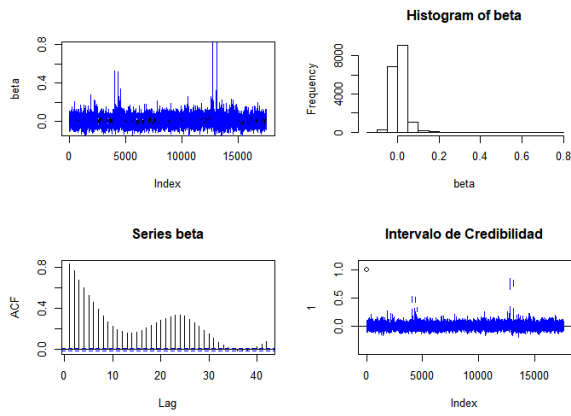


Figura 9.5:  $\beta_2$  MLD 1 4000 estimaciones eliminadas

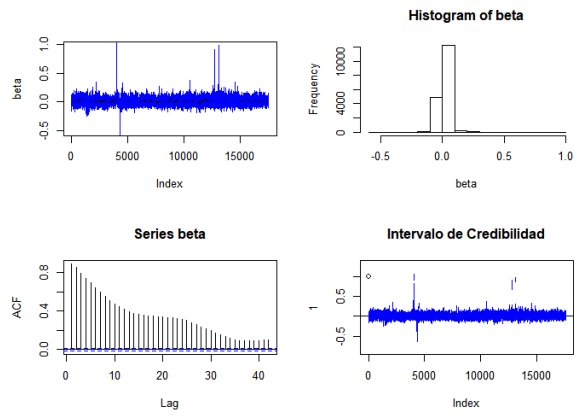


Figura 9.6:  $\beta_3$  MLD 1 4000 estimaciones eliminadas

	Estimado	Intervalo de credibilidad	
$W_{\beta_0}$	1,90E-01	1,53E-01	1,82E-01
$W_{\beta_1}$	8,79E-04	8,85E-04	1,02E-04
$W_{\beta_2}$	6,97E-04	5,88E-04	6,74E-04
$W_{\beta_3}$	7,62E-04	8,54E-04	1,00E-03
$V$	2,26E-04	2,25E-04	2,31E-04

Cuadro 9.1: Estimación parámetros MLD 1

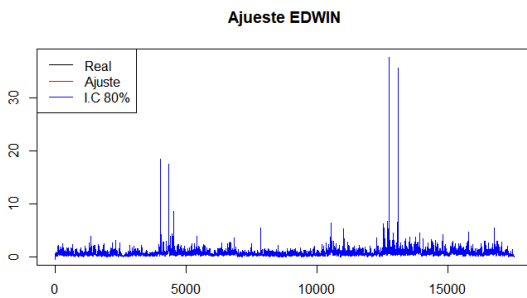


Figura 9.7: Ajuste MLD 2

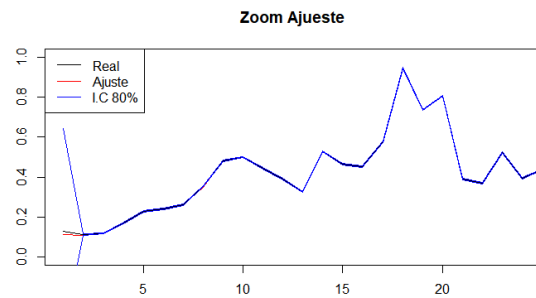


Figura 9.8: Zoom Ajuste MLD 2

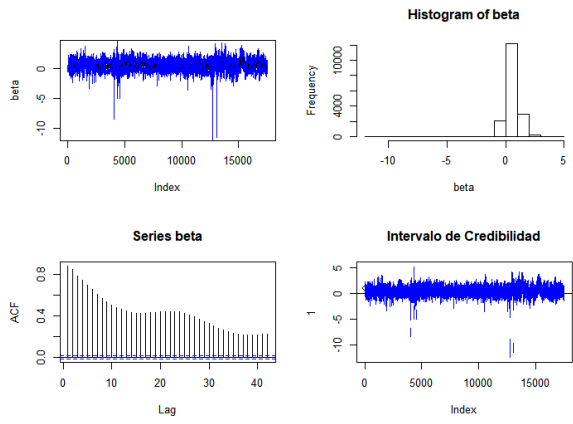


Figura 9.9:  $\beta_0$  MLD 2

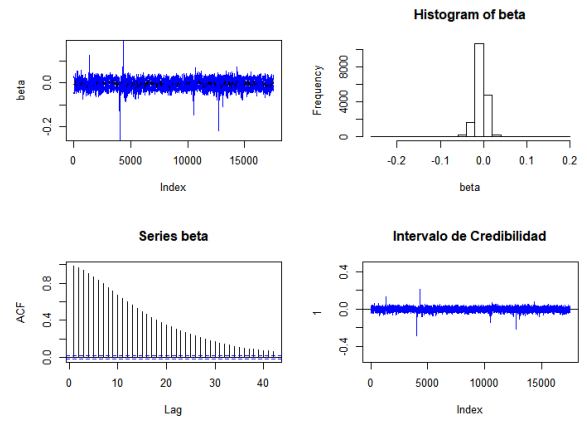


Figura 9.10:  $\beta_1$  MLD 2

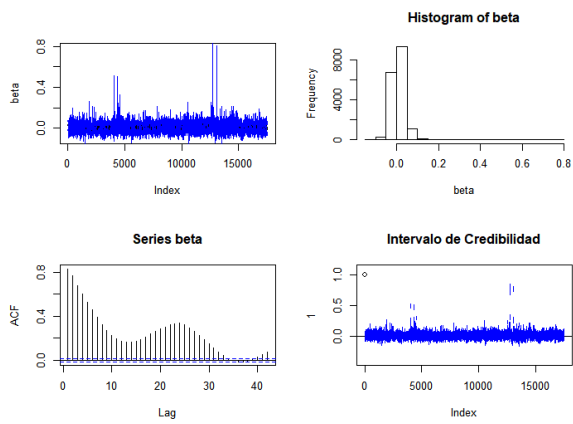


Figura 9.11:  $\beta_2$  MLD 2

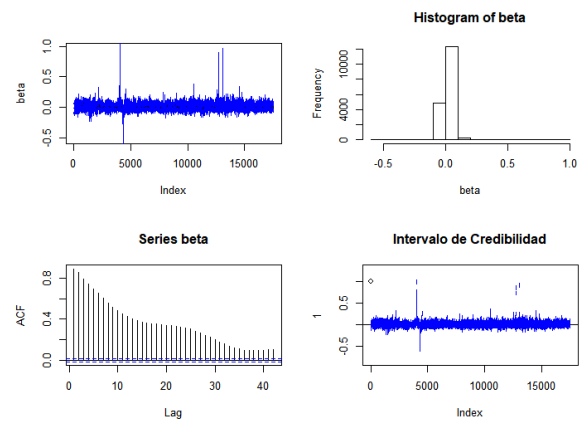


Figura 9.12:  $\beta_3$  MLD 2

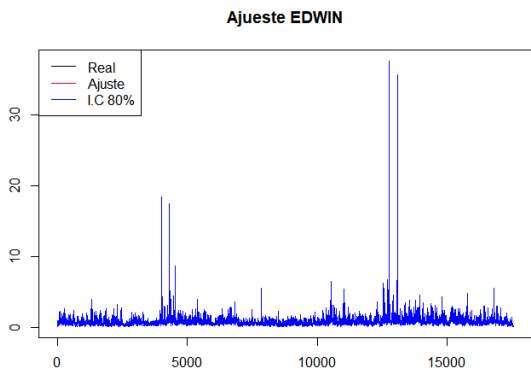


Figura 9.13: Ajuste MLD 3

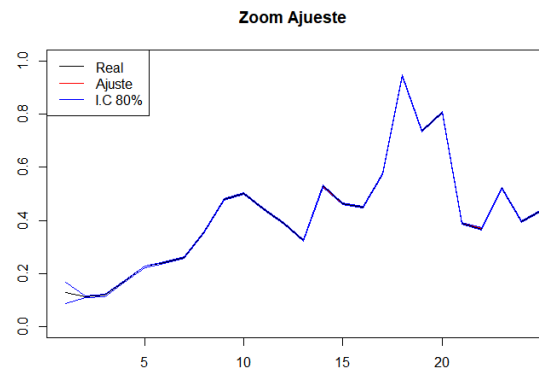


Figura 9.14: Zoom Ajuste MLD 3

	Estimado	Intervalo de credibilidad	
$W_{\beta_0}$	3,68E-04	2,70E-04	3,39E-04
$W_{\phi_1}$	2,35E-04	1,99E-04	2,42E-04
$W_{\beta_1}$	2,10E-05	2,00E-05	2,14E-05
$W_{\beta_2}$	1,12E-05	1,07E-05	1,13E-05
$W_{\beta_3}$	4,07E-05	3,86E-05	4,22E-05
$V$	2,30E-05	2,25E-05	2,31E-05

Cuadro 9.2: Estimación parámetros MLD 3

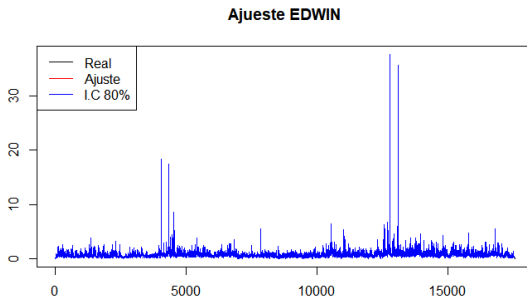


Figura 9.15: Ajuste MLD 4

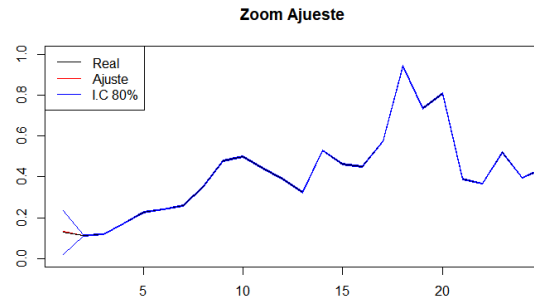


Figura 9.16: Zoom Ajuste MLD 4

	Estimado	Intervalo de credibilidad	
$W_{\beta_0}$	3,46E-04	2,77E-04	3,66E-04
$W_{\phi_1}$	2,03E-04	1,88E-04	2,38E-04
$W_{\phi_2}$	2,33E-04	1,96E-04	2,55E-04
$W_{\beta_1}$	2,05E-05	2,02E-05	2,20E-05
$W_{\beta_2}$	1,19E-05	1,10E-05	1,18E-05
$W_{\beta_3}$	3,81E-05	3,87E-05	4,32E-05
$V$	2,28E-06	2,24E-06	2,32E-06

Cuadro 9.3: Estimación parámetros MLD 4

DIA/HORA	Predichos	IC 80 %	DIA/HORA	Predichos	IC 80 %
01-01-2018 0:00	0,939	0,278 : 1,601	02-01-2018 12:00	0,401	-0,356 : 1,157
01-01-2018 1:00	0,8	0,113 : 1,486	02-01-2018 13:00	0,337	-0,42 : 1,093
01-01-2018 2:00	0,916	0,173 : 1,66	02-01-2018 14:00	0,23	-0,526 : 0,987
01-01-2018 3:00	0,654	-0,097 : 1,405	02-01-2018 15:00	0,183	-0,574 : 0,94
01-01-2018 4:00	0,507	-0,247 : 1,262	02-01-2018 16:00	-0,045	-0,801 : 0,712
01-01-2018 5:00	0,378	-0,377 : 1,134	02-01-2018 17:00	0,038	-0,719 : 0,794
01-01-2018 6:00	0,399	-0,357 : 1,155	02-01-2018 18:00	0	-0,757 : 0,756
01-01-2018 7:00	0,369	-0,387 : 1,126	02-01-2018 19:00	-0,055	-0,812 : 0,702
01-01-2018 8:00	0,325	-0,431 : 1,081	02-01-2018 20:00	-0,046	-0,802 : 0,711
01-01-2018 9:00	0,382	-0,375 : 1,138	02-01-2018 21:00	-0,108	-0,865 : 0,649
01-01-2018 10:00	0,501	-0,256 : 1,257	02-01-2018 22:00	-0,101	-0,858 : 0,656
01-01-2018 11:00	0,595	-0,161 : 1,352	02-01-2018 23:00	0,078	-0,679 : 0,834
01-01-2018 12:00	0,541	-0,215 : 1,297	03-01-2018 0:00	0,076	-0,681 : 0,832
01-01-2018 13:00	0,663	-0,093 : 1,419	03-01-2018 1:00	0,14	-0,617 : 0,897
01-01-2018 14:00	0,591	-0,165 : 1,348	03-01-2018 2:00	-0,006	-0,763 : 0,75
01-01-2018 15:00	0,74	-0,017 : 1,496	03-01-2018 3:00	-0,108	-0,865 : 0,649
01-01-2018 16:00	0,757	0 : 1,513	03-01-2018 4:00	-0,153	-0,91 : 0,604
01-01-2018 17:00	0,792	0,036 : 1,549	03-01-2018 5:00	-0,166	-0,922 : 0,591
01-01-2018 18:00	0,673	-0,083 : 1,43	03-01-2018 6:00	-0,166	-0,922 : 0,591
01-01-2018 19:00	0,68	-0,076 : 1,437	03-01-2018 7:00	-0,195	-0,952 : 0,562
01-01-2018 20:00	0,671	-0,086 : 1,427	03-01-2018 8:00	-0,369	-1,126 : 0,388
01-01-2018 21:00	0,582	-0,175 : 1,338	03-01-2018 9:00	-0,292	-1,049 : 0,465
01-01-2018 22:00	0,508	-0,249 : 1,264	03-01-2018 10:00	-0,285	-1,041 : 0,472
01-01-2018 23:00	0,522	-0,234 : 1,279	03-01-2018 11:00	-0,185	-0,942 : 0,572
02-01-2018 0:00	0,528	-0,229 : 1,284	03-01-2018 12:00	-0,214	-0,971 : 0,542
02-01-2018 1:00	0,493	-0,264 : 1,249	03-01-2018 13:00	-0,212	-0,969 : 0,545
02-01-2018 2:00	0,646	-0,11 : 1,403	03-01-2018 14:00	-0,167	-0,924 : 0,59
02-01-2018 3:00	0,445	-0,311 : 1,202	03-01-2018 15:00	-0,1	-0,857 : 0,657
02-01-2018 4:00	0,398	-0,358 : 1,155	03-01-2018 16:00	-0,072	-0,829 : 0,685
02-01-2018 5:00	0,468	-0,289 : 1,224	03-01-2018 17:00	-0,256	-1,013 : 0,501
02-01-2018 6:00	0,463	-0,294 : 1,219	03-01-2018 18:00	-0,279	-1,036 : 0,478
02-01-2018 7:00	0,246	-0,511 : 1,002	03-01-2018 19:00	-0,27	-1,027 : 0,487
02-01-2018 8:00	0,168	-0,588 : 0,925	03-01-2018 20:00	-0,238	-0,995 : 0,519
02-01-2018 9:00	0,149	-0,608 : 0,905	03-01-2018 21:00	-0,186	-0,943 : 0,571
02-01-2018 10:00	0,181	-0,576 : 0,938	03-01-2018 22:00	-0,344	-1,101 : 0,413
02-01-2018 11:00	0,266	-0,491 : 1,022	03-01-2018 23:00	-0,299	-1,056 : 0,458

Cuadro 9.4: Predicciones 72 horas ARIMAX 3

DIA/HORA	Predichos	IC 80 %	DIA/HORA	Predichos	IC 80 %
01-01-2018 0:00	0,806	0,133 : 1,478	02-01-2018 12:00	0,591	-0,169 : 1,352
01-01-2018 1:00	0,747	0,051 : 1,443	02-01-2018 13:00	0,538	-0,223 : 1,298
01-01-2018 2:00	0,793	0,044 : 1,543	02-01-2018 14:00	0,454	-0,306 : 1,215
01-01-2018 3:00	0,537	-0,219 : 1,294	02-01-2018 15:00	0,347	-0,414 : 1,107
01-01-2018 4:00	0,42	-0,339 : 1,179	02-01-2018 16:00	0,176	-0,585 : 0,936
01-01-2018 5:00	0,259	-0,501 : 1,018	02-01-2018 17:00	0,312	-0,449 : 1,072
01-01-2018 6:00	0,303	-0,457 : 1,063	02-01-2018 18:00	0,196	-0,564 : 0,957
01-01-2018 7:00	0,275	-0,485 : 1,035	02-01-2018 19:00	0,25	-0,511 : 1,01
01-01-2018 8:00	0,321	-0,439 : 1,081	02-01-2018 20:00	0,242	-0,519 : 1,002
01-01-2018 9:00	0,412	-0,348 : 1,172	02-01-2018 21:00	0,13	-0,63 : 0,891
01-01-2018 10:00	0,566	-0,194 : 1,326	02-01-2018 22:00	0,084	-0,677 : 0,844
01-01-2018 11:00	0,724	-0,036 : 1,484	02-01-2018 23:00	0,214	-0,547 : 0,974
01-01-2018 12:00	0,663	-0,097 : 1,423	03-01-2018 0:00	0,223	-0,537 : 0,984
01-01-2018 13:00	0,799	0,039 : 1,559	03-01-2018 1:00	0,268	-0,493 : 1,029
01-01-2018 14:00	0,801	0,041 : 1,561	03-01-2018 2:00	0,136	-0,625 : 0,897
01-01-2018 15:00	0,923	0,163 : 1,683	03-01-2018 3:00	0,024	-0,736 : 0,785
01-01-2018 16:00	0,86	0,1 : 1,62	03-01-2018 4:00	-0,044	-0,805 : 0,716
01-01-2018 17:00	0,937	0,177 : 1,697	03-01-2018 5:00	-0,067	-0,827 : 0,694
01-01-2018 18:00	0,831	0,071 : 1,591	03-01-2018 6:00	-0,067	-0,827 : 0,694
01-01-2018 19:00	0,679	-0,081 : 1,439	03-01-2018 7:00	-0,083	-0,844 : 0,677
01-01-2018 20:00	0,686	-0,074 : 1,447	03-01-2018 8:00	-0,184	-0,945 : 0,577
01-01-2018 21:00	0,646	-0,115 : 1,406	03-01-2018 9:00	-0,117	-0,878 : 0,644
01-01-2018 22:00	0,609	-0,152 : 1,369	03-01-2018 10:00	-0,115	-0,876 : 0,645
01-01-2018 23:00	0,564	-0,196 : 1,324	03-01-2018 11:00	0,063	-0,697 : 0,824
02-01-2018 0:00	0,527	-0,233 : 1,287	03-01-2018 12:00	0,104	-0,657 : 0,865
02-01-2018 1:00	0,442	-0,319 : 1,202	03-01-2018 13:00	0,086	-0,675 : 0,846
02-01-2018 2:00	0,527	-0,233 : 1,287	03-01-2018 14:00	0,154	-0,606 : 0,915
02-01-2018 3:00	0,334	-0,426 : 1,094	03-01-2018 15:00	0,181	-0,58 : 0,941
02-01-2018 4:00	0,332	-0,428 : 1,092	03-01-2018 16:00	0,207	-0,554 : 0,968
02-01-2018 5:00	0,397	-0,363 : 1,157	03-01-2018 17:00	0,008	-0,753 : 0,769
02-01-2018 6:00	0,377	-0,384 : 1,137	03-01-2018 18:00	-0,104	-0,865 : 0,657
02-01-2018 7:00	0,255	-0,506 : 1,015	03-01-2018 19:00	-0,111	-0,872 : 0,649
02-01-2018 8:00	0,196	-0,564 : 0,957	03-01-2018 20:00	-0,113	-0,874 : 0,648
02-01-2018 9:00	0,221	-0,54 : 0,981	03-01-2018 21:00	-0,091	-0,852 : 0,67
02-01-2018 10:00	0,267	-0,493 : 1,028	03-01-2018 22:00	-0,197	-0,958 : 0,564
02-01-2018 11:00	0,385	-0,376 : 1,145	03-01-2018 23:00	-0,176	-0,937 : 0,585

Cuadro 9.5: Predicciones 72 horas ARIMAX 4

# Bibliografía

- [1] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [2] Claudio Cárdenas, Carlos Sovier, Ulda Pérez, and CA Sergio González. Consultas de urgencia general y por causa respiratoria en la red de establecimientos del sistema nacional de servicios de salud (snss): un modelo predictivo en el servicio de salud de chiloé. *Revista chilena de enfermedades respiratorias*, 30(3):133–141, 2014.
- [3] Chris K Carter and Robert Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.
- [4] Departamento de Estadísticas e Información de Salud. <http://www.deis.cl/estadisticas-atencionesurgencia/>, 2017.
- [5] Sylvia Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of time series analysis*, 15(2):183–202, 1994.
- [6] Dani Gamerman. Markov chain monte carlo for dynamic generalised linear models. *Biometrika*, 85(1):215–227, 1998.
- [7] Nicolás Ariel Garrido. Mejora en los procesos de planificación, control y monitoreo del servicio de urgencia del hospital clínico de la universidad de chile. *Repositorio Universidad de Chile*, 2015.
- [8] P Jeffrey Harrison and Colin F Stevens. Bayesian forecasting. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 205–247, 1976.
- [9] Hospital Clínico San Borja Arriarán. <http://hcsba.cl/wp/como-funcionan-los-servicios-de-urgencia/>, 2016.
- [10] Hospital de Urgencia Asistencia Pública. [http://huap.redsalud.gob.cl/wrdprss\\_minsal/wp-content/uploads/2017/04/Cuenta-Publica-2016-V3.pdf](http://huap.redsalud.gob.cl/wrdprss_minsal/wp-content/uploads/2017/04/Cuenta-Publica-2016-V3.pdf), 2017.
- [11] Erik B Kulstad, K Michael Hart, and Simon Waghchoure. Occupancy rates and emergency department work index scores correlate with leaving without being seen. *Western Journal of Emergency Medicine*, 11(4):324, 2010.
- [12] Melissa L McCarthy, Scott L Zeger, Ru Ding, Dominik Aronsky, Nathan R Hoot, and Gabor D Kelen. The challenge of predicting demand for emergency department services. *Academic Emergency Medicine*, 15(4):337–346, 2008.
- [13] Pedro Alberto Morettin and Clélia Maria de Castro Toloí. *Modelos para previsão de séries temporais*, volume 1. Instituto de matematica pura e aplicada, 1981.

- [14] Farzad Rahmani, Farzin Rezazadeh, Alireza Ala, Maryam Soleimanpour, Robab Mehdi-zadeh Esfanjani, and Hassan Soleimanpour. Evaluation of overcrowding of emergency department in imam reza hospital in 2015 by implementing 2 scales: Nedocs and edwin. *Iranian Red Crescent Medical Journal*, 19(6), 2017.
- [15] RJ Salway, R Valenzuela, JM Shoenberger, WK Mallon, and A Viccellio. Congestión en el servicio de urgencia: Respuestas basadas en evidencias a preguntas frecuentes. *Revista Médica Clínica Las Condes*, 28(2):220–227, 2017.
- [16] Emilio Santelices and José Luis Santelices. Descripción y análisis del sistema de red de urgencia (rdu) en chile. recomendaciones desde una mirada sistémica. *Revista Médica Clínica Las Condes*, 28(2):186–198, 2017.
- [17] David J Spiegelhalter, Nicola G Best, Bradley P Carlin, and Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.
- [18] James H Stock and Mark W Watson. Introduction to econometrics: Global edition. *US: Pearson Education*, 2012.
- [19] Henri Theil. Linear aggregation of economic relations. 1954.
- [20] Steven J Weiss, Robert Derlet, Jeanine Arndahl, Amy A Ernst, John Richards, Madonna Fernández-Frankelton, Robert Schwab, Thomas O Stair, Peter Vicellio, David Levy, et al. Estimating the degree of emergency department overcrowding in academic medical centers: results of the national ed overcrowding study (nedocs). *Academic Emergency Medicine*, 11(1):38–50, 2004.
- [21] Mike West and Jeff Harrison. *Bayesian Forecasting and Dynamic Models*. Springer, 2 edition, 1997.