



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

ANÁLISIS DEL CRECIMIENTO ECONÓMICO A TRAVÉS DEL
INDICADOR GLOBAL DE LA ACTIVIDAD ECONÓMICA (IGAE)
BAJO LA METODOLOGÍA DE BOX-JENKINS Y REDES
NEURONALES ARTIFICIALES.

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

LICENCIADO EN ACTUARÍA

PRESENTA:

JULIO ADRIÁN BONILLA ALARCON

DIRECTOR DE TESIS:

DR. BULMARO JUÁREZ HERNÁNDEZ

PUEBLA, MAYO 2022.

Índice general

1. Conceptos Macroeconómicos y Crecimiento Económico.	13
1.1. Macroeconomía.	13
1.2. Conceptos macroeconómicos.	14
1.2.1. Agregados Económicos.	14
1.2.2. Producto Interno Bruto (PIB).	14
1.2.3. Producto Nacional Bruto (PNB).	15
1.3. Crecimiento económico	15
1.4. Indicadores del crecimiento económico.	15
1.4.1. Definiciones básicas para comprender el índice de crecimiento económico.	16
1.5. Indicador Global de la Actividad Económica (IGAE).	17
1.5.1. Concepto de Indicador Global de la Actividad Económica (IGAE).	17
1.5.2. Producto Interno Bruto (PIB) contra el Indicador Global de la Actividad Económica (IGAE)	17
1.5.3. Cálculo del IGAE	18
1.5.4. Números Índices Utilizados	18
1.5.5. Datos IGAE	20
2. Modelos de series de tiempo	21
2.1. Procesos estocásticos.	21
2.2. Serie de tiempo	22
2.3. Autocovarianza y Autocorrelación	24
2.3.1. Autocovarianza, autocorrelación y autocorrelación parcial	25
2.3.2. Series de tiempo estrictamente y débilmente estacionarias	28
2.4. Series de tiempo no estacionarias	29
2.5. Ejemplos de series de tiempo estacionarias	29
2.5.1. Procesos de Ruido Blanco	29
2.5.2. Ruido i.i.d (independiente e idénticamente distribuido).	31
2.5.3. Caminata Aleatoria.	31
2.6. Función de autocorrelación muestral (FAM)	32
2.7. Función de Autocorrelación Parcial Muestral (FACMP)	32
2.8. Modelo Autorregresivo de Promedios Móviles (ARMA)	33
2.8.1. Introducción	33

2.8.2.	Procesos de media móvil de primer orden o Proceso MA(1)	34
2.8.3.	Procesos de media móvil de segundo orden o Proceso MA(2)	36
2.8.4.	Procesos de media móvil de orden q o Proceso MA(q)	38
2.8.5.	Estacionariedad del Proceso MA	38
2.8.6.	Ejemplo MA(q)	39
2.8.7.	Proceso Autorregresivo de primer orden o Proceso AR(1)	41
2.8.8.	Proceso Autorregresivo de Segundo Orden O Proceso AR(2)	43
2.8.9.	Proceso Autorregresivo de Orden p o Proceso AR(p)	45
2.9.	Proceso Autorregresivo de Medias Móviles (ARMA)	46
2.9.1.	ARMA(1,1)	46
2.10.	Modelo Autorregresivo de Medias Móviles Integrado <i>ARIMA</i>	49
2.10.1.	Modelo autorregresivo integrado de medias móviles de orden p, d, q <i>ARIMA</i> (p, d, q)	50
2.10.2.	Test de Dickey-Fuller o de Raíz unitaria.	50
2.10.3.	Modelo <i>ARIMA</i> (0, 1, 1) o <i>IMA</i> (1, 1)	51
2.10.4.	Varianza y autocovarianza del modelo <i>ARIMA</i>	51
2.11.	Metodología Box-Jenkins	52
2.11.1.	Estacionariedad	53
2.11.2.	Identificación.	54
2.11.3.	Estimación.	55
2.11.4.	Evaluación del modelo	55
2.11.5.	Pronóstico	55
2.12.	Ejemplo	56
3.	Redes Neuronales Artificiales	65
3.1.	Conceptos Básicos de Redes Neuronales	66
3.1.1.	Ventajas de la red neuronal	68
3.1.2.	La neurona (Unidad)	68
3.2.	Función de Activación	69
3.2.1.	Función Lineal	69
3.2.2.	Función Escalón	70
3.2.3.	Función Rectificadora Lineal Unificada (ReLU)	71
3.2.4.	Función sigmoide	72
3.3.	Función de costo (loss function)	75
3.3.1.	Error cuadrático medio	75
3.3.2.	Error medio absoluto	76
3.3.3.	Las funciones Log loss, logistic loss o cross-entropy loss	76
3.4.	Múltiples capas	77
3.5.	Entrenamiento	78
3.5.1.	Retropropagación (Backpropagation)	78
3.6.	Preprocesado	80
3.6.1.	Imputación de valores ausentes	80
3.6.2.	Exclusión de variables con varianza próxima a cero	81

3.6.3.	Estandarización y escalado de variables numéricas	81
3.6.4.	Binarización de las variables cualitativas	82
3.7.	Número y tamaño de capas	82
3.8.	Tasa de aprendizaje	82
3.9.	Tipos de optimizadores de la función de costo	83
3.9.1.	Gradiente descendente	83
3.9.2.	Descenso acelerado de Nesterov (NAG)	84
3.9.3.	Descenso de Gradiente Adaptable (ADAGRAD)	84
3.9.4.	Momentun Adaptable (ADAM)	85
3.9.5.	Importancia de la reducción del sesgo	86
3.10.	Ejemplo	87
4.	Caso de estudio	97
4.1.	Metodología del cálculo	100
4.2.	Análisis de los datos del IGAE usando la metodología de Box-Jenkins (Modelo <i>SARIMA</i>).	101
4.2.1.	Análisis del modelo sin intervención	102
4.3.	Análisis de la serie de datos del IGAE con Redes Neuronales Artificiales.	114
4.3.1.	Análisis de las muestras para la creación de la Red.	114
4.3.2.	Algunas gráficas descriptivas de los datos.	116
4.3.3.	Preparación de los datos	117
4.3.4.	Creación de la Red Neuronal Artificial.	118
4.3.5.	Pronósticos del IGAE	121
5.	Análisis y conclusiones	125
5.0.1.	Análisis y Resultados	127
5.0.2.	Conclusiones.	133
	Referencias	133

Índice de figuras

2.1. Ejemplo de gráfica de autocorrelación.	27
2.2. Ruido Blanco.	30
2.3. Caminata Aleatoria.	31
2.4. Ruido Blanco.	36
2.5. Llegada de barcos a puertos mexicanos.	39
2.6. Descomposición de la serie.	40
2.7. Medias móviles de la serie.	40
2.8. Proceso Autorregresivo de orden 1.	42
2.9. Trayectoria para un proceso ARMA(1,1).	48
2.10. Metodología Box-Jenkins.	53
2.11. Gráfico Serie de Tiempo.	57
2.12. Gráfico de la Serie de Tiempo con una diferencia.	58
2.13. Gráfico Serie de Tiempo con dos diferencias.	59
2.14. FACP de la serie con dos diferencias.	60
2.15. Diagnostico del modelo propuesto.	61
2.16. Gráficas de Residuales del Modelo.	62
2.17. Pronóstico	63
3.1. Representación de una red neuronal.	66
3.2. Representación de una red neuronal que es equivalente a un modelo lineal.	67
3.3. Comportamiento de una neurona artificial.	69
3.4. Representación función lineal.	70
3.5. Representación función escalón.	71
3.6. Representación función ReLU.	72
3.7. Representación función Sigmoide.	73
3.8. Representación función Tangente Hiperbólica	73
3.9. Representación de una red neuronal feed-forward (multi-layer perceptron).	77
3.10. Representación del modelo NAG.	84
3.11. Gráfico de Datos	88
3.12. Gráfico de Medias Mensuales de los años 2017 y 2018.	90
3.13. Gráfico de Datos del Comportamiento de la Red Neuronal contra los Datos Originales.	93

3.14. Curva de pérdida.	93
3.15. Pronóstico de ventas en los próximos 7 días.	96
4.1. Actividades económicas dominantes en México [6].	99
4.2. Serie de tiempo obtenida para el IGAE.	101
4.3. Representación de las series después de su transformación.	102
4.4. Serie X_t^1 con una diferencia.	103
4.5. FAM y FAMP de la serie X_t^1	103
4.6. Diagrama de cajas y bigotes.	104
4.7. Gráfico FAM, FAMP y comportamiento del $IGAE^{\frac{1}{4}}$ nivel estacional.	105
4.8. Gráfico FAM, FAMP y comportamiento del $IGAE^{\frac{1}{4}}$ con una diferencia no estacional.	106
4.9. $IGAE^{\frac{1}{4}}$ con una diferencia en el nivel no estacional y en el estacional.	107
4.10. Pronóstico de Enero a Diciembre 2019.	110
4.11. Comparación de los datos originales con la propuesta obtenida por el modelo.	111
4.12. Pronostico IGAE para Enero a Diciembre del 2020.	112
4.13. Serie original y propuesta del 2020.	112
4.14. Gráfico de residuales del modelo SARIMA(4,0,1)(2,1,2) ₁₂	113
4.15. Comportamiento de los datos de Enero de 1993 a Diciembre de 2019.	114
4.16. Gráfica de la medias a través del tiempo.	116
4.17. Gráfica de la medias a través del tiempo.	116
4.18. Estructura de la Red Neuronal con $n = 12$ y $m = 12$	119
4.19. Comparación de información.	120
4.20. Resultados de otros modelos de entrenamiento.	120
4.21. Comparación de los datos reales y pronosticados para el IGAE mensual del 2019.	122
4.22. Pronóstico obtenido a través de redes Neuronales IGEA 2020.	123
5.1. Comparación de los pronósticos de las metodologías Box-Jenkins y Redes Neuronales para 2019.	128
5.2. Comparación de resultados obtenidos con resultados INEGI 2019.	130
5.3. Diagramas de densidad generadas por el método Kernel.	130
5.4. Gráficos de pronósticos obtenidos a través de la metodología Box-Jenkins y Redes Neuronales para el año 2020.	132
5.5. Comparación de ambas metodologías con datos preliminares 2020.	132

Índice de cuadros

2.1. Clasificación de procesos estocásticos.	22
2.2. Series de tiempo en diferentes campos científicos.	23
2.3. Resumen FAC y FACP para los diferentes procesos.	49
2.4. Observaciones del petróleo maya de Enero de 2013 a Diciembre de 2019.	56
2.5. Serie de Tiempo Transformada.	56
2.6. Prueba Dickey-Fuller	57
2.7. Prueba Dickey-Fuller para observaciones con una diferencia.	58
2.8. Prueba Dickey-Fuller para observaciones con dos diferencias.	59
2.9. Modelos <i>ARIMA</i> propuestos para obtener un pronóstico.	60
2.10. Parámetros del modelo.	61
2.11. Prueba L-Jung Box.	62
2.12. Pronostico próximos 9 meses.	63
3.1. Tipos de funciones de activación [12].	74
3.2. Ventas diarias [18].	87
3.3. datos anuales.	88
3.4. Estadísticas de ventas.	89
3.5. Medias mensuales de ventas.	89
3.6. Matriz de entradas y salidas.	91
3.7. Resultados del Entrenamiento.	92
3.8. Datos de la mitad de los días del mes de noviembre del 2018.	94
3.9. Matriz de entradas y salidas.	94
3.10. Pronóstico primer semana de Diciembre.	95
4.1. Clasificación de actividades económicas.	100
4.2. Prueba Dickey-Fuller para la serie con una diferencia.	104
4.3. Prueba Dickey-Fuller para la serie transformada por la raíz cuarta.	105
4.4. Prueba Dickey-Fuller para la serie con una diferencia.	106
4.5. Prueba Dickey-Fuller para la serie con una diferencia.	107
4.6. Criterio de Akaike	108
4.7. Valores obtenido del modelo $ARIMA(4, 0, 1)(2, 1, 2)_{12}$	109
4.8. Pronóstico de Enero a Diciembre del 2019	110
4.9. Pronostico Enero a Diciembre del 2020	111

4.10. Primeras observaciones.	115
4.11. Cantidad de observaciones por año.	115
4.12. Resumen estadístico de los datos observados del IGAE.	115
4.13. Matriz de entradas y salidas.	118
4.14. Primeras observaciones.	119
4.15. Pronóstico Enero a Diciembre del 2019.	121
4.16. Matriz de entradas y salidas.	122
4.17. Datos obtenidos por la red neuronal para el pronóstico del año 2020. . .	123
5.1. Pronósticos de Enero a Diciembre del 2019, usando la metodología de Box-Jenkins y Redes Neuronales.	128
5.2. Información arrojada por el INEGI y resultados obtenidos de la metodología Box-Jenkins y Redes Neuronales Artificiales para el año 2019. .	129
5.3. Comparación de resultados obtenidos por el pronóstico de Enero a Diciembre del 2020.	131

Introducción

La búsqueda de nuevas técnicas y métodos que permitan mejorar la eficiencia en la ejecución de diferentes tipos de tareas, siempre ha sido una característica del ser humano, ya que uno de sus principales objetivos ha sido la optimización del tiempo, la eficiencia de las cosas y automatizar las operaciones mecánicas. Un caso de estudio muy importante para la humanidad siempre ha sido el manejo del sector económico en especial el crecimiento económico de un país. Puesto que el crecimiento económico tiene un fin sin fin de formas de medirse, ya que existen diversos indicadores los cuales miden tal crecimiento, la mejora de estos indicadores lleva, teóricamente, a un aumento en los estándares de vida de la población [6].

En la mayoría de los países existen desde hace bastante tiempo algunos indicadores generales sobre el comportamiento de la coyuntura mexicana, con información sobre la evaluación regularmente mensual de la producción industrial; los resultados del comercio exterior, el desempeño de las variables fiscales, financieras y monetarias, así como índices de precios al productor o al consumidor. En los años noventa se publicó por primera vez una variable que media la actividad económica denominado Indicador Global de la Actividad Económica (IGAE) [10].

El Indicador Global de la Actividad Económica permite conocer y dar seguimiento a la evolución mensual del sector real de la economía [11], este indicador ofrece información oportuna, completa y coherente de la evolución económica de México, dicha información nos ayuda para la toma de diferentes decisiones sobre el país. Se elabora utilizando el mismo marco conceptual y metodológico empleado en el cálculo de las Cuentas de Bienes y Servicios del Sistema de Cuentas Nacionales de México (SCNM), conceptos, fuentes de datos y compilación (CNT) del Fondo Monetario Internacional.

A través del tiempo el uso de análisis y pronósticos se han convertido de vital importancia para cualquier situación importante ya que, dadas algunas condiciones históricas y actuales, les permite conocer algunas condiciones en el futuro.

Pronosticar es muy importante, ya que las predicciones de hechos futuros se pueden incorporar al proceso de toma de decisiones. Al predecir hechos que ocurrirán en el futuro, quien elabora los pronósticos debe confiar en la información de los hechos que han ocurrido en el pasado y que su de su interes. Es decir, con objeto de preparar

un pronóstico, quien lo elabora tiene que analizar la información anterior y basar la predicción en los resultados del análisis [1], [2], [4] [5], [6], [7] [8].

Una serie de tiempo la podemos definir como un conjunto de datos estadísticos que se recopilan, se observan o registran en intervalos de tiempo regulares, es una forma estructurada de representar datos.

El estudio de series de tiempo es un estudio importante para desarrollar modelos estadísticos para la explicación del comportamiento de una variable aleatoria que varía con el tiempo, con la distancia, o según el índice; y que permiten estimar datos futuros de dicha variable aleatoria, cabe recalcar que en un pronóstico siempre existe un grado de incertidumbre, en pocas palabras un pronóstico siempre contiene un nivel de error, por lo que es de vital importancia medir dichos errores para ver si la metodología de dicho pronóstico es adecuada [2], [3], [4], [5], [8].

Nuevas técnicas y métodos que han permitido mejorar la ejecución de diferentes tipos de tareas han llevado al desarrollo, relativamente reciente, del método llamado Redes Neuronales Artificiales (RNA).

Las RNA representan un nuevo paradigma en el manejo y análisis de la información inspirada en el funcionamiento de las miles de millones de neuronas biológicas que el ser humano posee en su sistema nervioso.

Las RNA son sistemas que aproximan a nivel básico todo este complejo hardware biológico, resultando en un esquema artificial semi-paramétrico que permite el análisis de variables financieras y económicas en las que el mecanismo de la relación subyacente es especialmente no lineal. En este sentido, la unidad básica de una RNA es la neurona, denominada también procesador elemental, ésta representa un dispositivo simple de cálculo que recibe señales exteriores de otras neuronas y proporciona una respuesta a dichos impulsos [12].

Capítulo 1

Conceptos Macroeconómicos y Crecimiento Económico.

1.1. Macroeconomía.

Una primera definición de macroeconomía se puede tomar como el estudio de los agregados económicos. Con base en esto, se puede entender cómo funciona la economía, y obtener recomendaciones de política económica.

El estudio de la macroeconomía siempre ha estado ligado a sus implicancias de política. Éstas pueden ir desde el extremo donde se plantea que no hay nada que hacer, pues lo que observamos en la realidad no son más que respuestas óptimas de las empresas y hogares a cambios en la economía, hasta otro extremo donde se ven alarmantes señales de desequilibrios que sería necesario corregir con medidas de política económica. Para adoptar cualquier posición, primero debemos entender la realidad.

Definición 1.1. (Macroeconomía). La palabra macro proviene del griego "makros" que significa grande. La macroeconomía como parte de la economía estudia los grandes agregados económicos: la producción, el empleo, la inversión, los precios, las importaciones, exportaciones, entre otros. Considera la suma de las actividades económicas que desempeñan todas las unidades económicas individuales, es decir, se encarga del análisis del comportamiento económico colectivo.

La macroeconomía ayuda a explicar los acontecimientos económicos así como también elaborar medidas cuyo objetivo es mejorar los resultados económicos [6].

1.2. Conceptos macroeconómicos.

1.2.1. Agregados Económicos.

Dentro de esta estructura, los agregados macroeconómicos son los que se encargan del estudio del conjunto de la actividad económica y de magnitudes globales. Su propósito de determinar las condiciones generales de crecimiento y equilibrio de la economía. Además, con los agregados macroeconómicos se pueden detectar los “Cuellos de botella” de las economías nacionales, destacando las debilidades sectoriales y regionales.

Los agregados macroeconómicos nos muestran las tendencias de las economías nacionales, su crecimiento real, su constitución sectorial y otra serie de indicaciones de gran importancia, como la evolución de la carga tributaria y las realizaciones de los sectores público y privado.

También se puede detectar el ritmo del ingreso de la colectividad, los niveles de consumo agregado, la evolución del ahorro total, la acumulación de capital, la distribución del ingreso entre los diferentes factores productivos, de manera más formal podemos decir.

Definición 1.2. (Agregado Económico). Estudia todas aquellas tendencias en conjunto de las decisiones tomadas por los individuos. Se encargan del estudio en conjunto de las actividades económicas, como también de las magnitudes globales, para poder determinar el crecimiento y el equilibrio de la economía a la que analiza.

Nos muestra cuáles son las tendencias de la economía nacional, y también cuál es su crecimiento real, tanto del sector público como del sector privado.

1.2.2. Producto Interno Bruto (PIB).

El PIB, posiblemente es el indicador económico más conocido por la población, ya que es la medida más comprensiva de la actividad económica de un país y por lo tanto, el mejor indicador de su comportamiento general y del tamaño en sí de la economía.

Definición 1.3. (PIB). Es un indicador económico que refleja el valor monetario de todos los bienes y servicios finales producidos por un país o región en un determinado periodo de tiempo.

Las palabras clave son final y producidos. Al referirse a bienes finales, significa los que se producen para consumo final y no intermedios para el uso de otro bien final. La idea es captar el valor agregado en cada etapa de la producción y, así, evitar una doble contabilidad.

1.2.3. Producto Nacional Bruto (PNB).

Otro agregado económico de importancia es el Producto Nacional Bruto (PNB), este agregado permite medir cuál es el ingreso total producido por una nación y se puede definir como sigue.

Definición 1.4. (PNB). El Producto Nacional Bruto (PNB) es el indicador que mide el ingreso económico de los residentes nacionales que se encuentran en el país y en el extranjero.

1.3. Crecimiento económico

El crecimiento económico es de vital importancia para lograr el desarrollo económico y social, lo que constituye una de las metas principales de todo país o nación. Este implica el crecimiento significativo de los ingresos y de las formas de vida de los individuos de una sociedad. Medir el crecimiento sostenido de cierto país o nación de la economía y de algunos sectores productivos más representativos de tal economía, permite establecer que tan lejos o que tan cerca se está del desarrollo social.

¿Por qué es importante?

El crecimiento económico de un país habla del bienestar [6], la forma más tradicional de medir esto es a través del Producto Interno Bruto (PIB). El crecimiento económico medido así se considera muy bueno, por que guarda relación con la cantidad de bienes materiales disponibles y mejoría del nivel de vida de las personas.

1.4. Indicadores del crecimiento económico.

Casi siempre, la actividad económica de un país tiende a expandirse en el tiempo. Sin embargo, no crece todo el tiempo, si no que muestra periodos de expansión y recesión conocidos como ciclos económicos.

Lo ideal sería que la actividad económica de un país se extendiera a través del tiempo a un buen ritmo, de tal forma que se genere empleo suficiente para los que buscan nuevas oportunidades, pero sin crear presiones inflacionarias. El crecimiento económico produce una demanda de empleos, mientras que el aumento de la población genera una oferta laboral. En general, se busca que la economía crezca lo suficiente para tener una tasa baja de desempleo. Sin embargo, si el crecimiento es muy elevado, la demanda puede exceder la oferta laboral y el ajuste ocasiona presiones inflacionarias. Existen una serie de indicadores que son de gran importancia desde hace muchos años, por que anticipan la diferencia de la actividad económica, proporcionan información relativa a los encargos de tomar decisiones de política económica.

Los indicadores son elementales para evaluar, dar seguimiento y predecir tendencias de la situación de un país, un estado o una región en lo referente a su economía, sociedad, desarrollo humano, etc. Así como para valorar el desempeño institucional encaminado a lograr las metas y objetivos fijados en cada uno de los ámbitos de acción de los programas de gobierno.

1.4.1. Definiciones básicas para comprender el índice de crecimiento económico.

Definición 1.5. (Producción Agregada). Es la suma de toda la producción de un país determinado.

Definición 1.6. (Inflación). Es un proceso provocado por el desequilibrio existente entre la producción y la demanda es la causa de una subida continua de los precios de la mayor parte de los productos y servicios, y una pérdida del valor del dinero para poder adquirirlos o hacer uso de ellos [10].

Definición 1.7. (Tasa de desempleo). El INEGI se refiere a la tasa de desempleo como la desocupación y menciona que es el porcentaje de la población económicamente activa que se encuentra desempleada pero que busca trabajo activamente en ciertas zonas de la población [10].

Definición 1.8. (Índice nacional de precios al consumidor “INPC”). El Índice nacional de precios al consumidor es un indicador económico que se emplea recurrentemente, cuya finalidad es medir a través del tiempo la variación de los precios de una canasta fija de bienes y servicios representativas del consumo de los hogares [10].

El INPC es el instrumento estadístico por medio del cual se mide el fenómeno económico que se conoce como inflación.

Definición 1.9. (Producto Interno Bruto Real “PIBR”). Es un indicador que nos muestra la cantidad de bienes y servicios finales generados con los factores de producción de un país y un periodo de tiempo dado, aunque estos factores se encuentren fuera de dicho país [6].

Definición 1.10. (Producto Nacional Bruto Real “PNBR”). Es el valor monetario de todos los bienes y servicios producidos por un país o una economía valorados a precios constantes, es decir, según los precios del año que se toma como base o en las comparaciones.

1.5. Indicador Global de la Actividad Económica (IGAE).

El PIB es la medida más comprensiva de la actividad económica de un país y por lo mismo, el mejor indicador de su comportamiento general y del tamaño en sí de la economía, es decir, es el indicador de los indicadores económicos. Pero existe otro indicador de suma importancia que se encarga de medir el crecimiento económico y este es el Indicador Global de la Actividad Económica (IGAE).

Para poder medir entre un indicador comprensivo de la actividad económica del país (el PIB) y la necesidad de contar con información más oportuna, el INEGI construyó uno indicador mensual con información preliminar y parcial, que utiliza el mismo esquema conceptual y metodológico de las cuentas nacionales del país, de tal forma que sus resultados presentan una alta correlación con los del Producto Interno Bruto (PIB) trimestral. Por lo mismo, el IGAE constituye un excelente indicador de comportamiento del producto.

1.5.1. Concepto de Indicador Global de la Actividad Económica (IGAE).

Definición 1.11. (IGAE). El Indicador Global de la Actividad Económica permite conocer y dar seguimiento a la evolución mensual del sector real de la economía. Para su cálculo se utilizan: el esquema conceptual, los criterios metodológicos, la clasificación de actividades económicas y las fuentes de información, que se emplean en los cálculos anuales y trimestrales del Producto Interno Bruto.

Este cálculo se alinea con las cifras anuales utilizando la técnica Denton. Incorpora a las Actividades Primarias, Secundarias y Terciarias, a excepción de: la pesca, el aprovechamiento forestal, los corporativos y otras actividades de servicios.

1.5.2. Producto Interno Bruto (PIB) contra el Indicador Global de la Actividad Económica (IGAE)

Para poder hacer la comparación se deben revisar una gran lista de puntos que ayudan a determinar cual indicador es el mejor, el valor del PIB se calcula en miles de millones de pesos y representan la totalidad de la producción del país para un periodo determinado.

El IGAE se presenta como un índice mensual con el 2013 (mismo año base que tiene el PIB) igual a 100. Aun así, se debe subrayar que el método de cálculo tanto del total del IGAE como de los totales de las actividades primarias, secundarias y terciarias es igual al del PIB, es decir, se obtienen por suma de los valores a precios constantes (base 2013) de las clases de actividad, expresándolos como números índices para su publicación. Por lo mismo, el IGAE no se construye mediante índices ponderados fijos.

Si se quiere comparar el IGAE con el PIB, primero se debe trimestralizar al tomar el promedio de los tres índices mensuales para cada trimestre y, después, convertir al PIB en un índice con base 2013 = 100. Una de las ventajas principales que tiene el IGAE sobre el PIB es que el tiempo de su emisión, es mensual, lo que permite obtener información a menor plazo respecto al PIB [6].

1.5.3. Cálculo del IGAE

El método utilizado para calcular el Indicador Global de la Actividad Económica es el de la producción, la base del cálculo consiste en elaborar índices mensuales de volumen físico de la producción para cada una de las clases seleccionadas, con base fija en el año 2013, con estos índices se realiza la extrapolación del valor agregado bruto del año base, asumiendo el supuesto de que el valor agregado bruto tiene el mismo comportamiento que el valor bruto de la producción.

El consumo intermedio a precios constantes, se calculó a partir de la aplicación de relaciones de insumo-producto fijas, tomadas de la serie de las cuentas anuales, admitiendo una constancia tecnológica; el valor bruto de la producción constante se obtuvo de sumar el valor agregado bruto constante y el valor del consumo intermedio constante.

El cálculo del índice de volumen físico de las actividades industriales, según el origen de las mismas, consiste en elaborar índices mensuales de volumen de la producción real para cada una de las clases que cuentan con información de cantidades producidas, valores de producción y precios a nivel de producto.

1.5.4. Números Índices Utilizados

Definición 1.12. (Número Índice). Un número índice es una medida estadística que recoge la evolución relativa en el periodo t de una magnitud económica, es un valor relativo con una base igual al 100.0% o un múltiplo de 100 tal como el 10, 1 o 1000. También permite comparar una magnitud económica en una zona geográfica respecto de una zona de referencia [10].

Los números índice se usan como un indicador para el cambio relativo de una cosa o grupo de cosas; en el caso de la Contabilidad Nacional se aplica a la producción, al consumo intermedio, al valor agregado, a la inversión, al consumo, a la demanda efectiva, al personal ocupado, etc. En general un número índice es la relación porcentual que mide el cambio de un tiempo a otro en precio, cantidad, valor o algún otro elemento de interés.

Los números índice pueden ser construidos para distintos períodos de tiempo, que van de alta frecuencia (diarios, semanales, mensuales, trimestrales, etc.) a los anuales. Para el caso del cálculo del Indicador Global de la Actividad Económica se elaboran índices de volumen físico de la producción tipo Laspeyres, índices de valor (ingreso real o gasto real) e índices simples basados en un indicador relacionado, expresados base 2013 = 100. índice de volumen físico de la producción tipo Laspeyres [10].

$$Q_{0,n} = \frac{\sum p_0 q_n}{\sum p_0 q_0 / 12} * 100 = \frac{\sum p_0 q_0 \left(\frac{q_n}{q_0}\right)}{\sum p_0 q_0} * 100. \quad (1.1)$$

Donde:

$Q_{0,n}$ = Índice del volumen de producción del periodo n con relación al periodo 0.

0 = Año base.

n = Periodo de referencia.

q_n = Cantidad de un bien producido durante el periodo n (estudio).

q_0 = Cantidad de un bien producido durante el periodo 0 (base).

p_0 = Precio de un bien correspondiente al periodo base.

La observación mensual real de cada clase queda resuelta aplicando la mecánica expuesta a la producción detallada por productos o línea de productos, ya sea que se trate de producción primaria o incluya producción secundaria.

Además del índice del total y sus componentes agropecuario, industrial y servicios, se incluyen 16 actividades, los cuatro índices agregados 4 sectores en las actividades industriales y 8 grupos de sectores de servicios. Bajo estas características, los resultados del IGAE mejoran su correlación con los del PIB Trimestral, por lo cual se constituye en un excelente indicador oportuno del comportamiento mensual del Producto [6].

1.5.5. Datos IGAE

Las cifras mensuales del IGAE se encuentran disponibles en el portal web del Instituto Nacional de Estadística y Geografía, en el apartado del banco de información económica (BIE), desde del mes de Enero de 1993 y se muestran en índices de volumen físico de producción con base fija en el año 2013.

En éste trabajo se toman los datos del IGAE para hacer un análisis estadístico con el propósito de realizar un pronóstico de dicha información, es decir que se tendrá conocimiento anticipado de tal indicador en períodos en los que no se tienen mediciones disponibles, lo que significa que se predicen los valores históricos al futuro. La metodología que se seguirá es la de Box-Jenkins y Redes Neuronales Artificiales, lo cual consiste de un trabajo de series de tiempo aplicado a los datos del indicador IGAE. Al ser un indicador de la actividad económica, forma parte de un análisis macroeconómico, este análisis nos llevará a una serie de conclusiones para analizar entre algunas cosas el impacto de diferentes impuestos y los planes de gasto.

Actualmente conocer las condiciones económicas medianamente el indicador IGAE es de mucha utilidad para algunas áreas de la ciencia en especial el área de la económica y los hacedores de política.

Capítulo 2

Modelos de series de tiempo

2.1. Procesos estocásticos.

Un fenómeno estadístico que ocurre en el tiempo de acuerdo a leyes de probabilidad es llamado un proceso estocástico. Las series de tiempo al ser analizadas entonces pueden ser vistas como una particular realización de esta, producida por un mecanismo de probabilidad subyacente del sistema bajo estudio. En palabras más simples, una serie de tiempo se considerará como una realización de un proceso estocástico, es por eso la importancia de darle una definición formal.

Definición 2.1. (Proceso estocástico). Un proceso estocástico es una colección de variables aleatorias X_t , definidas en un espacio de probabilidad (Ω, F, P) , con $t \in T$, donde t pertenece al conjunto de subíndices $T \subset \mathbb{R}$ (donde T puede ser un conjunto numerable o no-numerable) de tal forma que a cada elemento del conjunto le corresponde una y sólo una variable aleatoria.

Si tenemos que el conjunto de subíndices T , es un subconjunto de los números enteros \mathbb{Z} se dice que es un proceso de tiempo discreto, donde X_t representa el valor del proceso al tiempo t . De igual forma cuando T es un subconjunto no numerable de los números reales entonces se presenta un proceso a tiempo continuo, donde el subíndice t de cada variable aleatoria X_t representa el instante de tiempo en que es observada.

En la Tabla 2.1 se muestran las diferentes clasificaciones de los estados de acuerdo a la naturaleza de las variables.

Tabla 2.1: Clasificación de procesos estocásticos.

$T \backslash X$	Discreta	Continua
Discreta	Cadenas de Markov, Procesos de ramificación	Serie de tiempo
Continua	Procesos de Poisson, Teoría de colas, Procesos de nacimiento y muerte	Procesos Brownianos

Una serie de tiempo es una realización de un proceso estocástico en tiempo discreto, donde los elementos de T están ordenados y corresponden a instantes equidistantes del tiempo.

2.2. Serie de tiempo

Siempre ha existido una enorme curiosidad por tratar de conocer sucesos futuros para así poder prever algún suceso extraño y así poder eliminarlo antes de que este haga un cambio significativo. En la actualidad, esta curiosidad se trata de satisfacer de una manera formal usando herramientas matemáticas, en particular herramientas probabilistas. Este estudio se fundamenta en sucesos ocurridos en el pasado. Esto es así, puesto que un nuevo tipo de inferencia estadística que se hace acerca del futuro de alguna variable o compuesto cambiante se basa en sucesos previos. Una de las técnicas más relevante para hacer inferencias sobre el futuro basado en lo sucedido en el pasado, es la exploración de series de tiempo.

Actualmente el análisis de series de tiempo tiene muchas aplicaciones en diversos campos científicos, son incontables las aplicaciones que tienen la posibilidad de ser citadas, en diversas zonas del entendimiento, como por ejemplo, en economía, física, geofísica, química, electricidad, en demografía, en marketing, en telecomunicaciones, en transporte, etcétera.

En la Tabla 2.2 se muestra algunas aplicaciones que tienen las series de tiempo en algunos campos científicos.

Tabla 2.2: Series de tiempo en diferentes campos científicos.

Series De Tiempo	Ejemplos
Series Económicas	-Precio de un artículo -Tasas de desempleo -Tasa de inflación -Índice de precios
Series Físicas Meteorológicas	-Precipitación del agua -Temperatura máxima diaria -Velocidad del viento
Geofísica	-Series sismológicas
Series demográficas	-Tasa de crecimiento de la población -Tasa de natalidad, mortalidad -Resultados de censos poblacionales
Series de marketing	-Series de demanda, gastos, ofertas
Series de telecomunicación	-Análisis de señales
Series de transporte	-Series de tráfico

Definición 2.2. (Serie de tiempo.) Una serie de tiempo es una secuencia de observaciones X_t sobre intervalos de tiempo separados de manera regular. Mas aun una serie de tiempo $\{X_t\}$ es un proceso estocástico donde el índice está relacionado con el tiempo específico T .

Graficar la base de datos que conforman una series de tiempo muchas veces resulta de gran utilidad para llevar acabo un correcto análisis ya que mediante esta gráfica es posible observar muchas características de las cuales las mas comunes son discontinuidades, irregularidades o datos atípicos.

Regularmente las series de tiempo se encuentran en base de datos que casi siempre se pueden observar de manera anual, semestral, cuatrimestral, trimestral, bimestral o mensual sin omitir casos como quincenales o diarios aunque estos últimos son menos comunes. A menudo, los datos de las series de tiempo se examinan con la esperanza de descubrir algún patrón permanente que se pueda aprovechar para preparar un pronóstico. Con objeto de identificar dicho patrón es conveniente muchas veces pensar que la serie de tiempo consta de varios componentes que a continuación se definen.

Definición 2.3. (Tendencia). La tendencia hace referencia al desplazamiento hacia arriba o hacia bajo que caracteriza a las series de tiempo con en relación a una época de tiempo. Esta refleja el incremento o declinación de extensa duración en las series de tiempo.

Definición 2.4. (Ciclo.) El ciclo se refiere a los movimientos hacia arriba o hacia

abajo alrededor de los niveles de tendencia. Estas fluctuaciones tienen una duración entre 2 a 10 años.

Definición 2.5. (Variaciones estacionales.) Las variaciones estacionales son patrones periódicos en una serie de tiempo, en palabras simples, son tendencias que se repiten por algún factor que va guiado a través de los datos.

Definición 2.6. (Fluctuaciones irregulares.) Las fluctuaciones irregulares son movimientos erráticos que siguen un patrón indefinido o regular.

Cabe resaltar que los componentes definidos anteriormente no siempre se presentan solos ya que estos pueden presentarse en combinación o bien todos juntos.

Ya descritos los componentes que influyen en una serie de tiempo, la podemos representar mediante la ecuación.

$$X_t = m_t + s_t + \epsilon_t. \quad (2.1)$$

Donde:

- t es el tiempo.
- X_t es la variable observada
- m_t es la tendencia del modelo.
- s_t es el componente estacional, que constituya la señal o parte determinística.
- ϵ_t es el error, ruido o parte aleatoria del modelo.

2.3. Autocovarianza y Autocorrelación

Anteriormente se mencionó que las series de tiempos son realizaciones cuyo objetivo es determinar el comportamiento de ciertos datos y con estos tratar de predecir o pronosticar el futuro, para lograr una predicción adecuada no es posible usar cualquier proceso estocástico, es decir, este debería de contar con cierta seguridad, pero si a cada momento de tiempo se observa un comportamiento distinto y desequilibrado, éste proceso no va a ser de gran utilidad para poder hacer lo que se requiere.

Se dice que un proceso estocástico se puede caracterizar cuando se puede determinar el conjunto de funciones de distribución para cada conjunto finito de variables aleatorias del proceso, sin embargo, suele ser complejo determinar las características de un proceso por medio de su función de distribución, por ello, la caracterización se suele hacer a través de los dos primeros momentos (media y varianza) de cada X_t , y la covarianza de X_t y X_s .

Definición 2.7. (Media). Se define la esperanza matemática de un proceso estocástico $\{X_t\}$ al tiempo t , como $\mu_x(t)$ tal que

$$\mu_x(t) = \mathbb{E}[X_t] = \int_{-\infty}^{\infty} x f_t(x) dx, \text{ si } \int_{-\infty}^{\infty} |x| f_t(x) dx < \infty. \quad (2.2)$$

Lo anterior representa el valor esperado de un proceso al tiempo t , que no siempre es igual en cada valor t .

Definición 2.8. (Varianza). Se denomina varianza de un proceso estocástico $\{X_t\}$ al tiempo t , como $\sigma^2(t)$, tal que

$$\sigma^2(t) = Var[X_t] = \mathbb{E}[(X_t - \mu(t))^2] = \int_{-\infty}^{\infty} (x - \mu(t))^2 f_t(x) dx \quad (2.3)$$

si

$$\int_{-\infty}^{\infty} (x - \mu(t))^2 f_t(x) dx < \infty.$$

Una serie de tiempo se puede ver como una muestra aleatoria de observaciones de un proceso estocástico, es así como se determinan los parámetros anteriormente mencionados utilizando los datos de la serie de tiempo, a su vez de la misma forma se obtienen la media muestral y la varianza muestral que a continuación se definen.

Definición 2.9. (Media Muestral). La media muestral se define como

$$\bar{x} = \hat{\mu}_x = \frac{1}{T} \sum_{i=1}^T x_i, \text{ donde, } 1 \leq i \leq T. \quad (2.4)$$

Y es una estimación puntual de la media de la población μ .

Definición 2.10. (Varianza Muestral). La varianza muestral se define como

$$s^2 = \hat{\sigma}_x = \frac{1}{T} \sum_{i=1}^T (x_i - \bar{x})^2. \quad (2.5)$$

Y es una estimación puntual de la varianza de la población σ^2 .

2.3.1. Autocovarianza, autocorrelación y autocorrelación parcial

En general el comportamiento de una variable aleatoria X se caracteriza a través de su función de densidad $f(x)$. De la misma manera, dos variables aleatorias se describen a través de una función de densidad conjunta. En las series de tiempo existe cierta correlación entre las observaciones.

Se entiende por covarianza el costo que indica el nivel de alteración conjunta de dos variables aleatorias en interacción a sus medias. Es el dato primordial para decidir si hay una dependencia entre las dos variables. Una covarianza positiva sugiere que una vez que una variable crece la otra variable también lo hace, o sea, poseen una interacción directa. Sin embargo, una covarianza negativa sugiere que una vez que una variable crece la otra decrece, manteniendo una relación inversa. En palabras más sencillas la covarianza es el valor que refleja en qué cuantía dos variables aleatorias varían de forma conjunta respecto a sus medias.

Definición 2.11. (Estacionariedad). Las series temporales son estacionarias, cuando la media y la variabilidad se mantienen constantes a lo largo del tiempo y la serie no sigue una tendencia.

Definición 2.12. (Covarianza). La covarianza está definida como

$$\begin{aligned} Cov(X_t, X_{t+k}) &= Cov(X_{t-t}, X_{t+k-t}) \\ &= Cov(X_0, X_k) \\ &= Cov(X_0, X_{|k|}) \\ &= \gamma_k. \end{aligned} \tag{2.6}$$

De esta forma se denota que la autocovarianza sólo depende del lapso k .

Definición 2.13. (Autocovarianza). La función de autocovarianzas de un proceso estocástico es una función que describe las covarianzas entre las variables del proceso en cada par de instantes:

$$\gamma_k = Cov(X_t, X_{t+k}) = \mathbb{E}[(X_t - \mu)(X_{t+k} - \mu)]. \tag{2.7}$$

Con la colección de valores γ_k , $K = 1, 2, 3, \dots$ que se denomina función de autocovarianza ya antes mencionada por la cual se obtiene el coeficiente de autocorrelación.

En ocasiones en una serie de tiempo acontece, que los valores que toma una variable en el tiempo no son independientes entre sí, sino que un valor determinado depende de los valores anteriores. Entendemos por correlación el valor que nos indica si dos variables están o no relacionadas en forma lineal.

Definición 2.14. (Coeficiente de Autocorrelación). La función de autocorrelación de un proceso estocástico es una función de dos instantes que describe las correlaciones entre las variables en un par de instantes cualesquiera:

$$\rho_k = \frac{\mathbb{E}[(X_t - \mu)(X_{t+k} - \mu)]}{\sqrt{\mathbb{E}[(X_t - \mu)^2]\mathbb{E}[(X_{t+k} - \mu)^2]}} = \frac{Cov(X_t, X_{t+k})}{Var(X_t)} = \frac{\gamma_k}{\gamma_0}. \tag{2.8}$$

Recordando que $Var(X_t) = Var(X_{t+k}) = \gamma_0$.

La colección de los valores de ρ_k , $k = 1, 2, 3, \dots$ es llamada función de autocorrelación. Con lo anterior, las funciones de autocovarianza γ_k y autocorrelación tiene las siguientes propiedades [6]:

1. $\gamma_0 = \text{Var}(X_t)$; $\rho_0 = 1$.
2. $|\gamma_k| \leq \gamma_0$; $|\rho_k| \leq 1$.
3. $\gamma_k = \gamma_{-k}$ y $\rho_k = \rho_{-k}$ para todo k .
4. Las funciones γ_k y la función de autocorrelación ρ_k son semidefinidas positivas en el sentido que

$$\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \gamma_{|t_i - t_j|} \geq 0, \quad \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho_{|t_i - t_j|} \geq 0$$

para cualquier conjunto de puntos t_1, t_2, \dots, t_n y cualquier número $\alpha_1, \alpha_2, \dots, \alpha_n$.

En la Figura 2.1 se muestra un ejemplo del comportamiento de la función de autocorrelación.

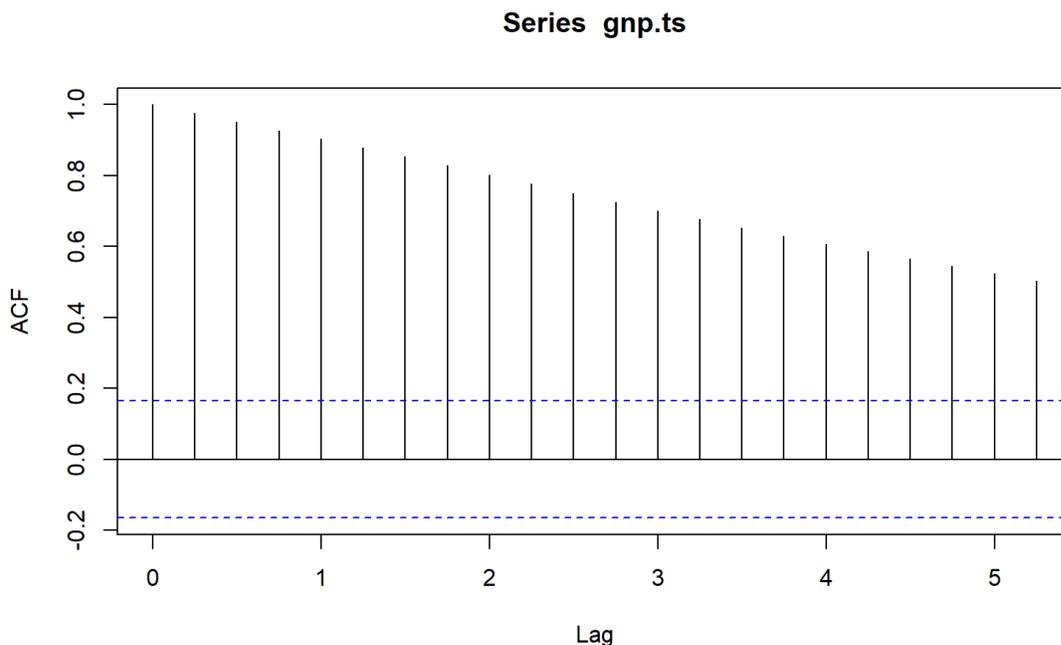


Figura 2.1: Ejemplo de gráfica de autocorrelación.

Definición 2.15. (Función de autocorrelación parcial (FACP)). La autocorrelación parcial de un procesos estacionario, se denota por $\phi_{k,k}$ con $k = 1, 2, \dots$ es la correlación entre X_t y X_{t+k} después de que su dependencia lineal mutua en las que intervienen las variables $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$ se han eliminado. Tenemos que:

$$\phi_{1,1} = \text{Corr}(X_{t+1}, X_t) = \rho_1,$$

$$\phi_{k,k} = \text{Corr}(X_t, X_{t+k} | x_{t+1} \dots x_{t+k-1})$$

las variables X_t y X_{t+k} no están correlacionadas con $X_{t+1}, \dots, X_{t+k-1}$ es decir $\phi_{k,k}$ es la correlación de X_t y X_{t+k} en una dependencia lineal [6].

2.3.2. Series de tiempo estrictamente y débilmente estacionarias

Una serie de tiempo se dice que es estacionaria cuando es estable a lo largo del tiempo, es decir, cuando la media, varianza y covarianza son constantes en el tiempo. Es de suma importancia entender este concepto, se utilizarán los conceptos de estacionariedad estricta y estacionariedad débil.

Definición 2.16. (Serie de tiempo estrictamente estacionaria). Una serie de tiempo X_t , se dice que es un proceso estrictamente estacionario si para cualesquiera dos colecciones finitas $\{X_t, X_{t+1}, \dots, X_{t+k}\}$ y $\{X_{t+h}, X_{t+h+1}, \dots, X_{t+h+k}\}$ tienen la misma distribución para cualquier h en los enteros. Esto significa que:

$$\mathbb{P}\{X_t \leq c_1, \dots, X_{t+k} \leq c_k\} = \mathbb{P}\{X_{t+h} \leq c_1, \dots, X_{t+h+k} \leq c_k\} \quad (2.9)$$

o bien:

$$F_{X_t, X_{t+1}, \dots, X_{t+k}}(c_1, \dots, c_k) = F_{X_{t+h}, X_{t+h+1}, \dots, X_{t+h+k}}(c_1, \dots, c_k).$$

Para toda $k = 1, 2, \dots$, para todo c_1, c_2, \dots, c_k y todos los tiempo desplazados $h = 0, \pm 1, \pm 2, \dots$

En otras palabras tenemos que la distribución de probabilidad conjunta de las observaciones $\{X_t, X_{t+1}, \dots, X_{t+k}\}$ son exactamente la mismas y solo dependen del lapso k [1].

Definición 2.17. (Serie de tiempo débilmente estacionario). Se dice que una serie de tiempo es un proceso débilmente estacionario, si X_t , es un proceso de varianza finita constante tal que cumple con lo siguiente:

1. La función del valor de la media, μ_x es constante y no depende del tiempo t , es decir, $\mathbb{E}(X_t) = \mu, \quad \forall t \in \mathbb{R}, \quad \mu_x < \infty$.
2. La función del valor de la varianza, σ^2 , es constante y no depende del tiempo t , es decir, $\text{Var}(X_t) = \sigma^2, \quad \forall t \in \mathbb{R}, \quad \sigma^2 < \infty$.
3. La función de autocovarianza, $\gamma(k)$, solo depende de su lapso k , es decir, $\text{Cov}(X_t, X_{t+k}) = \gamma_k, \quad \forall t, k \in \mathbb{R}$.

En resumen si una serie de tiempo es estacionaria, su media, su varianza y su auto-covarianza permanecen iguales sin importar el momento en el cual se midan; es decir, son invariantes respecto al tiempo.

Podríamos afirmar que dado cualquier instante de tiempo, el comportamiento que la serie tiene es el mismo sin importar el lapso en que ésta sea observada.

2.4. Series de tiempo no estacionarias

Son series en las cuales la tendencia y/o variabilidad cambian en el tiempo. Los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante.

Definición 2.18. (Serie de tiempo no estacionaria). Se dice que una serie es no estacionaria si se presentan alguna de las siguientes propiedades:

1. El valor de la media, μ_x no es constante.
2. El valor de la varianza, σ^2 , no es constante.
3. Ocurren los dos puntos anteriores en el mismo momento.
4. La covarianza entre cualquier par de variables aleatorias depende tanto de "t" como de "k".

En resumen las series no estacionarias pueden mostrar cambios de varianza, también pueden mostrar tendencia, es decir que la media crece o baja a lo largo de cierto periodo de tiempo y además, pueden presentar efectos estacionales, es decir que el comportamiento de la serie es parecido en ciertos momentos de tiempo. En éste tipo de series sólo se puede estudiar su comportamiento durante un periodo a considerar. Por lo que es imposible hacer un pronostico para otros periodos.

2.5. Ejemplos de series de tiempo estacionarias

2.5.1. Procesos de Ruido Blanco

Supongamos una serie de tiempo denotado por X_t , decimos que es un proceso **estocástico puramente aleatorio** o **proceso estocástico de ruido blanco o caminata aleatoria**, si éste tiene las siguientes propiedades:

1. $\mathbb{E}[X_t] = 0, \forall t.$
2. $Var[X_t] = \mathbb{E}[(X_t - \mu_t)^2]. = \mathbb{E}[(X_t - \mu)]^2 = \mathbb{E}[(X_t)^2]. = \sigma^2.$
3. $Cov[X_t, X_s] = \mathbb{E}[(X_t - \mu_t)(X_t - \mu_s)] = \mathbb{E}[(X_t - \mu)(X_s - \mu)] = \mathbb{E}[X_t X_s] = 0.$

En resumen, un proceso X_t es denominada ruido blanco si su valor promedio es cero, tiene varianza finita y constante, y no le importan sus valores pasados, de tal forma que su valor presente no se ve afectado por sus valores pasados. Por sus propiedades, este proceso es débilmente estacionario o simplemente estacionario si todas las variables aleatorias tiene una media cero, una varianza σ^2 y no existe correlación entre ella y se denotara como $X_t \sim \mathbf{RB}(0, \sigma^2)$.

Con función de autocovarianza

$$\gamma_k = \begin{cases} \sigma^2, & \text{si } k = 0, \\ 0, & \text{si } k \neq 0. \end{cases}$$

Función de autocorrelación

$$\rho_k = \begin{cases} 1 & \text{si } k = 0, \\ 0 & \text{si } k \neq 0 \end{cases}$$

Y función de autocorrelación parcial

$$\phi_{k,k} = \begin{cases} 1 & \text{si } k = 0, \\ 0 & \text{si } k \neq 0 \end{cases}$$

En la figura 2.2 se muestra la trayectoria que realiza un Ruido Blanco.

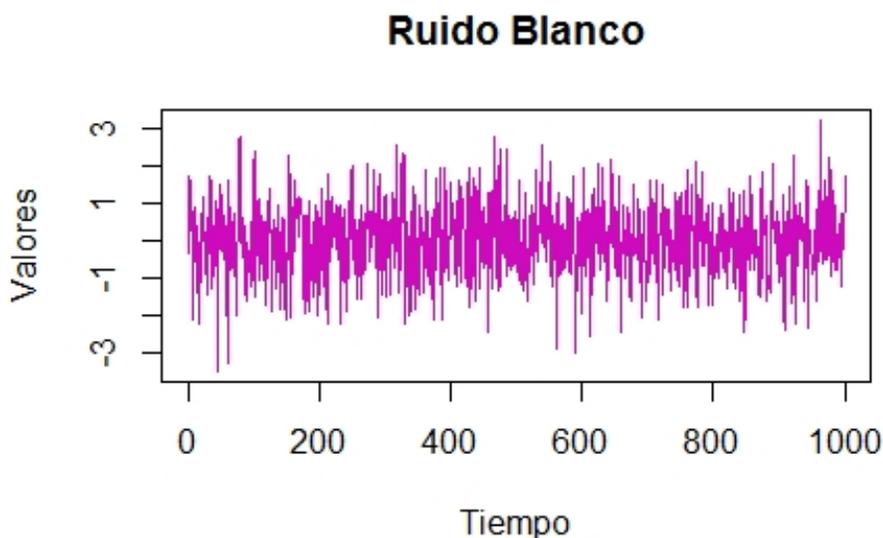


Figura 2.2: Ruido Blanco.

2.5.2. Ruido i.i.d (independiente e idénticamente distribuido).

Otro ejemplo de proceso estacionario es una sucesión \mathbf{X}_t de variables aleatorias independientes idénticamente distribuidas (**v.a.i.i.d**) también denominado **ruido i.i.d**, $\mathbf{X}_t \sim \text{IID}(\mu, \sigma^2)$, donde la propiedad de distribución idéntica significa que a cada uno de los valores de la sucesión tiene la misma distribución de probabilidad y en particular tendrán su media y su varianza iguales.

2.5.3. Caminata Aleatoria.

Sea una caminata Aleatoria o camino Aleatorio, abreviado como RW (Random Walks), es una formalización matemática de la trayectoria que resulta de hacer una serie de bastantes pasos aleatorios y definida como $\{S_n, n = 0, 1, 2, \dots\}$ la cual esta definida por la suma de variables aleatorias idénticamente distribuidas y se dice que tiene una media cero, cuando:

$$S_t = \begin{cases} 0, & \text{si } t = 0 \\ X_1 + X_2 + \dots + X_t, & \text{si } t = 1, 2, \dots \end{cases}$$

y $\{X_t, t = 0, 1, 2, \dots\}$ es un ruido i.i.d con media cero. En la figura 2.3 se muestra la trayectoria que realiza una Caminata Aleatoria.

Algunos ejemplos claros de Caminata Aleatoria es la ruta trazada de una molécula mientras viaja por un líquido o un gas, el camino que sigue un animal en busca de alimento, el precio de una acción fluctuante entre otras pueden tratarse como una caminata aleatoria.

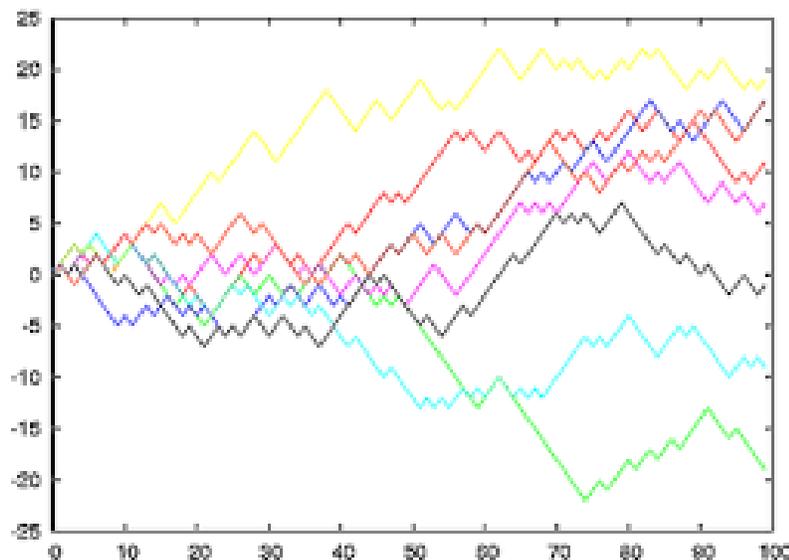


Figura 2.3: Caminata Aleatoria.

2.6. Función de autocorrelación muestral (FAM)

Se han mencionado los parámetros para detectar la dependencia que tienen las observaciones, ahora se muestran los estimadores de las funciones de autocorrelación y autocorrelación parcial, comúnmente a sus gráficas se les conoce como **Correlogramas**.

Se pueden observar las funciones de autocovarianza y autocorrelación de una serie x_1, x_2, \dots, x_T de observaciones como se muestra a continuación.

El estimador de la función de autocovarianza del lapso k es

$$c_k = \hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} (x_t - \bar{x})(x_{t+k} - \bar{x}), \quad k = 0, 1, 2, \dots, T - k, \quad \text{con } k < T.$$

Y el estimador de la función de autocorrelación estimada por la función de autocorrelación muestral (**FAC muestral**)

$$r_k = \hat{\rho}_k = \frac{c_k}{c_0}, \quad k = 0, 1, 2, \dots, T - k, \quad \text{con } k < T.$$

La función de autocorrelación muestral también se aplica a series de tiempo no estacionarias mientras que la función de autocorrelación solo aplica para series de tiempo estacionarias.

Una buena generalización que se tiene para calcular una confiable FAC es que se requieren al menos 50 observaciones, y las autocorrelaciones individuales de la muestra deben estar calculados hasta el periodo k [6].

2.7. Función de Autocorrelación Parcial Muestral (FACMP)

La función de autocorrelación parcial muestral $\hat{\phi}_{k,k}$ está dada por un método recursivo, empezando con $\hat{\phi}_{1,1} = \hat{\rho}_1$ y así calcular $\hat{\phi}_{k,k}$.

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{i=1}^k \hat{\phi}_{ki} \hat{\rho}_{k+1-i}}{1 - \sum_{i=1}^k \hat{\phi}_{ki} \hat{\rho}_i}.$$

y

$$\hat{\phi}_{k+1,i} = \hat{\phi}_{ki} - \hat{\phi}_{k+1,k+1} \hat{\phi}_{k,k+1-i} \quad \text{con } i = 1, \dots, k.$$

siempre que el proceso sea una secuencia de ruido blanco, la varianza de $\hat{\phi}_{k,k}$ puede ser aproximado por

$$\text{Var}(\hat{\phi}_{k,k}) \simeq \frac{1}{n}.$$

2.8. Modelo Autorregresivo de Promedios Móviles (ARMA)

2.8.1. Introducción

Los procesos autoregresivos se originaron en el trabajo de Cochrane y Orcutt de 1994, mediante el cual se analizó los residuales de una regresión clásica como un proceso autoregresivo. En este apartado se explica como se forman modelos lineales basados en la teoría de regresión clásica.

En los modelos de descomposición $X_t = M_t + S_t + \epsilon_t$, $t = 1, 2, \dots$ se estima $\hat{\epsilon}_t$ y se determina si es o no un ruido blanco mediante algunas pruebas que más adelante se explicaran. Si $\hat{\epsilon}_t$ no fuese un ruido blanco, el siguiente paso es modelar dicha componente mediante tres posibles modelos.

1. Medias Móviles de orden q, MA(q).
2. Autoregresivos de orden p, AR(p).
3. Medias Móviles de Autoregresivos, ARMA(p,q).

Estos tres tipos de modelos varían en su capacidad de capturar distintos tipos de comportamiento de autoregresión.

Comenzaremos explicando el primer modelo con sus diferentes características para poder desarrollar una comprensión básica de las propiedades de los modelos necesarios.

Definición 2.19. (Proceso Lineal). Sea X_t una serie de tiempo, se dice que es un proceso lineal si éste se puede representar como

$$X_t = \sum_{i=-\infty}^{\infty} \theta_i Z_{t-i} \quad (2.10)$$

para toda t, donde Z_t es un ruido blanco y θ_i es una secuencia de constantes con $\sum_{i=-\infty}^{\infty} |\theta_i| < \infty$

La condición $\sum_{i=-\infty}^{\infty} |\theta_i| < \infty$ garantiza que la suma infinita de la Ecuación (2.10) converge.

Lo anterior se simplificará en términos de B que es el operador de retraso, esto último surge de algunos métodos diferentes ([1] [4] y [5]) se define el operador ∇ de diferencias de lapso $k = 1$, por

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t. \quad (2.11)$$

Así

$$BX_t = X_{t-1}.$$

Se define a la potencia del operador B como $B^j(X_t)$ y $\nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t))$, $j \geq 1$ con $\nabla^0(X_t) = X_t$ polinomios en B y en ∇ son manipulados de la misma forma como una función polinómica de variables reales, por ejemplo:

$$\begin{aligned}\nabla^2 X_t &= \nabla(\nabla(X_t)) \\ &= (1 - B)(1 - B)X_t \\ &= (1 - 2B + B^2)X_t \\ &= X_t - 2X_{t-1} + X_{t-2}.\end{aligned}$$

Tomando en cuenta que $BZ_t = Z_{t-1}$, el proceso lineal se puede expresar como

$$X_t = \theta(B)Z_t.$$

donde $\theta(B) = \sum_{i=-\infty}^{\infty} \theta_i B^i$. De lo anterior se desprenden algunos caso particulares, que reciben nombres específicos, por ejemplo, cuando $\theta_i = 0$ para toda $i < 0$, a ese proceso se le conoce como **Proceso de medias móviles o MA(∞)**.

2.8.2. Procesos de media móvil de primer orden o Proceso MA(1)

Este proceso es denominado proceso de medias móviles de orden 1, se denota como MA(1) y tiene como estructura:

$$X_t = \mu + \epsilon_t + \theta\epsilon_{t-1}, \quad t = 0, \pm 1, \pm 2, \dots \quad (2.12)$$

donde ϵ_t es un ruido blanco.

Se puede observar que $\mathbb{E}[X_t] = \mu$ y que su $Var[X_t] = (1 + \theta^2)\sigma_\epsilon^2$

$$\begin{aligned}\mathbb{E}[X_t] &= \mu + \mathbb{E}[\epsilon_t + \theta\epsilon_{t-1}] \\ &= \mu + \mathbb{E}[\epsilon_t] + \mathbb{E}[\theta\epsilon_{t-1}] \\ &= \mu \quad \forall t \in \mathbb{Z}\end{aligned} \quad (2.13)$$

y su varianza

$$\begin{aligned}\gamma_0 &= Var[X_t] \\ &= \mathbb{E}[(X_t - \mu)^2] = \mathbb{E}[(\mu + \epsilon_t + \theta\epsilon_{t-1} - \mu)^2] \\ &= \mathbb{E}[(\epsilon_t + \theta\epsilon_{t-1})^2] \\ &= \mathbb{E}[\epsilon_t^2] + \mathbb{E}[2\epsilon_t\theta\epsilon_{t-1}] + \mathbb{E}[\theta^2\epsilon_{t-1}^2] \\ &= \sigma^2 + \theta^2\sigma^2 \\ &= \sigma^2(1 + \theta^2).\end{aligned} \quad (2.14)$$

En cuanto la función de autocovarianza para γ_1 , se tiene:

$$\begin{aligned}
\gamma_1 &= Cov(X_t, X_{t-1}) \\
&= \mathbb{E}[(X_t - \mu)(X_{t-1} - \mu)] \\
&= \mathbb{E}[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-1} + \theta\epsilon_{t-2})] \\
&= \mathbb{E}[\epsilon_t\epsilon_{t-1} + \theta\epsilon_{t-1}^2 + \theta\epsilon_t\epsilon_{t-2} + \theta^2\epsilon_{t-1}\epsilon_{t-2}] \\
&= \mathbb{E}[\epsilon_t\epsilon_{t-1}] + \theta\mathbb{E}[\epsilon_{t-1}^2] + \theta\mathbb{E}[\epsilon_t\epsilon_{t-2}] + \theta^2\mathbb{E}[\epsilon_{t-1}\epsilon_{t-2}] \\
&= \theta\sigma^2.
\end{aligned} \tag{2.15}$$

Para γ_2 se tiene que:

$$\begin{aligned}
\gamma_2 &= Cov(X_t, X_{t-2}) \\
&= \mathbb{E}[(X_t - \mu)(X_{t-2} - \mu)] \\
&= \mathbb{E}[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-2} + \theta\epsilon_{t-3})] \\
&= \mathbb{E}[\epsilon_t\epsilon_{t-2} + \theta\epsilon_{t-1}\epsilon_{t-2} + \theta\epsilon_t\epsilon_{t-3} + \theta^2\epsilon_{t-1}\epsilon_{t-3}] \\
&= 0.
\end{aligned} \tag{2.16}$$

Finalmente tenemos que

$$\gamma_k = \begin{cases} (1 + \theta^2)\sigma_\epsilon^2, & \text{si } k = 0, \\ -\theta\sigma_\epsilon^2, & \text{si } k = 1. \\ 0, & \text{si } k \geq 2. \end{cases} \tag{2.17}$$

y la serie MA(1) es estacionaria, con función de autocorrelación:

$$\rho_k = \begin{cases} 1, & \text{si } k = 0, \\ \frac{\theta}{1+\theta^2}, & \text{si } k = 1. \\ 0, & \text{si } k \geq 2. \end{cases} \tag{2.18}$$

Se define la función $g(\theta) = -\frac{\theta}{(1+\theta^2)}$ y vemos que es monótona decreciente en θ y se puede observar que el máximo valor absoluto que puede tomar ρ_1 en un modelo MA(1) es $\frac{1}{2}$ y este es el único valor no nulo de su función de autocorrelación simple, siendo negativo si $\theta > 0$ y positivo en caso contrario.

De acuerdo a la función de autocorrelación parcial se tiene que para un modelo MA(1) en general es

$$\phi_{k,k} = \frac{-(\theta)^k(1 - \theta^2)}{1 - \theta^{2(k+1)}}. \tag{2.19}$$

Obsérvese que la FACP de un proceso MA(1), contrariamente a su FAC, no se corta después del lapso (retardo) uno, ésta más bien se desvanece exponencialmente dependiendo del signo de θ_1 . En la figura 2.4 se muestra las diferentes trayectorias para el proceso MA(1) y el comportamiento de FAC y FACP.

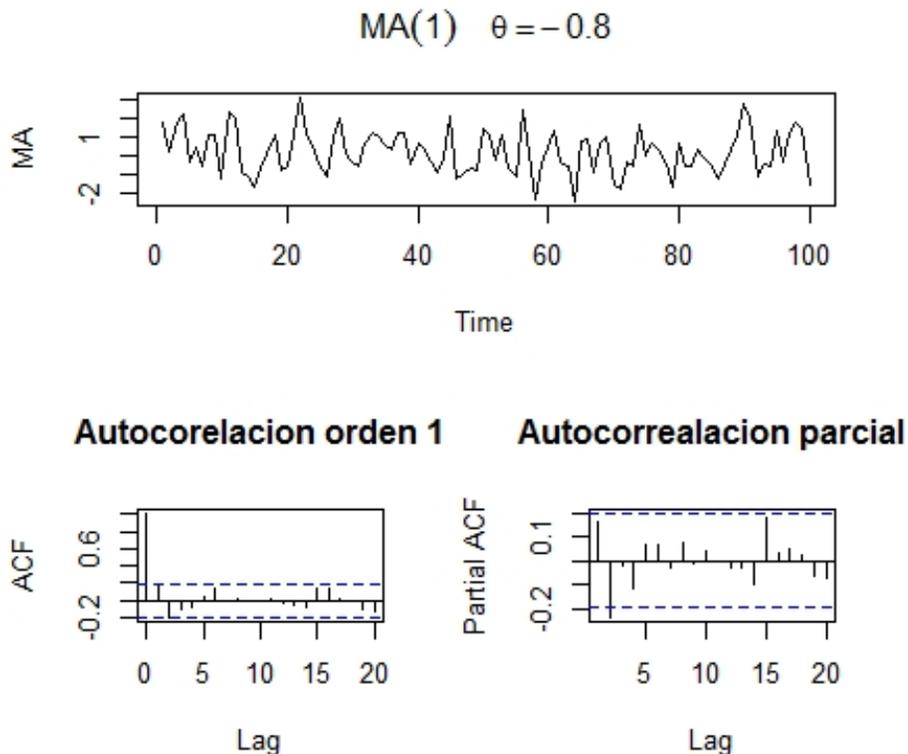


Figura 2.4: Ruido Blanco.

2.8.3. Procesos de media móvil de segundo orden o Proceso MA(2)

Se presentan las características más importantes del proceso de medias móviles de orden 2 o MA(2) el cual está definido como:

$$X_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

donde ϵ_t es un ruido blanco.

Se observa que $\mathbb{E}[X_t] = \mu$ y que su $Var[X_t] = (1 + \theta_1^2 + \theta_2^2)\sigma^2$

$$\begin{aligned} \mathbb{E}[X_t] &= \mu + \mathbb{E}[\epsilon_t] + \theta_1 \mathbb{E}[\epsilon_{t-1}] + \theta_2 \mathbb{E}[\epsilon_{t-2}] \\ &= \mu, \quad \forall t. \end{aligned} \tag{2.20}$$

Y su varianza

$$\begin{aligned}
\gamma_0 &= Var[X_t] \\
&= \mathbb{E}[(X_t - \mu)^2] \\
&= \mathbb{E}[(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2})^2] \\
&= \mathbb{E}[\epsilon_t^2 + \theta_1^2\epsilon_{t-1}^2 + \theta_2^2\epsilon_{t-2}^2 + \text{productos cruzados}] \\
&= \sigma^2 + \theta_1^2\sigma^2 + \theta_2^2\sigma^2 + 0 \\
&= \sigma^2(1 + \theta_1^2 + \theta_2^2).
\end{aligned} \tag{2.21}$$

En cuanto su función de autocovarianza para γ_1 , se tiene:

$$\begin{aligned}
\gamma_1 &= Cov(X_t, X_{t-1}) \\
&= \mathbb{E}[(X_t - \mu)(X_{t-1} - \mu)] \\
&= \mathbb{E}[(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2})(\epsilon_{t-1} + \theta_1\epsilon_{t-2} + \theta_2\epsilon_{t-3})] \\
&= \mathbb{E}[\theta_1\epsilon_{t-1}^2 + \theta_1\theta_2\epsilon_{t-2}^2 + \text{productos cruzados}] \\
&= (\theta_1 + \theta_1\theta_2)\sigma^2.
\end{aligned} \tag{2.22}$$

Para γ_2 se tiene:

$$\begin{aligned}
\gamma_2 &= Cov(X_t, X_{t-2}) \\
&= \mathbb{E}[(X_t - \mu)(X_{t-2} - \mu)] \\
&= \mathbb{E}[(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2})(\epsilon_{t-2} + \theta_1\epsilon_{t-3} + \theta_2\epsilon_{t-4})] \\
&= \mathbb{E}[\theta_2\epsilon_{t-2}^2 + \text{productos cruzados}] \\
&= \theta_2\sigma^2.
\end{aligned} \tag{2.23}$$

Finalmente tenemos que:

$$\gamma_k = \begin{cases} (1 + \theta_1^2 + \theta_2^2)\sigma^2, & \text{si } k = 1, \\ -\theta_1\sigma^2, & \text{si } k = 2. \\ 0, & \text{si } k \geq 3. \end{cases} \tag{2.24}$$

Tenemos que la serie MA(2) es estacionaria, con función de autocorrelación:

$$\gamma_k = \begin{cases} 1 & \text{si } k = 0, \\ \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}, & \text{si } k = 1. \\ \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}, & \text{si } k \geq 2. \\ 0, & \text{si } k \geq 2. \end{cases} \tag{2.25}$$

2.8.4. Procesos de media móvil de orden q o Proceso MA(q)

Este proceso es denominado proceso de medias móviles de orden q , que se denota como MA(q) y tiene como estructura:

$$X_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \dots + \theta_q\epsilon_{t-q} \quad (2.26)$$

donde ϵ_t es un ruido blanco. Se puede comprobar que $\mathbb{E}[X_t] = \mu$ y que su $Var[X_t] = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)\sigma^2$.

En cuanto a su función de autocovarianza para γ_1 se tiene:

$$\gamma_1 = (\theta_1 + \theta_1\theta_2 + \theta_2\theta_3 + \dots + \theta_{q-1}\theta_q)\sigma^2$$

$$\gamma_2 = (\theta_2 + \theta_1\theta_3 + \theta_2\theta_4 + \dots + \theta_{q-2}\theta_q)\sigma^2$$

$$\gamma_s = (\theta_s + \theta_1\theta_{1+s} + \theta_2\theta_{2+s} + \dots + \theta_{q-s}\theta_q)\sigma^2 \quad si \quad s < q$$

$$\gamma_s = \theta_s\sigma^2 \quad si \quad s = q$$

$$\gamma_s = 0 \quad si \quad s > q.$$

De la covarianza anterior se obtiene la función de autocorrelación, tomando en cuenta que:

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2}, & \text{para } k = 1, 2, \dots, q \\ 0, & \text{para } k > q. \end{cases}$$

La característica de la función de autocorrelación es muy útil para identificar a un modelo MA. En una base de datos especifica cuando el lapso sea mayor a q , no siempre va a tomar el valor de cero pero lo esperado es que llegue a ser muy pequeño en valores absolutos después del lapso q .

2.8.5. Estacionariedad del Proceso MA

Los procesos MA al ser una suma de procesos estacionarios, son siempre estacionarios. Siempre y cuando las raíces de su ecuación característica no tenga módulo mayor que la unidad, se da la paradoja de que el efecto de las observaciones pasadas aumentaría con el tiempo.

Los procesos MA son estacionarios siempre que cumplan ciertas condiciones:

- La propiedad de estacionariedad se cumple dependiendo de los valores de los parámetros.

- En caso de los $MA(q)$, se requiere que la suma del valor absoluto de los θ sea menor que ∞ .

$$\sum_{j=0}^{\infty} |\theta_j| < \infty.$$

Esto es,

$$\sum_{j=0}^q |\theta_j| < \infty.$$

Asumiendo que $\theta_0 = 1$.

2.8.6. Ejemplo $MA(q)$

A continuación se muestra un ejemplo sobre el crecimiento que ha tenido México en la llegada de barcos en los últimos años en la Coordinación General de Puertos y Marina Mercante.

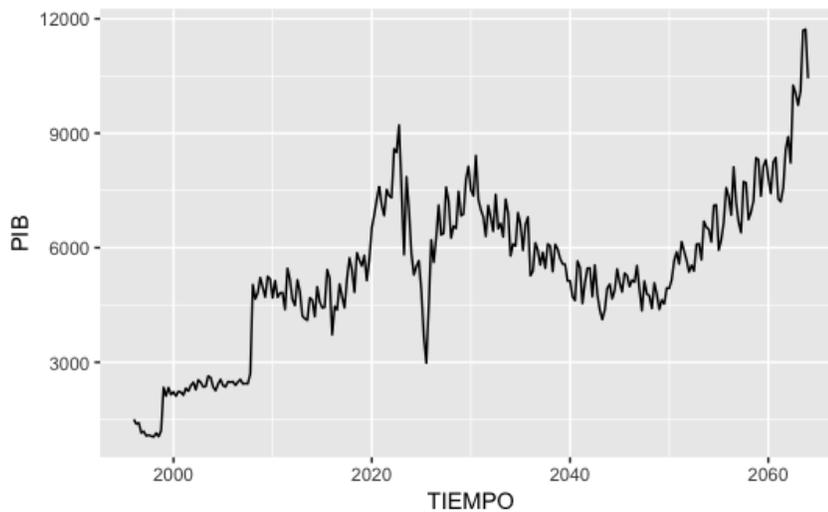


Figura 2.5: Llegada de barcos a puertos mexicanos.

En la Figura 2.5 se observa que es una serie que cuenta con una tendencia creciente y tiene un periodo cíclico que tiene una tendencia creciente.

En la Figura 2.6 se muestra la descomposición de la serie del PIB de las embarcaciones en sus componentes (Tendencia, Ciclo y Estacionalidad).

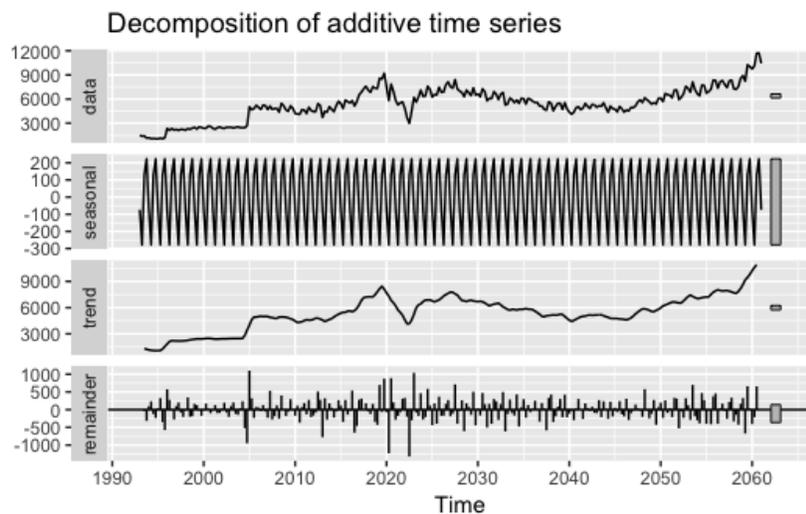


Figura 2.6: Descomposición de la serie.

Finalmente en la Figura 2.7 se muestran las medias móviles $MA(3)$ $MA(5)$, $MA(7)$ y $MA(9)$.

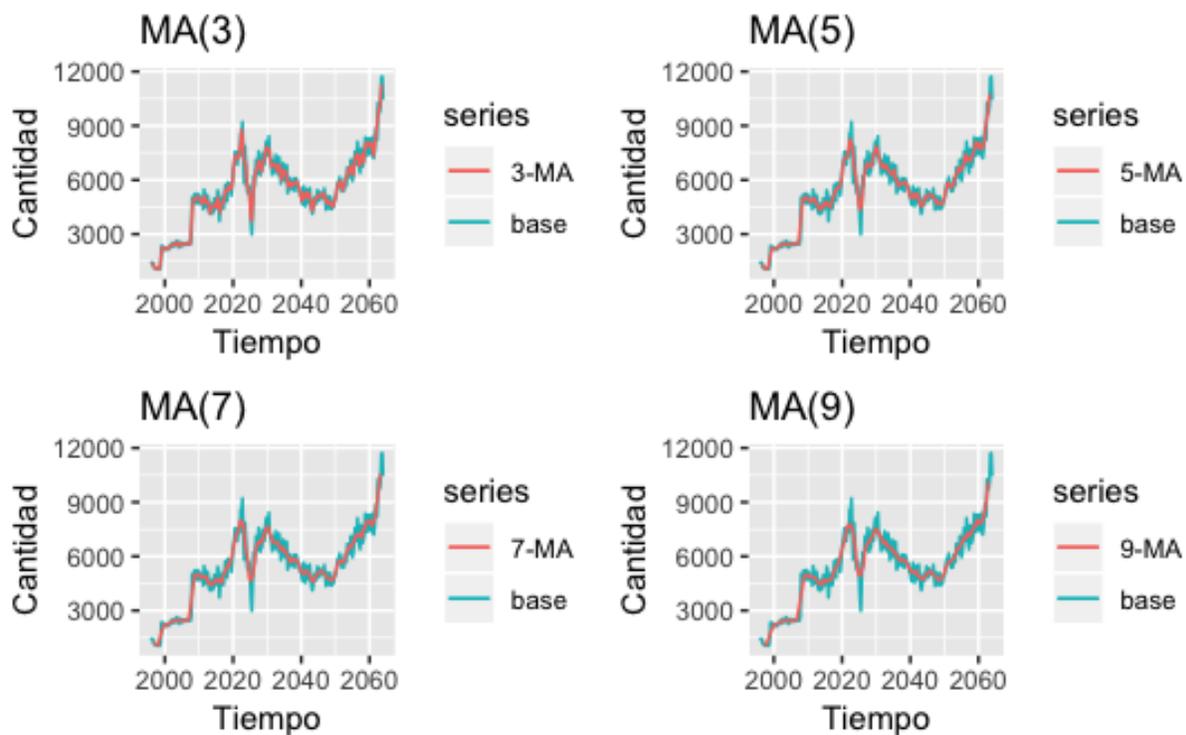


Figura 2.7: Medias móviles de la serie.

2.8.7. Proceso Autorregresivo de primer orden o Proceso AR(1)

Este proceso es denominado proceso autorregresivo de orden 1, y es denotado como AR(1) y tiene como estructura

$$X_t = c + \phi X_{t-1} + \epsilon_t \quad \text{con } |\phi| < 1. \quad (2.27)$$

Se puede observar que $\mathbb{E}[X_t] = \mu = \frac{c}{1-\phi}$.

Ahora despejando $c = \mu(1 - \phi)$, y remplazando en el modelo AR(1),

$$X_t = \mu(1 - \phi) + \phi X_{t-1} + \epsilon_t$$

Reordenando los términos tenemos el modelo AR(1) en desviaciones:

$$\begin{aligned} X_t - \mu &= \phi(X_{t-1} - \mu) + \epsilon_t \\ \tilde{X}_t &= \phi\tilde{X}_{t-1} + \epsilon_t \end{aligned}$$

y su varianza

$$\begin{aligned} \gamma_0 &= \text{Var}[X_t] \\ &= \mathbb{E}[(X_t - \mu)^2] = \mathbb{E}[(\tilde{X}_t)^2] = \mathbb{E}[(\phi\tilde{X}_{t-1} + \epsilon_t)^2] \\ &= \mathbb{E}[\phi^2\tilde{X}_{t-1}^2 + \epsilon_t^2 + 2\phi\tilde{X}_{t-1}\epsilon_t] \\ &= \phi^2\mathbb{E}[\tilde{X}_{t-1}^2] + \mathbb{E}[\epsilon_t^2] + 2\phi\mathbb{E}[\tilde{X}_{t-1}\epsilon_t] \\ &= \phi^2\gamma_0 + \sigma^2 \\ \gamma_0 &= \frac{\sigma^2}{1 - \phi^2}. \end{aligned} \quad (2.28)$$

Sus autocovarianzas están dadas por:

$$\begin{aligned} \gamma_1 &= \mathbb{E}[(X_t - \mu)(X_{t-1} - \mu)] \\ &= \mathbb{E}[\tilde{X}_t\tilde{X}_{t-1}] = \mathbb{E}[(\phi\tilde{X}_{t-1} + \epsilon_t)\tilde{X}_{t-1}] = \phi\mathbb{E}[\tilde{X}_{t-1}^2] + \mathbb{E}[\epsilon_t\tilde{X}_{t-1}] \\ &= \phi\gamma_0. \end{aligned} \quad (2.29)$$

Para γ_2

$$\begin{aligned} \gamma_2 &= \mathbb{E}[(x_t - \mu)(X_{t-2} - \mu)] \\ &= \mathbb{E}[\tilde{X}_t\tilde{X}_{t-2}] = \mathbb{E}[(\phi\tilde{X}_{t-1} + \epsilon_t)\tilde{X}_{t-2}] \\ &= \phi\mathbb{E}[\tilde{X}_{t-1}\tilde{X}_{t-2}] + \mathbb{E}[\epsilon_t\tilde{X}_{t-2}] \\ &= \phi\gamma_1 \\ &= \phi^2\gamma_0. \end{aligned} \quad (2.30)$$

Ahora si $s > 1$, se tiene que:

$$\begin{aligned} \gamma_s &= \phi\gamma_{s-1} \\ &= \phi^s\gamma_0. \end{aligned} \quad (2.31)$$

Para calcular las autocorrelaciones, al tratarse de series estacionarias bajo un mismo supuesto se divide todo por γ_0 , entonces se tiene que:

$$\rho_s = \phi^s \quad \text{para } s = 0, 1, 2, \dots \quad (2.32)$$

Las autocorrelaciones, como se muestran, van decayendo ya que se asume que $|\phi| < 1$ las potencias de ϕ van a hacer que ϕ^s sea cada vez mas pequeño cuando s tiende a infinito.

Las autocorrelaciones van a ser una sucesión de valores que van ir decayendo acercándose a cero asintóticamente

En la Figura 2.8 se muestra las diferentes trayectorias para el proceso autorregresivo de orden 1 AR(1) y el comportamiento de FAC y FACP.

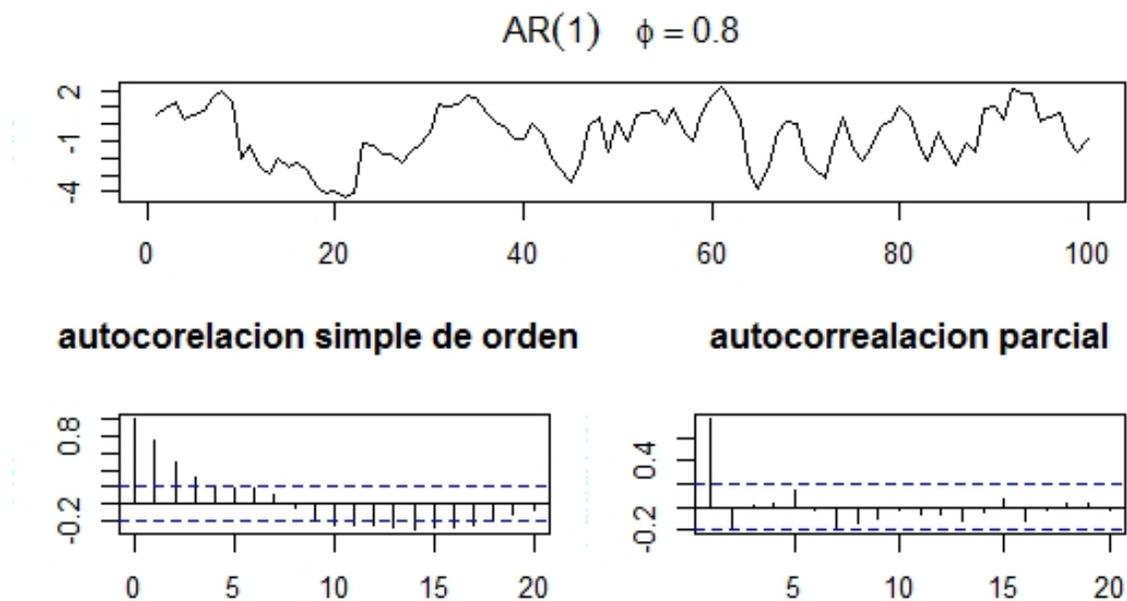


Figura 2.8: Proceso Autorregresivo de orden 1.

2.8.8. Proceso Autorregresivo de Segundo Orden O Proceso AR(2)

Este proceso es denominado proceso autorregresivo de orden 2, y se considera que la serie satisface

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t. \quad (2.33)$$

Se observa que su media está dada por

$$\mathbb{E}[X_t] = c + \phi_1 \mathbb{E}[X_{t-1}] + \phi_2 \mathbb{E}[X_{t-2}] + \mathbb{E}[\epsilon_t].$$

$$\mu = c + \phi_1 \mu + \phi_2 \mu.$$

Despejando se obtiene que

$$\mathbb{E}[X_t] = \mu = \frac{c}{1 - \phi_1 - \phi_2} \quad \phi_1 + \phi_2 \neq 1. \quad (2.34)$$

Donde

$$c = \mu(1 - \phi_1 - \phi_2)$$

$$X_t = \mu(1 - \phi_1 - \phi_2) + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t.$$

Con lo cual el modelo AR(2) usando desviaciones queda de la siguiente manera

$$X_t - \mu = \phi_1 (X_{t-1} - \mu) + \phi_2 (X_{t-2} - \mu) + \epsilon_t.$$

$$\tilde{X}_t = \phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + \epsilon_t.$$

Su varianza

$$\begin{aligned} \gamma_0 &= \text{Var}(X_t) = \mathbb{E}[\tilde{X}_t^2] \\ &= \mathbb{E}[(\phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + \epsilon_t)^2] \\ &= \mathbb{E}[\phi_1^2 \tilde{X}_{t-1}^2 + \phi_2^2 \tilde{X}_{t-2}^2 + \epsilon_t^2 + 2\phi_1 \phi_2 \tilde{X}_{t-1} \tilde{X}_{t-2} + 2\phi_1 \tilde{X}_{t-1} \epsilon_t + 2\phi_2 \tilde{X}_{t-2} \epsilon_t] \\ &= \phi_1^2 \gamma_0 + \phi_2^2 \gamma_0 + \sigma^2 + 2\phi_1 \phi_2 \gamma_1 + 0 + 0 \\ &= \phi_1^2 \gamma_0 + \phi_2^2 \gamma_0 + \sigma^2 + 2\phi_1 \phi_2 \gamma_1. \end{aligned} \quad (2.35)$$

Aún no se puede obtener una expresión definitiva ya que está en función de γ_1 .

Calculando γ_1 se obtiene:

$$\begin{aligned} \gamma_1 &= \mathbb{E}[(X_t - \mu)(X_{t-1} - \mu)] = \mathbb{E}[\tilde{X}_t \tilde{X}_{t-1}] \\ &= \mathbb{E}[(\phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + \epsilon_t) \tilde{X}_{t-1}] \\ &= \mathbb{E}[\phi_1 \tilde{X}_{t-1}^2 + \phi_2 \tilde{X}_{t-1} \tilde{X}_{t-2} + \epsilon_t \tilde{X}_{t-1}] \\ \gamma_1 &= \phi_1 \gamma_0 + \phi_2 \gamma_1. \end{aligned} \quad (2.36)$$

Resolviendo la Ecuación (2.36) se tiene que

$$\gamma_1 = \frac{\phi_1}{1 - \phi_2} \gamma_0 \quad (2.37)$$

Ahora reemplazando (2.37) en (2.35) se tiene finalmente que:

$$\begin{aligned} \gamma_0 &= \phi_1^2 \gamma_0 + \phi_2^2 \gamma_0 + \sigma^2 + 2\phi_1 \phi_2 \left(\frac{\phi_1}{1 - \phi_2} \right) \gamma_0 \\ &= \frac{\sigma^2}{1 - \phi_1^2 - \phi_2^2 - 2\phi_1 \phi_2 \left(\frac{\phi_1}{1 - \phi_2} \right)} \\ &= \frac{(1 - \phi_2) \sigma^2}{(1 + \phi_2) [(1 - \phi_2)^2 - \phi_1^2]}. \end{aligned} \quad (2.38)$$

Para la segunda autocovarianza se tiene

$$\begin{aligned} \gamma_2 &= \mathbb{E}[(X_t - \mu)(X_{t-2} - \mu)] = \mathbb{E}[\tilde{X}_t \tilde{X}_{t-2}] \\ &= \mathbb{E}[(\phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + \epsilon_t) \tilde{X}_{t-2}] \\ &= \phi_1 \gamma_1 + \phi_2 \gamma_0. \end{aligned} \quad (2.39)$$

En el caso de la tercera autocovarianza se tiene que

$$\begin{aligned} \gamma_3 &= \mathbb{E}[(X_t - \mu)(X_{t-3} - \mu)] = \mathbb{E}[\tilde{X}_t \tilde{X}_{t-3}] \\ &= \mathbb{E}[(\phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + \epsilon_t) \tilde{X}_{t-3}] \\ &= \phi_1 \gamma_2 + \phi_2 \gamma_1. \end{aligned} \quad (2.40)$$

Generalizando se tiene que

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} \quad \text{para todo } j \geq 2.$$

Resumiendo,

$$\rho_k = \begin{cases} \frac{\gamma_0}{\gamma_0} = 1, & \text{si } k = 0. \\ \frac{\phi_1}{1 - \phi_2}, & \text{si } k = 1. \\ \phi_1 \rho_1 + \phi_2, & \text{si } k = 2. \\ \dots & \\ \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, & \text{si } k > 2. \end{cases} \quad (2.41)$$

2.8.9. Proceso Autorregresivo de Orden p o Proceso AR(p)

Este proceso es denominado proceso de autorregresivos de orden p que se denotó como AR(p) y tiene como estructura:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t. \quad (2.42)$$

Luego, haciendo la conversión a desviaciones se tiene que:

$$\begin{aligned} X_t - \mu &= \phi_1 (X_{t-1} - \mu) + \phi_2 (X_{t-2} - \mu) + \dots + \phi_p (X_{t-p} - \mu) + \epsilon_t \\ \tilde{X} &= \phi_1 \tilde{X}_{t-1} + \phi_2 \tilde{X}_{t-2} + \dots + \phi_p \tilde{X}_{t-p} + \epsilon_t. \end{aligned}$$

Generalizando lo anterior se tiene que su media es igual a

$$\mathbb{E}[X_t] = \frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_p} \quad \text{con} \quad \phi_1 + \phi_2 + \dots + \phi_p \neq 1. \quad (2.43)$$

Para calcular la varianza se multiplica el AR(p) en desvíos por \tilde{X}_t y se toma $\mathbb{E}[\cdot]$ entonces se tiene que:

$$\text{Var}[X_t] = \gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma^2. \quad (2.44)$$

Autocovarianzas: se multiplica el AR(p) en desvíos por \tilde{X}_{t-1} y tomamos la esperanza $\mathbb{E}[\cdot]$

$$\gamma_1 = \Phi_1 \gamma_0 + \Phi_2 \gamma_1 + \Phi_3 \gamma_2 + \dots + \Phi_p \gamma_{p-1}.$$

Se multiplica el AR(p) en desvíos por $\tilde{X}_{t-2}, \tilde{X}_{t-3}$ etc y se toma la $\mathbb{E}[\cdot]$ por lo tanto tendríamos que:

$$\gamma_2 = \Phi_1 \gamma_1 + \Phi_2 \gamma_0 + \Phi_3 \gamma_1 + \dots + \Phi_p \gamma_{p-2}. \quad (2.45)$$

$$\gamma_p = \Phi_1 \gamma_{p-1} + \Phi_2 \gamma_{p-2} + \Phi_3 \gamma_{p-3} + \dots + \Phi_p \gamma_0. \quad (2.46)$$

De las ecuaciones anteriores se obtienen $\gamma_1, \gamma_2, \dots, \gamma_p$ en función de γ_0 .

Finalmente se calculan las autocorrelaciones dividiendo todo entre γ_0 y se tiene que:

$$\rho_k = \begin{cases} \phi_1 + \phi_2 \rho_1 + \phi_3 \rho_2 + \dots + \phi_p \rho_{p-1}, & \text{si } k = 1. \\ \phi_1 \rho_1 + \phi_2 \rho_0 + \dots + \phi_p \rho_{p-2} & \text{si } k = 2. \\ \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p, & \text{si } k = p. \\ \phi_1 \rho_{s-1} + \phi_2 \rho_{s-2} + \dots + \phi_p \rho_{s-p}, & \text{si } k > p. \end{cases} \quad (2.47)$$

2.9. Proceso Autorregresivo de Medias Móviles (AR-MA)

La combinación de los procesos de $AR(q)$ y $MA(p)$ da lugar a procesos mixtos a los cuales se les denomina como $ARMA(p,q)$ y los cuales se definen como sigue:

Definición 2.20. Se dice que (X_t) es un proceso $ARMA(p,q)$ si es estacionario y si para cada t , se cumple:

$$X_t = c + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}. \quad (2.48)$$

Donde (ϵ_t) es un ruido blanco.

EL modelo $ARMA(p,q)$ es estacionario si cumple que:

1. La parte $MA(p)$ es estacionaria.
2. La parte $AR(q)$ es estacionaria

2.9.1. ARMA(1,1)

La estructura ARMA mas sencilla es el modelo $ARMA(1,1)$:

$$X_t = c + \phi X_{t-1} + \epsilon_t + \theta \epsilon_{t-1}. \quad (2.49)$$

Que dependerá de cuatro parámetros desconocidos ϕ , θ , σ^2 y c . El proceso $ARMA(1,1)$, es estacionario cuando $|\phi| < 1$, e invertible cuando $|\theta| < 1$. Distintos modelos pertenecientes a esta familia pueden escribirse sin más que variar los órdenes (p, q) de los componentes autorregresivos y de medias móviles.

Calculando su media se tiene que:

$$\mathbb{E}[X_t] = c + \phi \mathbb{E}[X_{t-1}] + \mathbb{E}[\epsilon_t] + \theta \mathbb{E}[\epsilon_{t-1}]$$

$$\mu = c + \phi \mu$$

$$\mu = \frac{c}{1 - \phi}.$$

En desviaciones se tiene que el modelo $ARMA(1,1)$ es

$$X_t - \mu = \phi(X_{t-1} - \mu) + \epsilon_t + \theta \epsilon_{t-1}$$

$$\tilde{X}_t = \phi \tilde{X}_{t-1} + \epsilon_t + \theta \epsilon_{t-1}.$$

Calculando su varianza:

$$\begin{aligned} \gamma_0 &= \mathbb{E}[(X_t - \mu)^2] = \mathbb{E}[\tilde{X}_t^2] \\ &= \mathbb{E}[(\phi \tilde{X}_{t-1} + \epsilon_t + \theta \epsilon_{t-1})^2] \\ &= \mathbb{E}[\phi^2 \tilde{X}_{t-1}^2 + \epsilon_t^2 + \theta^2 \epsilon_{t-1}^2 + 2\phi\theta \tilde{X}_{t-1} \epsilon_{t-1} + 2\phi \tilde{X}_{t-1} \epsilon_t + 2\theta \epsilon_t \epsilon_{t-1}] \\ \gamma_0 &= \phi^2 \gamma_0 + \sigma^2 + \theta^2 \sigma^2 + 2\phi\theta \sigma^2 \\ &= \frac{\sigma^2(1 + \theta^2 + 2\phi\theta)}{1 - \phi^2}. \end{aligned} \quad (2.50)$$

Calculando sus autocovarianzas se tiene:

$$\begin{aligned}
 \gamma_1 &= \mathbb{E}[\tilde{X}_t \tilde{X}_{t-1}] \\
 &= \mathbb{E}[(\phi \tilde{X}_{t-1}^2 + \epsilon_t + \theta \epsilon_{t-1}) \tilde{X}_{t-1}] \\
 &= \phi \mathbb{E}[\tilde{X}_{t-1}^2] + \mathbb{E}[\epsilon_t \tilde{X}_{t-1} + \theta \mathbb{E}[\epsilon_{t-1} \tilde{X}_{t-1}]] \\
 \gamma_1 &= \phi \gamma_0 + \theta \sigma^2.
 \end{aligned} \tag{2.51}$$

Para el caso de γ_2 se tiene:

$$\begin{aligned}
 \gamma_2 &= \mathbb{E}[\tilde{X}_t \tilde{X}_{t-2}] \\
 &= \mathbb{E}[(\phi \tilde{X}_{t-1} + \epsilon_t + \theta \epsilon_{t-1}) \tilde{X}_{t-2}] \\
 &= \phi \mathbb{E}[\tilde{X}_{t-1} \tilde{X}_{t-2}] + \mathbb{E}[\epsilon_t \tilde{X}_{t-2} + \theta \mathbb{E}[\epsilon_{t-1} \tilde{X}_{t-2}]] \\
 \gamma_2 &= \phi \gamma_1 \\
 &= \phi^2 \gamma_0 + \phi \theta \sigma^2.
 \end{aligned} \tag{2.52}$$

Generalizando se tiene:

$$\begin{aligned}
 \gamma_j &= \phi \gamma_{j-1} \quad \forall j \geq 2 \\
 &= \phi^j \gamma_0 + \phi^{j-1} \theta \sigma^2.
 \end{aligned}$$

Finalmente obtenemos las autocorrelaciones dividiendo todo entre γ_0 y tenemos que:

$$\rho_j = \begin{cases} \frac{\gamma_1}{\gamma_0} = \phi + \frac{\theta(1-\phi^2)}{1+\theta^2+2\phi\theta}, & \text{si } j = 1. \\ \phi^2 + \phi \frac{\theta(1-\theta^2)}{1+\theta^2+2\phi\theta}, & \text{si } j = 2. \\ \phi \rho_{j-1}, & \text{si } j \geq 2. \end{cases} \tag{2.53}$$

La función de autocorrelación de un proceso $ARMA(1,1)$ comienza en el valor ρ_1 , y a partir de él, decrece a una tasa ϕ . Es decir, dicha función de autocorrelación se comporta, a partir de $k = 1$, como la función de autocorrelación de un proceso $AR(1)$.

En la Figura 2.9 se puede observar las diferentes trayectorias para un proceso $ARMA(1,1)$ y el comportamiento de FAC y FACP, dependiendo de los signos de ϕ_1 y θ_1 .

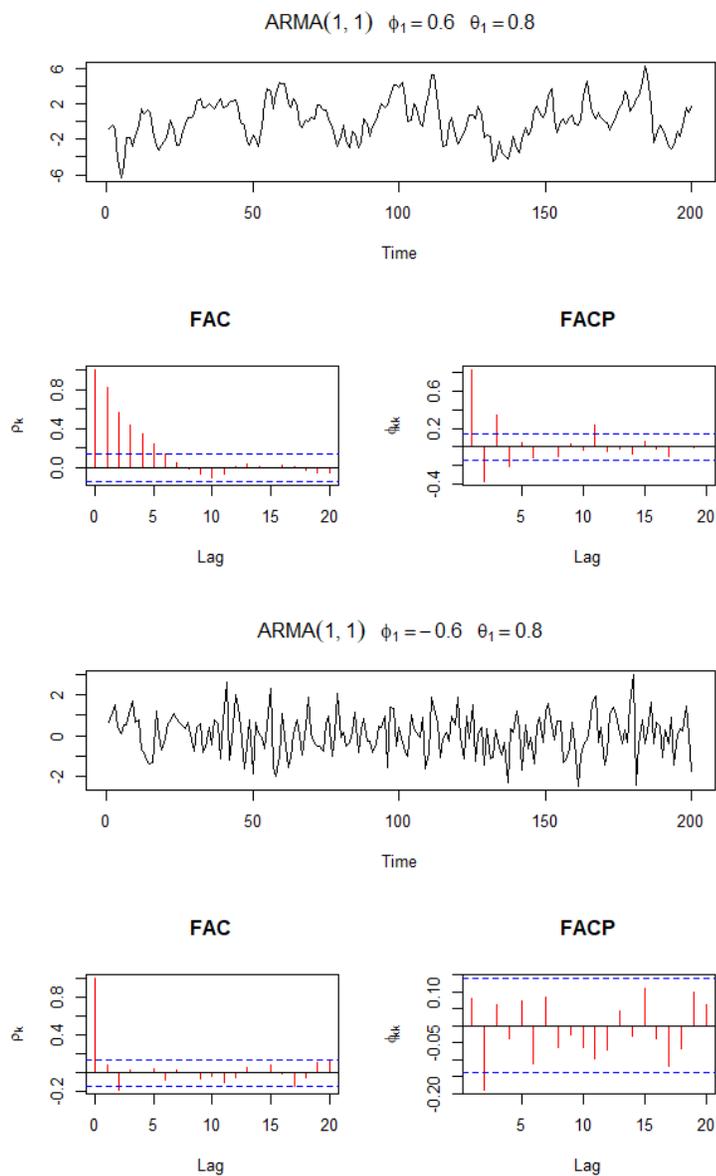


Figura 2.9: Trayectoria para un proceso ARMA(1,1).

Por último, en la Tabla 2.3 se da una descripción de la FAC Y FACP para cada uno de los procesos vistos anteriormente.

Tabla 2.3: Resumen FAC y FACP para los diferentes procesos.

Proceso	FAC	FACP
Media Móvil de orden p	Se corta después del lapso q	Resulta una mezcla de decaimientos exponenciales y ondas sinusoidales amortiguadas, dependiendo de los signos.
Proceso autorregresivo de orden p	Se extingue.	Se corta después del lapso p .
Proceso autorregresivo de medias móviles de orden $(1, 1)$	Se desvanece de manera exponencial.	Se corta de manera que domina el decremento exponencial amortiguado.
Proceso autorregresivo de medias móviles de orden (p, q)	Se extingue.	Se extingue .

2.10. Modelo Autorregresivo de Medias Móviles Integrado *ARIMA*

Diversos investigadores han tratado de desarrollar y aplicar nuevas propuestas para la modelación de series de tiempo, los cuales han logrado probar su eficiencia ya que son capaces de tratar con cualquier patrón de datos y producir pronósticos precisos en base a ciertas características específicas. Las nuevas propuestas de análisis se implementaron en un principio a un número reducido de problemas. Específicamente los modelos desarrollados en las últimas décadas son los llamados autorregresivos (AR), de medias móviles (MA), integrados (I), así como las posibles combinaciones entre estos (ARMA y ARIMA).

La principal diferencia entre estos modelos y los clásicos es el enfoque estocástico que se le da a las series de tiempo, en vez de tratarlas de forma determinista. Bajo este enfoque se concibe a la serie de tiempo como un conjunto de valores de tipo aleatorio, generados a partir de un proceso totalmente desconocido, es decir, se concibe a la serie como un proceso estocástico. Así mismo, derivado del desconocimiento del proceso generador de los datos, el objetivo de este enfoque es tratar de identificar el modelo probabilístico que represente las características principales del comportamiento de la serie.

2.10.1. Modelo autorregresivo integrado de medias móviles de orden p, d, q $ARIMA(p, d, q)$

Al igual que el modelo anterior ($ARMA$) este modelo es la combinación de los modelos de autorregresivos y el de medias móviles con la diferencia de incluir un proceso de restablecimiento de inestabilidad original presente en nuestra serie de tiempo al cual se le denomina integración d . La forma general de un modelo $ARIMA$ se presenta a continuación.

$$X'_t = c + \phi_1 X'_{t-1} + \dots + \phi_p X'_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}. \quad (2.54)$$

En donde X' , expresa que sobre la serie original X_t se han aplicado d diferencias. En la estimación de los modelos $ARIMA$ el problema parte de identificar el modelo que mejor describa al fenómeno a predecir.

2.10.2. Test de Dickey-Fuller o de Raíz unitaria.

Definición 2.21. (Test de Dickey-Fuller) Es una prueba de raíz única que detecta estadísticamente la presencia de conducta tendencial estocástica en las series temporales de las variables mediante un contraste de hipótesis.

Es un test muy popular debido a su simpleza y facilidad con la que puede realizarse ya que se apoya en estimaciones por autorregresivos de mínimos cuadrados ordinario. A continuación se presentan algunas variables del test.

Test de Dickey-Fuller sin intercepto ni tendencia.

Es este caso se considera un modelo $AR(1)$ sin intercepto

$$X_t = \phi X_{t-1} + \epsilon_t.$$

En este modelo simple tendríamos una raíz unitaria si es que el valor ϕ fuera igual a uno.

A continuación se estima X_t en función de su valor rezagado y se realiza una prueba de hipótesis sobre si ϕ es igual a uno o no.

Es frecuente en este test hacer algunas transformaciones como la que sigue.

Se resta el valor X_{t-1} en ambos lados y se obtiene:

$$\begin{aligned} X_t - X_{t-1} &= \phi X_{t-1} - X_{t-1} + \epsilon_t \\ \nabla X_t &= (\phi - 1)X_{t-1} + \epsilon_t \\ \nabla X_t &= \alpha X_{t-1} + \epsilon_t. \end{aligned} \quad (2.55)$$

donde $\alpha = (\phi - 1)$ y ∇ es el operador de las primeras diferencias. Por lo tanto, en la práctica se calcula la Ecuación (2.55) y se prueba la hipótesis nula de que $\alpha = 0$,

entonces $\rho = 1$, es decir, tenemos una raíz unitaria, lo que significa que la serie de tiempo es no estacionaria.

Ahora $\hat{\Phi}$ es un estimador por Mínimos Cuadrados Ordinarios (*MCO*). EL estimador del error estándar de $\hat{\Phi}$ es

$$\hat{ES}(\hat{\Phi}) = S\left(\sum_2^n (\nabla X_{t-1} - \bar{X})^2\right)^{\frac{1}{2}}$$

donde $S^2 = \sum_{t=2}^n (\nabla X_t - \hat{\Phi} X_{t-1})^2 / (n - 3)$ y \bar{X} es la media de la serie.

2.10.3. Modelo *ARIMA*(0, 1, 1) o *IMA*(1, 1)

El modelo *ARIMA*(0,1,1) se presenta cuando se tiene $p = 0$, $d = 1$ y $q = 1$ y se expresa como sigue:

$$X_t = X_{t-1} + \epsilon_t - \theta\epsilon_{t-1} \quad (2.56)$$

donde $|\theta| < 1$. El proceso se reduce a un modelo *MA*(1) para la serie con primeras diferencias.

2.10.4. Varianza y autocovarianza del modelo *ARIMA*

Un proceso estacionario en la media no es necesariamente estacionario en la varianza y autocovarianza [6]. Sin embargo un proceso que no es estacionario en la media tampoco lo sera en su varianza y su covarianza. En particular, un fenómeno del modelo *ARIMA* es que a pesar de ser no estacionario, las características del proceso completo son determinadas siempre por un número finito de parámetros, por ejemplo si se consideran las variables X_1, X_2, \dots, X_n y se supone que se ajusta a un modelo *IMA*(1, 1)

$$X_t = X_{t-1} + \epsilon_t - \theta\epsilon_{t-1}.$$

de esta forma se obtienen n observaciones en la serie que se pueden escribir como sigue

$$\begin{aligned} X_t &= X_{t-1} + \epsilon_t - \theta\epsilon_{t-1} \\ &= X_{t-2} + \epsilon_t + (1 - \theta)\epsilon_{t-1} - \theta\epsilon_{t-2} \\ &\vdots \\ &= X_n + \epsilon_t + (1 - \theta)\epsilon_{t-1} + \dots + (1 - \theta)\epsilon_{n+1} - \theta\epsilon_n. \end{aligned} \quad (2.57)$$

De igual forma para $t - k > n$ Finalmente, con respecto al tiempo inicial n

$$\begin{aligned} Var(X_t) &= [1 + (t - n - 1)(1 - \theta)^2]\sigma^2 \\ Var(X_{t-k}) &= [1 + (t - k - n - 1)(1 - \theta)^2]\sigma^2 \\ Cov(X_{t-k}, X_t) &= [(1 - \theta) + (t - k - n - 1)(1 - \theta)^2]\sigma^2. \end{aligned} \quad (2.58)$$

Así la correlación es

$$\text{Corr}(X_{t-k}, X_t) = \frac{[(1-\theta) + (t-k-n-1)(1-\theta)^2]\sigma^2}{[1 + (t-n-1)(1-\theta)^2]\sigma^2 [[1 + (t-k-n-1)(1-\theta)^2]\sigma^2]} \quad (2.59)$$

De lo anterior se obtienen las siguientes observaciones

1. La varianza $\text{Var}(X_t)$ de un modelo *ARIMA* es dependiente del tiempo.
2. $\text{Var}(X_t) \neq \text{Var}(X_t)$ para $k \neq 0$.
3. La varianza $\text{Var}(X_t)$ no tiene límite cuando $t \rightarrow \infty$.
4. Si t es grande con respecto a n , entonces la correlación $\simeq 1$.

2.11. Metodología Box-Jenkins

A lo largo del tiempo se han creado una serie de distintas metodologías que tratan de elegir al modelo que mejor represente a una serie de tiempo. En la actualidad la metodología mas usada y difundida es la que propusieron **G.E.P Box y G.M Jenkins** en el año de 1970 en la cual lograron un gran avance en la identificación, ajuste y verificación de modelos *ARIMA*. Este metodología se basa en tratar de determinar cual es el comportamiento del proceso a través del tiempo bajo cierto modelo probabilístico. A continuación se anuncian algunos puntos importantes acerca de la naturaleza de los datos en el modelo *ARIMA*.

Se destacan en seguida algunas consideraciones importantes acerca de la naturaleza de los datos en una modelación *ARIMA*.

1. Aunque dicha metodología trata tanto con datos discretos y continuos, sólo se puede aplicar a datos igualmente espaciados en el tiempo, en intervalos discretos de tiempo.
2. Para elaborar un modelo *ARIMA* se requiere una cantidad mínima de datos. Se sugieren que el mínimo sea de 50 observaciones. Un modelo *ARIMA* se puede aplicar a una serie con menor tamaño, realizando con mas cuidado su interpretación.
3. Los modelos *ARIMA* son especialmente útiles en el tratamiento de series que presentan patrones estacionales.
4. Se asume que las perturbaciones aleatorias presentes, son independientes entre sí. No existe correlación entre ellas, por lo tanto ningún patrón modelable.

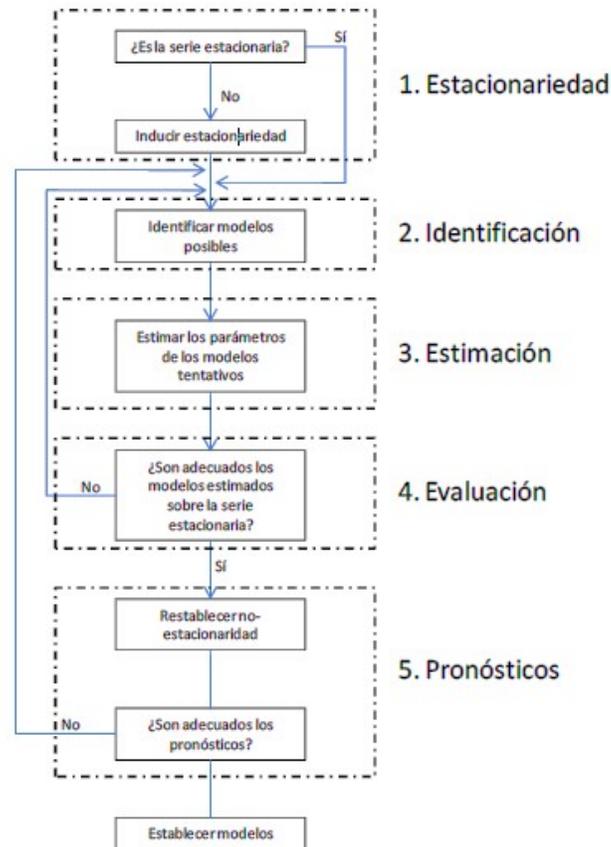


Figura 2.10: Metodología Box-Jenkins.

Dicho los puntos anteriores, se muestra en la figura 2.10 las etapas esenciales en la aplicación de esta metodología.

1. Estacionariedad.
2. Identificación.
3. Estimación.
4. Evaluación.
5. Pronóstico.

2.11.1. Estacionariedad

Anteriormente se explica que los procesos estocásticos se clasifican en estacionarios y no estacionarios. La estacionariedad está principalmente relacionada con la estabilidad de una serie. Un proceso estacionario se puede definir como una secuencia de datos que no presentan ningún cambio sistemático en la media ni en la varianza

así se dice que la serie es estable. De la definición mencionada anteriormente se puede decir que una serie de tiempo es estacionaria si su media y su varianza es esencialmente constantes a través del tiempo. Este primer concepto es de gran relevancia para la metodología Box-Jenkins ya que en caso de que los datos no se comporten de manera estable a lo largo del tiempo, es necesario aplicar una transformación a la serie para llevarla a ello. Existen dos formas de conocer si una serie es estacionaria: por medio del gráfico de la serie y por medio de la función de autocorrelación simple.

1. Inspección gráfica.

A través de una gráfica de la serie de tiempo, si se observa que los valores de la serie fluctúan respecto de una media constante, entonces es razonable pensar que la serie es estacionaria, si no ocurre lo anterior se llega a la conclusión de que la serie no es estacionaria. De manera visual se puede detallar si la serie es estacionaria si se detectan elevaciones o inclinaciones en las observaciones. Cualquier patrón de este tipo expresa que la serie es inestable.

2. Funciones de autocorrelación.

Cuando no es posible de manera visual determinar si la serie es estacionaria, usualmente se recurre a la función de autocorrelación simple, en especial en aquellas series con tendencia poco remarcada. El reconocimiento de la estabilidad se logra a partir de la variedad de comportamientos que esta función puede mostrar [8].

La función de autocorrelación simple se puede cortar ó truncar. Es decir que existe una espiga en el desfase k en la función de autocorrelación simple, si r_k es estadísticamente grande. Se considera que existe una espiga en la función si el valor absoluto de

$$t_{rk} = \frac{r_k}{s_{rk}}$$

es mayor que 2. La función de autocorrelación se trunca después del desfase k si no hay espigas en los desfases mayores que k en la función.

2.11.2. Identificación.

Una vez asegurado que la serie es estable a través del tiempo, el siguiente paso es la identificación del modelo probable que rige el proceso de serie de tiempo. Algunas ideas de esta fase son las siguientes:

1. La serie de tiempo que se encuentre en proceso de estudio cuenta con sus respectivas funciones de autocorrelación simple y parcial (FAS y FAP).
2. Cada una de las distintas formas *ARMA* posee se FAS y FAP teóricas asociadas al modelo.

3. Si la FAS y FAP calculadas de la serie a la que deseamos ajustar al modelo se asemejan a alguna o varias FAS y FAP teóricas, se puede concluir que el modelo *ARMA* es un modelo tentativo para la serie.

Se dice entonces que la identificación del modelo probable se realiza por medio de la comparación de las funciones de autocorrelación calculadas contra las teóricas.

2.11.3. Estimación.

Como su nombre lo dice, en esta etapa se estiman los coeficientes del modelo escogido tentativamente en el paso anterior, existen distintos criterios para dicho propósito. El principal criterio para el cálculo de los parámetros es el llamado **Estimación de Máxima Verosimilitud** ya que se ha demostrado que los parámetros estimados a través de este criterio reflejan con gran exactitud las características presentes en los datos de series de tiempo.

2.11.4. Evaluación del modelo

Una vez estimado los coeficientes del modelo propuesto, la siguiente etapa es la evaluación o verificación del mismo. En este paso se comprueba la eficiencia del modelo y se decide si es estadísticamente adecuado. Un modelo estadísticamente adecuado es aquel cuyos residuales son independientes entre si. Es decir, si los residuales son completamente aleatorios.

Para comprender mejor esta etapa, recordemos algunas ideas relevantes:

1. La idea básica de la metodología es modelar aquellos datos que estén correlacionados entre sí, mediante la combinación de términos *AR* y *MA*.
2. Los fenómenos reales siempre presentan perturbaciones aleatorias.
3. La metodología Box-Jenkins asume que las perturbaciones aleatorias pertenecientes a un proceso X_t son independientes entre si.

En la práctica las perturbaciones aleatorias no se pueden identificar directamente en una serie de tiempo. Sin embargo, los residuales del modelo *ARIMA* nos proporcionan un cálculo aproximado de las perturbaciones reales.

2.11.5. Pronóstico

La última fase de la metodología Box-Jenkins es pronosticar valores futuros de la serie de tiempo. El pronóstico de una serie de tiempo se hace por medio de predicciones o estimaciones puntuales. Así, una predicción puntual (\hat{X}_t) se define como el valor de la variable X en el tiempo t calculado por medio del modelo *ARIMA* ajustado en la serie.

2.12. Ejemplo

A continuación se muestra un ejemplo de la metodología Box-Jenkins en el análisis de datos de una serie de tiempo de los precios del petróleo maya que se muestran en la Tabla 2.4

Tabla 2.4: Observaciones del petróleo maya de Enero de 2013 a Diciembre de 2019.

Precios del petroleo Maya					
98.0	95.1	36.3	44.5	62.9	55.1
103.4	97.0	34.9	44.2	64.6	51.0
100.8	93.3	32.5	41.9	66.4	50.7
97.8	89.5	26.2	43.2	64.3	54.5
97.9	84.1	23.1	43.9	68.4	49.8
97.0	72.5	23.7	41.2	71.2	44.6
99.7	63.5	28.3	43.9	59.8	28.9
99.2	50.2	31.3	45.6	51.9	...

A través de una prueba de Box- Cox corregimos las varianzas desiguales creando una transformación con 0.5 y así obteniendo una nueva serie de observaciones que se muestran en la Tabla 2.5.

Tabla 2.5: Serie de Tiempo Transformada.

Precios del petróleo Maya					
17.8	17.5	10.0	11.3	13.9	12.8
18.3	17.7	9.8	11.3	14.1	12.3
18.1	17.3	9.4	11.0	14.3	12.2
17.8	16.9	8.2	11.1	14.0	12.8
17.8	16.3	7.6	11.2	14.5	12.1
17.7	15.0	7.7	10.8	14.9	11.4
18.0	13.9	8.6	11.2	13.5	8.8
17.9	12.2	9.2	11.5	12.4	
17.8	10.7	10.1	11.9	12.7	
17.4	11.6	10.6	12.0	13.1	
16.8	11.6	10.3	12.6	13.4	
16.9	12.0	10.3	12.7	13.8	
88.3	12.4	10.1	13.2	13.5	
90.8	12.5	10.8	13.0	13.1	
91.4	11.4	10.4	13.1	13.2	
93.9	10.4	11.0	13.3	12.1	

La Figura 2.11 se muestran las observaciones transformadas de la serie de tiempo de la Tabla 2.5.

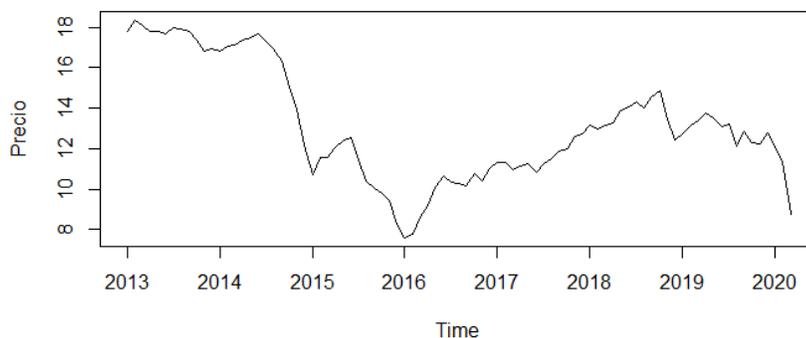


Figura 2.11: Gráfico Serie de Tiempo.

En la Figura 2.11 se aprecia claramente que esta serie no cuenta con una media y una varianza constante, lo cual se puede comprobar mediante una prueba de Dickey-Fuller, lo que lleva al siguiente juego de hipótesis.

$$H_0 : \gamma = 0 \quad \text{vs} \quad H_1 : \gamma < 0 \quad (2.60)$$

Con:

- H_0 : hipótesis nula es decir tiene raíz unitaria.
- H_1 hipótesis alternativa es decir no tiene raíz unitaria.

bajo el estadístico τ la región de rechazo es $RC = \{\tau_H : \tau_H \leq \tau_\alpha\}$, los resultados se muestran en la siguiente Tabla 2.6

Tabla 2.6: Prueba Dickey-Fuller

Para $\alpha = 0.01$ y $\tau_\alpha = -2.63$	
Prueba de hipótesis	
τ_H	-1.56
RC	$\tau_H \leq -2.63$
p-valor	0.7328
Conclusión	No se rechaza H_0

En la Tabla 2.6 se muestra que no se rechaza la hipótesis nula lo que indica que la serie no es estacionaria, por lo tanto se confirma que la serie no es estacionaria y bajo la metodología de Box-Jenkins se transforma a la serie para así volverla estacionaria, aplicando primero una diferencia a la serie.

Luego, al aplicar la transformación de una diferencia a las observaciones se obtiene la primer serie transformada cuyo gráfico se encuentra en la Figura 2.12

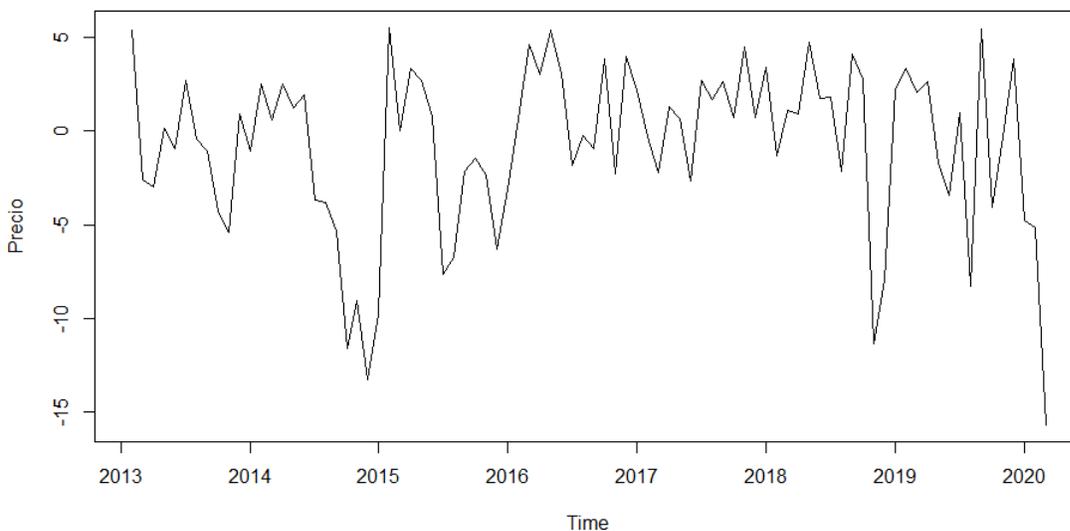


Figura 2.12: Gráfico de la Serie de Tiempo con una diferencia.

Ahora la serie parece tener una media y varianza casi constante pero se debe comprobar nuevamente a través del test de Dickey-Fuller, y el resultado se muestra en la tabla 2.7 Podemos observar que aún la serie no es estacionaria puesto que en

Tabla 2.7: Prueba Dickey-Fuller para observaciones con una diferencia.

Para $\alpha = 0.01$ y $\tau_\alpha = -2.63$	
Prueba de hipótesis	
τ_H	-3.14
RC	$\tau_H \leq -2.63$
p-valor	0.105
Conclusión	No se rechaza H_0

el cuadro anterior se muestra que el $p - value = 0.07162$ y que para que la serie sea estacionaria debe rechazarse la hipótesis nula del siguiente juego de hipótesis.

H_0 : No estacionaria (p-value > 0.05) vs H_1 : Es estacionaria (p-value < 0.05)

Por lo tanto se confirma que la serie no es estacionaria y bajo la metodología de Box-Jenkins se transforma a la serie para así volverla estacionaria, aplicando una segunda diferencia.

Luego, al aplicar la transformación de dos diferencia a las observaciones se obtiene la primer serie transformada cuyo gráfico se encuentra en la Figura 2.13

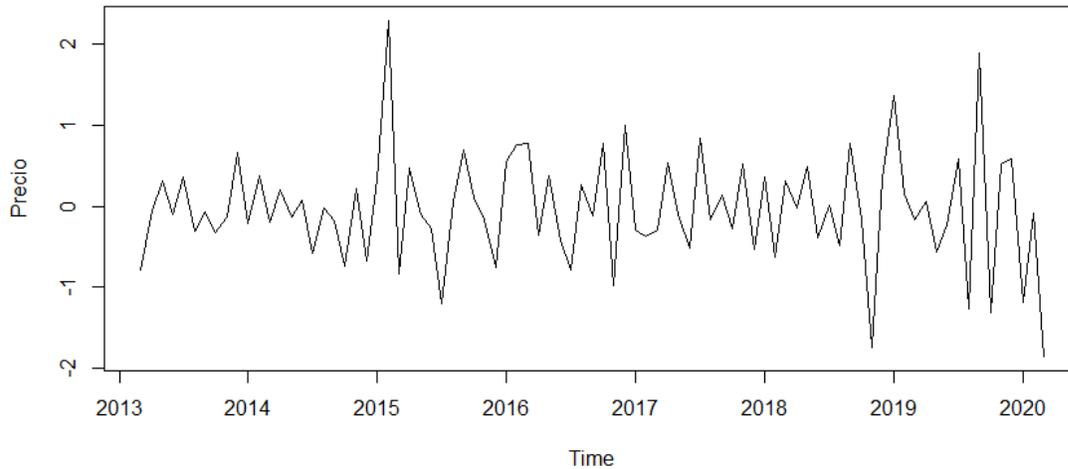


Figura 2.13: Gráfico Serie de Tiempo con dos diferencias.

Ahora la serie parece tener una media y una varianza mas constante pero nuevamente se debe comprobar a través del test de Dickey-Fuller, y el resultado se muestra en la tabla 2.8

Tabla 2.8: Prueba Dickey-Fuller para observaciones con dos diferencias.

Para $\alpha = 0.01$ y $\tau_\alpha = -2.63$	
Prueba de hipótesis	
τ_H	-5.1673
R.C	$\tau_H \leq -2.63$
p-valor	0.01
Conclusión	Se rechaza H_0

A través del cuadro anterior se muestra que la serie nos lleva al rechazo de la hipótesis nula, lo cual implica que la serie transformada puede considerarse como estacionaria. Además se analiza el comportamiento de su autocorrelograma y autocorrelograma parcial, en la Figura 2.14.

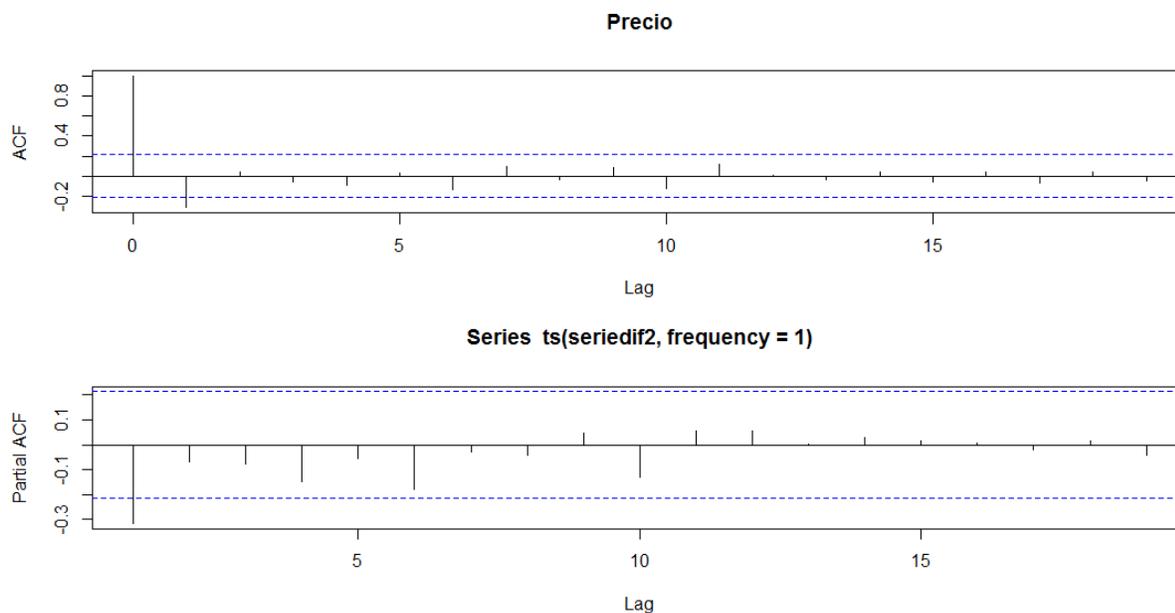


Figura 2.14: FACP de la serie con dos diferencias.

Notemos que el autocorrelograma parece tener un comportamiento de un modelo $MA(2)$. El autocorrelograma parcial tiene una espiga sobresaliente en el primer lapso por lo que podría corresponder a la $FACP$ de un modelo $AR(1)$. Se elige un modelo $ARIMA$ por tener dos diferencias y se piensa en un modelo combinado entre AR MA .

A continuación se prueban diferentes modelos para la serie, eligiendo al mejor mediante el Criterio de Akaike, el cual dice que el mejor modelo es aquel con el estadístico de menor tamaño y en el que se muestran en la Tabla 2.9.

Tabla 2.9: Modelos $ARIMA$ propuestos para obtener un pronóstico.

Modelo	AIC
ARIMA(1,2,1)	161.54
ARIMA(1,1,0)	158.32
ARIMA(2,2,2)	164.62
ARIMA(0,2,1)	167.46
ARIMA(1,2,2)	163.20
ARIMA(0,1,1)	161.71

A continuación se procede a la estimación de parámetros del modelo que es considerado como el mejor, en el ejemplo anterior se observa que el mejor modelo es $ARIMA(1, 1, 0)$ las estimaciones obtenidas mediante el Software R para series de tiempo se presentan en la Tabla 2.10

Tabla 2.10: Parámetros del modelo.

Parámetro	Valor	Error Estándar
$AR(1)$	0.4190	0.1103

forma

$$\begin{aligned} X_t &= c + \phi_1 X_{t-1} + \epsilon_t \\ &= 0.4190 X_{t-1} + \epsilon_t. \end{aligned} \quad (2.61)$$

Se muestra, los coeficientes para nuestros autorregresivos y medias móviles pero en este ejemplo no se cuenta con ninguna media móvil.

Por último se verifica que el modelo propuesto sea adecuado, por lo que se procede a analizar el comportamiento de los residuales, el cual determina si los residuales son aleatorios.

Para verificar se utiliza la prueba de Ljung-Box la cual contrasta el siguiente juego de hipótesis

$$H_0 : \text{El modelo es apropiado} \text{ vs } H_1 : \text{El modelo es inapropiado}$$

Se presenta el diagnostico del modelo propuesto en la Figura 2.15.

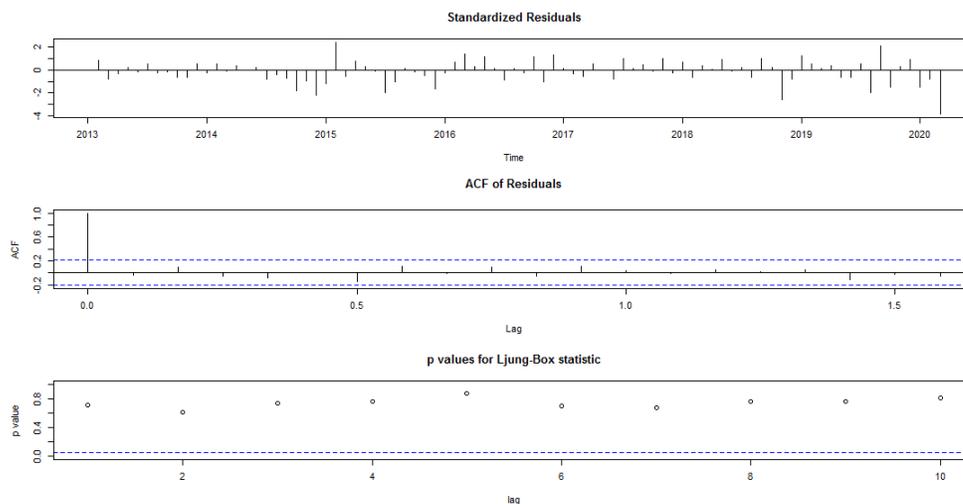


Figura 2.15: Diagnostico del modelo propuesto.

Recordemos que los valores estandarizados que se muestran en la parte superior de la Figura 2.16 deben parecerse a un ruido blanco que mas adelante se comprobaran con la prueba de ruido blanco, pero más aún se pueden observar los p -value en la parte inferior que es un estadístico que nos ayuda observar si existe un ruido blanco. Visualmente se puede observar que los p -values son mayores a 0.05 pero se comprueba con la prueba de Ljung-Box que se muestra en la Tabla 2.12.

Tabla 2.11: Prueba L-Jung Box.

Prueba L-Jung Box	
Prueba de hipótesis	
x-squared	0.13253
df	1
p-valor	0.7158
Conclusión	No Se rechaza H_0

Por lo que no se rechaza la hipótesis nula es decir se acepta la suficiencia del modelo, en la Figura 2.16 se muestra la gráfica e histograma de los residuales en los que se puede notar que cuentan con una media y varianza constante.

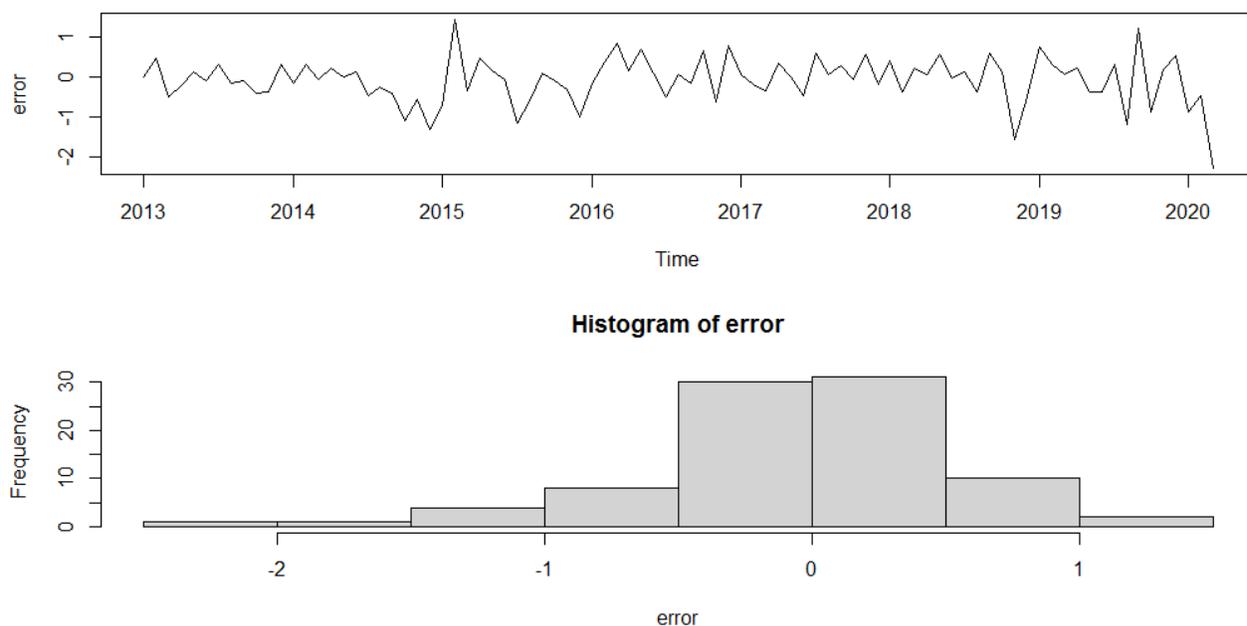


Figura 2.16: Gráficas de Residuales del Modelo.

Una vez que se ha comprobado la suficiencia se realiza el pronóstico para el siguiente año es decir se pronostican 9 valores. En la Figura 2.17 se muestra gráficamente el pronóstico, además del intervalo de confianza.

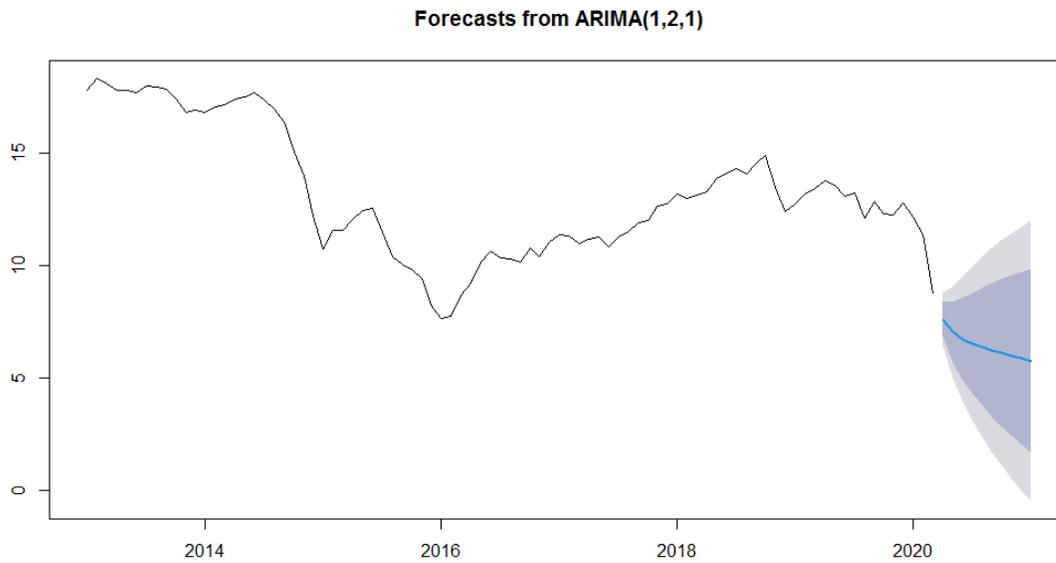


Figura 2.17: Pronóstico

La Tabla 2.12 indica el período e intervalo de confianza en el que se encuentra cada punto pronosticado

Tabla 2.12: Pronostico próximos 9 meses.

Fecha	Pronóstico	Lo 80	Hi 80	Lo 95	Hi 95
Apr 2020	7.586004	6.823559	8.348448	6.41994568	8.752062
May 2020	7.022253	5.689627	8.354878	4.98417783	9.060328
Jun 2020	6.714494	4.893504	8.535485	3.92952937	9.499459
Jul 2020	6.514825	4.270466	8.759185	3.08237422	9.947277
Aug 2020	6.360797	3.741618	8.979975	2.35510815	10.366485
Sep 2020	6.226039	3.268096	9.183982	1.70225512	10.749823
Oct 2020	6.099418	2.829904	9.368932	1.09912756	11.099709
Nov 2020	5.976233	2.416127	9.536339	0.53152145	11.420945
Dec 2020	5.854499	2.020379	9.688618	0.00928093	11.718278

Es así como se tiene al mejor modelo para pronosticar a los datos futuros, cabe señalar que efectivamente los datos fueron obtenidos de una simulación del modelo $ARIMA(1; 1; 0)$, comprobándose, al usar la metodología de Box-Jenkins, que efectivamente, se obtiene como mejor modelo aquel de donde se obtuvieron los datos por simulación.

Capítulo 3

Redes Neuronales Artificiales

A través del tiempo el ser humano se ha caracterizado por la búsqueda constante de nuevas vías para facilitar sus condiciones de vida. Todas estas búsquedas han servido para reducir significativamente el trabajo y más aun cuando se trata de tareas bastante complejas. Los progresos obtenidos han permitido dirigir estos esfuerzos a otros campos, como por ejemplo, a la construcción de máquinas calculadoras que ayuden a resolver de forma automática y rápida determinadas operaciones que resultan tediosas cuando se realizan a mano [16].

Estas máquinas permiten implementar fácilmente algoritmos para resolver multitud de problemas que antes resultaban difíciles de resolver. Sin embargo, se observa una limitación importante: ¿qué ocurre cuando el problema que se quiere resolver no admite un tratamiento algorítmico? Los desarrollos actuales de los científicos se dirigen al estudio de las capacidades humanas como una fuente de nuevas ideas para el diseño de las nuevas máquinas. Así, la inteligencia artificial es un intento por descubrir y describir aspectos de la inteligencia humana que pueden ser simulados mediante máquinas. Esta disciplina se ha desarrollado fuertemente en los últimos años teniendo aplicación en algunos campos como visión artificial, demostración de teoremas, procesamiento de información expresada mediante lenguajes humanos... etc [16].

El ámbito de las redes neuronales, es complejo y amplio. Durante los últimos años, el interés y la aplicación de este tipo de modelos han experimentado tal expansión que se ha convertido en una disciplina por sí misma. Las redes neuronales no son otra cosa más que emular una serie de características propias de los humanos, como memorizar y asociar hechos. En este capítulo se presenta de manera sencilla algunos conceptos básicos para el entendimiento de Redes Neuronales Artificiales y como éstas se pueden aplicar en nuestro caso de estudio [12], [13], [14], [15], [16].

3.1. Conceptos Básicos de Redes Neuronales

Existen una gran diversidad de formas de definir a las redes neuronales, desde definiciones simples hasta complejas en donde se trata de explicar su manejo, que consiste en:

1. Una nueva forma de computación, inspirada en modelos biológicos.
2. Un modelo matemático compuesto por un gran número de elementos procesales organizados en niveles.
3. Sistema de computación compuesto por un gran número de elementos simples, elementos de proceso interconectados, los cuales procesan información por medio de su estado dinámico como respuesta a entradas externas.

Definición 3.1. (Red Neuronal). Las redes neuronales son modelos creados al ordenar operaciones matemáticas siguiendo una determinada estructura. La forma más común de representar su estructura es mediante el uso de capas (*layers*), formadas a su vez por neuronas (*unidades, units o neurons*). Cada neurona, realiza una operación sencilla y está conectada a las neuronas de la capa anterior y de la capa siguiente mediante pesos, con el objetivo de regular la información que se propaga de una neurona a otra [12], [13],[14].

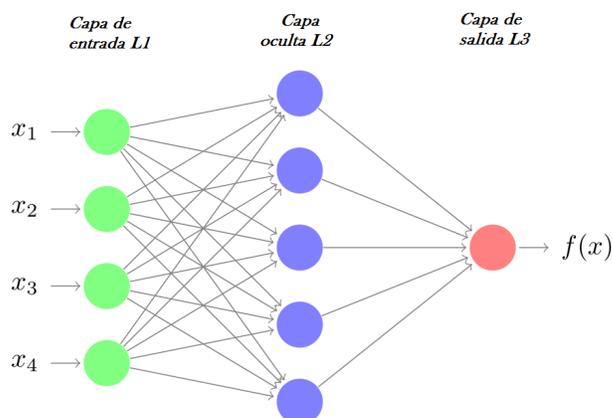


Figura 3.1: Representación de una red neuronal.

En la Figura 3.1 se muestra una red neuronal en donde la primera capa, se conoce como una capa de entrada (*input layer*) y recibe los datos en bruto, es decir, el valor de los predictores. La capa intermedia conocida como capa oculta (*hidden layer*), recibe los valores de la capa de entrada, ponderado por los pesos (flechas grises). La última capa (*output layer*), combina los valores que salen de la capa intermedia para generar la predicción.

Es muy fácil comprender la estructura de las redes a través de una representación de una red que es equivalente a un modelo de regresión lineal.

$$y = w_1x_1 + \dots + w_dx_d + b \quad (3.1)$$

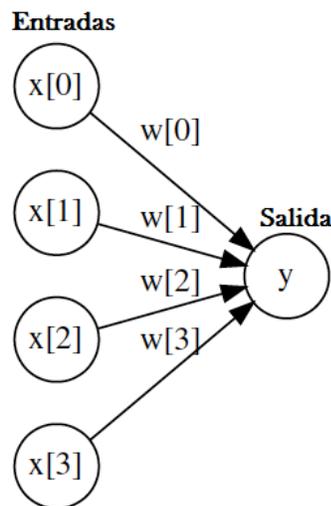


Figura 3.2: Representación de una red neuronal que es equivalente a un modelo lineal.

Cada neurona de la capa de entrada representa el valor de uno de los predictores. Las flechas representan los coeficientes de regresión, que en términos de redes se llaman pesos, y la neurona de salida representa el valor predicho. Para que esta representación equivalga a la ecuación de un modelo lineal, faltan dos cosas:

1. El sesgo del modelo.
2. Las operaciones de multiplicación y suma que combinan el valor de los predictores con los pesos del modelo.

La capa intermedia de una red tiene un valor de sesgo, pero suele omitirse en las representaciones gráficas. En cuanto a las operaciones matemáticas, es el elemento clave que ocurre dentro de las neuronas y conviene verlo con detalle [14].

3.1.1. Ventajas de la red neuronal

Las redes neuronales presentan un gran número de características semejantes al cerebro, un ejemplo claro es la capacidad de aprender de la experiencia, de generalizar cosas anteriores a nuevos casos, de obtener características esenciales a partir de entradas que representan información irrelevante.

Las redes neuronales permiten organizar por sí mismas lo aprendido, mientras que el aprendizaje es la modificación de cada elemento procesal, la *autoorganización* consiste en la modificación de la red neuronal completa para llevar a cabo un objetivo específico.

Además las redes neuronales tienen la capacidad de ser tolerantes a fallos. Fueron los primeros métodos computacionales con esta capacidad inherente. De este modo, si se produce un fallo en un número no muy grande de neuronas, aunque el comportamiento del sistema se ve influenciado, no sufre una caída repentina.

Una de las prioridades principales de las redes neuronales es que son capaces de operar en tiempo real. Las redes neuronales trabajan mediante conexiones en paralelo, lo que permite grandes velocidades de transmisión y respuesta casi al instante.

Por último, hay que destacar como ventaja la facilidad de inserción en la tecnología existente. De manera que, por ejemplo, resulta muy sencillo conseguir chips especializados que mejoren su capacidad.

En general, las redes neuronales tienen la capacidad de aprender a realizar tareas basadas en un entrenamiento inicial *Aprendizaje Adaptativo*. De esta forma, puede aprender a llevar a cabo ciertas tareas mediante el entrenamiento [16].

3.1.2. La neurona (Unidad)

La neurona es la principal unidad funcional dentro de estos modelos de redes. Dentro de cada neurona, ocurren la suma ponderada de sus entradas y la aplicación de una función de activación.

$$XW^t = (x_1, x_2, \dots, x_n) \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \sum_{i=1} x_i w_i. \quad (3.2)$$

En la primera parte, se multiplica cada valor de entrada x_i por su peso asociado w_i como se muestra en la Ecuación (3.2) y se suman junto con el sesgo. A este valor se le conoce como el valor neto de entrada a la neurona. Luego, este valor se pasa por una función, conocida como función de activación Ecuación (3.3), transforma el valor neto de entrada en un valor de salida.

$$\phi\left(\sum_{i=1} x_i w_i\right). \quad (3.3)$$

Si bien el valor que llega a la neurona, multiplicación de los pesos por las entradas, siempre es una combinación lineal, gracias a la función de activación, se pueden generar salidas muy diversas que se propagan, dicho valor puede ser la nueva entrada de una

neurona, formando así las redes neuronales, bien puede ser el resultado final o una nueva variable respuesta. Las respuestas obtenidas pueden ser una variable continua, una respuesta binaria o puede ser una respuesta categórica [14].

En la figura 3.4 se observa el comportamiento de una red neuronal artificial.

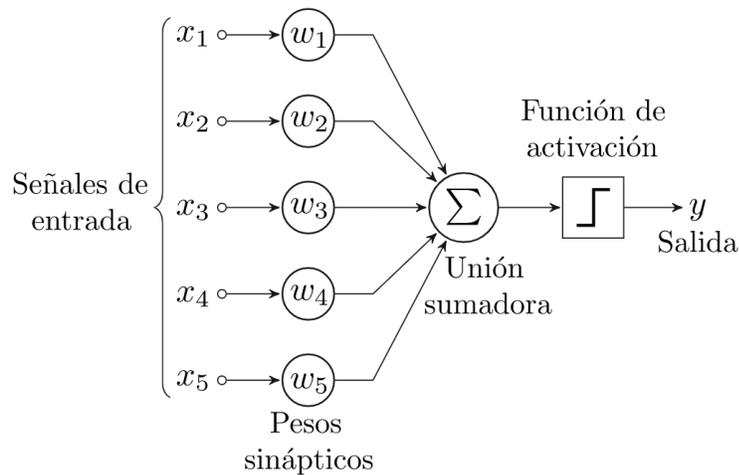


Figura 3.3: Comportamiento de una neurona artificial.

3.2. Función de Activación

Las funciones de activación controlan que información se propaga desde una capa a la siguiente (*forward propagation*).[13]. Estas funciones convierten el valor neto de entrada a la combinación neuronal de los entradas (*inputs*), pesos y sesgos, en un nuevo valor.

Es gracias a combinar funciones de activación no lineales con múltiples capas que los modelos de redes son capaces de aprender relaciones no lineales.

La gran mayoría de funciones de activación convierten el valor de entrada neto de la neurona en un valor dentro del rango $(0, 1)$ o $(-1, 1)$. Cuando el valor de activación de una neurona (salida de su función de activación) es cero, se dice que la neurona está inactiva, ya que no pasa ningún tipo de información a las siguientes neuronas. A continuación, se describen las funciones de activación más empleadas.

3.2.1. Función Lineal

Esta función también conocida como función identidad permite que la entrada sea igual a la salida, por lo que, aplicada a una red neuronal de varias capas se dice que es una regresión lineal. Así la función de activación lineal es,

$$f(x) = x, \quad \text{para } x \in [-\infty, \infty]. \quad (3.4)$$

Por lo tanto, esta función de activación lineal se usa si se requiere una regresión lineal como salida y de esta manera a la red neuronal que se le aplica tal función va a generar un valor único.

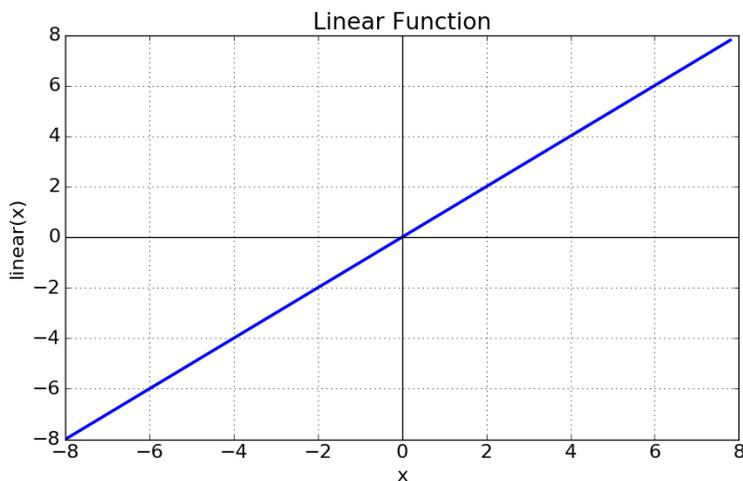


Figura 3.4: Representación función lineal.

3.2.2. Función Escalón

La función escalón, también conocida como función Umbral, indica que si x es menor que cero la neurona va a ser cero pero cuando es mayor igual a cero dará como resultado uno, y puede estar definida sobre $[0, 1]$ y $[-1, 1]$. Así la función escalón está dada por,

$$f(x) = \begin{cases} 1, & \text{si } x \geq 0 \text{ para } x \in [0, 1] \\ 0, & \text{si } x < 0 \text{ para } x \in [0, 1] \end{cases} \quad (3.5)$$

o bien

$$f(x) = \begin{cases} 1, & \text{si } x \geq 0 \text{ para } x \in [-1, 1] \\ -1, & \text{si } x < 0 \text{ para } x \in [-1, 1] \end{cases}$$

En la Figura 3.5 se muestra la representación de la función escalón,

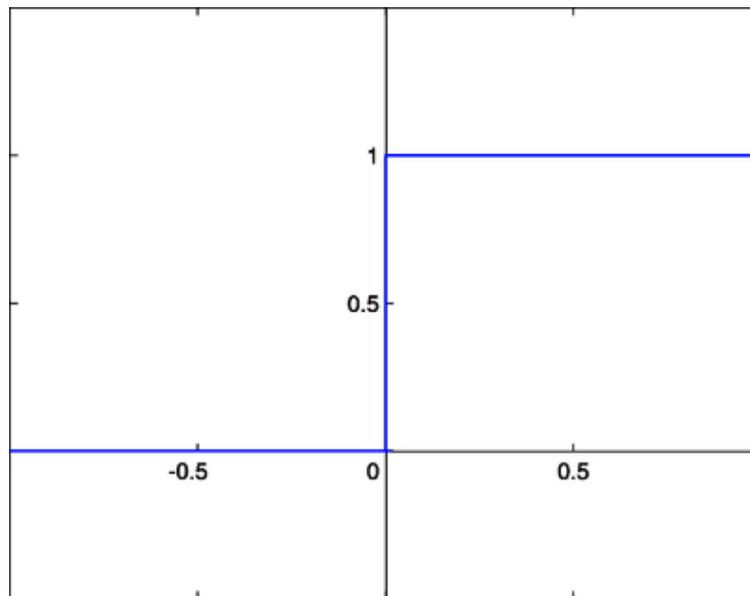


Figura 3.5: Representación función escalón.

3.2.3. Función Rectificadora Lineal Unificada (ReLU)

La función de activación ReLU aplica una transformación no lineal muy simple, activa la neurona solo si la entrada está por encima de cero. Mientras el valor de entrada está por debajo de cero, el valor de salida es cero, pero cuando es superior de cero, el valor de salida aumenta de forma lineal con el de entrada.

$$f(x) = \begin{cases} 0, & \text{si } x < 0 \\ x, & \text{si } 0 \leq x \leq 1 \\ 1, & \text{si } x > 1. \end{cases} \quad (3.6)$$

De esta forma, la función de activación retiene únicamente los valores positivos y descarta los valores negativos dándoles una activación de cero.

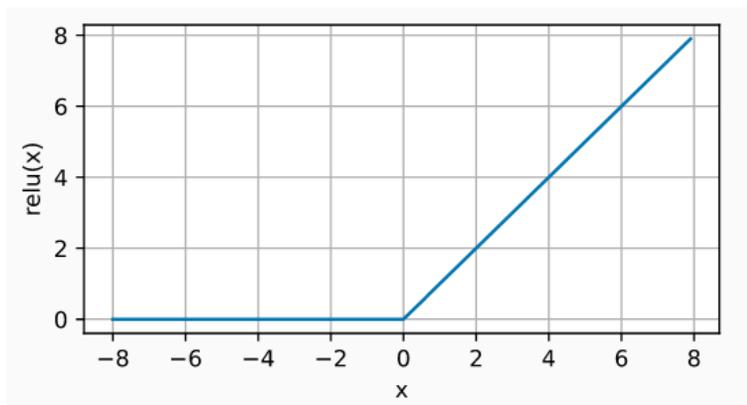


Figura 3.6: Representación función ReLU.

ReLU es con mucho la función de activación más empleada, por sus buenos resultados en aplicaciones diversas. La razón de esto reside en el comportamiento de su derivada (gradiente), que es cero o constante, evitando así un problema conocido como gradientes que desvanecen (*vanishing gradients*) que limita la capacidad de aprendizaje de los modelos de redes.

3.2.4. Función sigmoide

Esta función también conocida como función logística, está en un rango de valores de salida entre cero y uno por lo que la salida es interpretada como una probabilidad. Si se evalúa la función con valores de entrada negativos de valor absoluto alto, es decir si $x \leq 0$ la función será igual a cero, si se evalúa en cero la función dará 0.5 y en valores altos su valor es aproximadamente a 1. Por lo que esta función se usa en la última capa y se usa para clasificar datos en dos categorías. Actualmente la sigmoide no es una función muy utilizada debido a que no está centrada y esto afecta en el aprendizaje y entrenamiento de la neurona por lo que influye con el problema de desaparición de gradiente. Así, la función sigmoide está dada por,

$$f(x) = \frac{1}{1+e^{-x}}, \text{ para } x \in [0, 1]. \quad (3.7)$$

Aunque la función de activación sigmoide se utilizó mucho en los inicios de los modelos de redes, en la actualidad, suele preferirse la función ReLU.

Un caso en el que la función de activación sigmoide sigue siendo la función utilizada por defecto es en las neuronas de la capa de salida de los modelos de clasificación binaria, ya que su salida puede interpretarse como probabilidades.

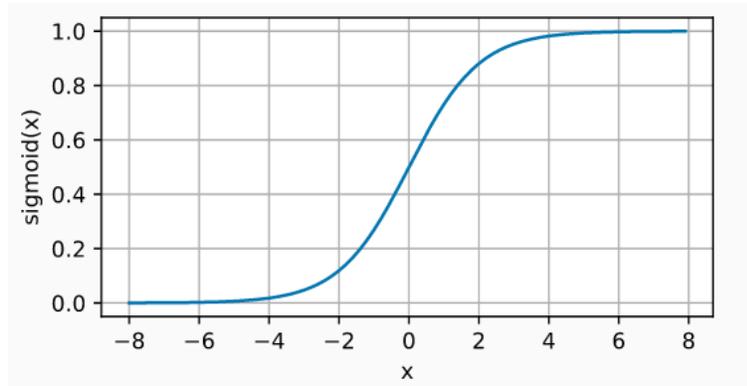


Figura 3.7: Representación función Sigmoide.

Tangente Hiperbólica (Tanh)

La función de activación Tanh, se comporta de forma similar a la función sigmoide, pero su salida está acotada en el rango $(-1, 1)$. Así la función tangente hiperbólica está dada por,

$$f(x) = \tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}. \quad (3.8)$$

Esta función de activación llamada tangente hiperbólica tiene un rango de valores de salida entre -1 y 1. Se dice que esta función es un escalamiento de la función logística, por lo que a pesar que esta función está centrada tiene un problema similar a la sigmoide debido al problema de desaparición del gradiente, que se da cuando en el entrenamiento se genera un error con el algoritmo de propagación hacia atrás y debido a esto el error se va propagando entre las capas, por lo que en cada iteración toma un valor pequeño y la red no puede obtener un buen aprendizaje.

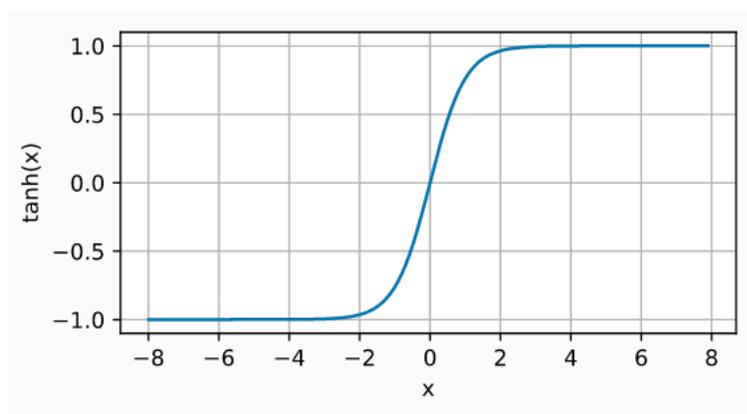
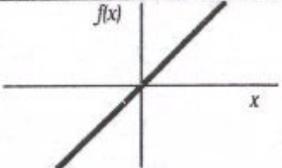
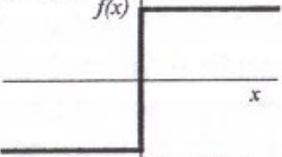
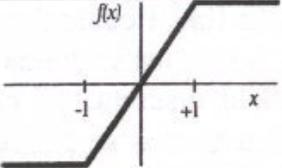
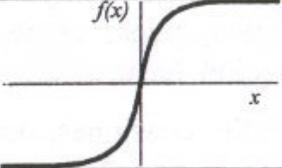
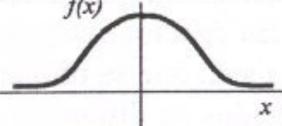
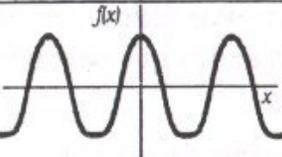


Figura 3.8: Representación función Tangente Hiperbólica

En la Figura 3.1 se muestra algunas de las funciones de activación más utilizados en los distintos modelos de redes neuronales.

Tabla 3.1: Tipos de funciones de activación [12].

	Función	Rango	Gráfica
Identidad	$y = x$	$[-\infty, +\infty]$	
Escalón	$y = \text{sign}(x)$ $y = H(x)$	$\{-1, +1\}$ $\{0, +1\}$	
Lineal a tramos	$y = \begin{cases} -1, & \text{si } x < -l \\ x, & \text{si } -l \leq x \leq +l \\ +1, & \text{si } x > +l \end{cases}$	$[-1, +1]$	
Sigmoidea	$y = \frac{1}{1 + e^{-x}}$ $y = \text{tgh}(x)$	$[0, +1]$ $[-1, +1]$	
Gaussiana	$y = Ae^{-Bx^2}$	$[0, +1]$	
Sinusoidal	$y = A \text{sen}(\omega x + \varphi)$	$[-1, +1]$	

3.3. Función de costo (loss function)

La función de costo, (l), [12] también llamada función de pérdida, *loss function* o *cost function*, es la encargada de cuantificar la distancia entre el valor real y el valor predicho por la red, en palabras más simples, su función es medir el error de la red al momento de realizar predicciones. En la mayoría de los casos, la función de costo arroja valores positivos. Cuanto más próxima a cero es el valor de costo, mejor son las predicciones de la red, siendo cero cuando los valores de predicción son iguales a los valores reales.

Dicha función tiene la capacidad de poder calcularse con una sola observación o para un conjunto de datos, el segundo caso, es el más utilizado para dirigir el entrenamiento de los modelos.

Dependiendo del tipo de problema, regresión o clasificación, es necesario utilizar una función de costo u otra. En problemas de regresión, las más utilizadas son el error cuadrático medio y el error absoluto medio. En problemas de clasificación suele emplearse la función logaritmo de pérdida (*log loss*), también llamada pérdida logística (*logistic loss* o *cross-entropy loss*) que a continuación se presentan.

3.3.1. Error cuadrático medio

El error cuadrático medio es la función más utilizada en problemas de regresión, para una determinada observación i , el error cuadrático se calcula como la diferencia al cuadrado entre el valor predicho \hat{y} y el valor real y

$$l^{(i)}(w, b) = (\hat{y}^{(i)} - y^{(i)})^2. \quad (3.9)$$

Las funciones de costo comúnmente se denotan como $l(w, b)$ para hacer referencia a que su valor depende de los pesos y sesgo del modelo, ya que estos son los que determinan el valor de las predicciones de $y^{(i)}$.

Entre algunas características se tienen.

1. Penaliza los valores que son muy grandes.
2. No es fácilmente interpretable.
3. Funcionan muy bien para optimizar regresiones en general.

Para cuantificar el error que comete el modelo para todo un conjunto de datos, simplemente se promedia el error de todas las N observaciones.

$$L(w, b) = \frac{1}{n} \sum_{i=1}^n l^{(i)}(w, b) = \frac{1}{n} \sum_{i=1}^n (\hat{y}^{(i)} - y^{(i)})^2. \quad (3.10)$$

Cuando el modelo se entrena utilizando el error cuadrático medio como función de costo, está aprendiendo a predecir la media de la variable respuesta [12].

3.3.2. Error medio absoluto

El error medio absoluto consiste en promediar el error absoluto de las predicciones.

$$L(w, b) = \frac{1}{n} \sum_{i=1}^n |\hat{y}^{(i)} - y^{(i)}|. \quad (3.11)$$

El error medio absoluto es más robusto frente a los valores atípicos que el error cuadrático medio. Esto significa que el entrenamiento del modelo se ve menos influenciado por datos anómalos que pueda haber en el conjunto de entrenamiento. Cuando un modelo se entrena utilizando el error absoluto medio como función de costo, está aprendiendo a predecir la mediana de la variable respuesta.

Entre algunas de las características que tienen se encuentran.

1. Es más difícil la diferenciación y la convergencia.
2. Penaliza menos los valores grandes.
3. Es más fácil de interpretar.

3.3.3. Las funciones Log loss, logistic loss o cross-entropy loss

En problemas de clasificación, la capa de salida utiliza como función de activación la función *softmax* y a través de esta función, la red devuelve una serie de valores que pueden interpretarse como la probabilidad de que la observación predicha pertenezca a cada una de las posibles clases.

Cuando la clasificación es de tipo binaria, donde la variable respuesta es 1 o 0, y $P = \Pr(Y=1)$, la función de costo *log-likelihood* se define como

$$L_{\log}(y, p) = -\log \Pr(y|p) = -(y \log(p)) + (1 - y) \log(1 - p). \quad (3.12)$$

Para problemas de clasificación con más de dos clases, esta fórmula se generaliza como sigue.

$$L_{\log}(Y, P) = -\log \Pr(Y|P) = -\frac{1}{N} \sum_{i=0}^{N-1} \sum_{K=0}^{K-1} y_{i,k} \log p_{i,k}. \quad (3.13)$$

En ambos casos, minimizar esta función equivale a que la probabilidad predicha para la clase correcta tienda a 1, y a 0 en las demás clases.

Entre algunas de las características que tienen se encuentran.

1. Más difícil en su diferenciación y convergencia.
2. Escala univariante.
3. Simétrica.
4. Es mas fácil de interpretar.

3.4. Múltiples capas

Las redes multicapas [14] son aquellas que disponen de un conjunto de neuronas agrupadas en varios (2, 3, etc.) niveles o capas. En estos casos, una forma para distinguir la capa a la que pertenece una neurona, consistiría en fijarse en el origen de las señales que recibe a la entrada y el destino de la señal de salida. Normalmente, todas las neuronas de una capa reciben señales de entrada desde otra capa anterior (la cual está más cerca a la entrada de la red), y envían señales de salida a una capa posterior (que está más cerca a la salida de la red). A estas conexiones se las denomina conexiones hacia adelante o *feedforward*.

Sin embargo, en un gran número de estas redes también existe la posibilidad de conectar la salida de las neuronas de capas posteriores a la entrada de capas anteriores; a estas conexiones se las denomina conexiones hacia atrás o *feedback*.

Estas dos posibilidades permiten distinguir entre dos tipos de redes con múltiples capas: las redes con conexiones hacia adelante o redes *feedforward*, y las redes que disponen de conexiones tanto hacia adelante como hacia atrás o redes *feedforward/feedback*.

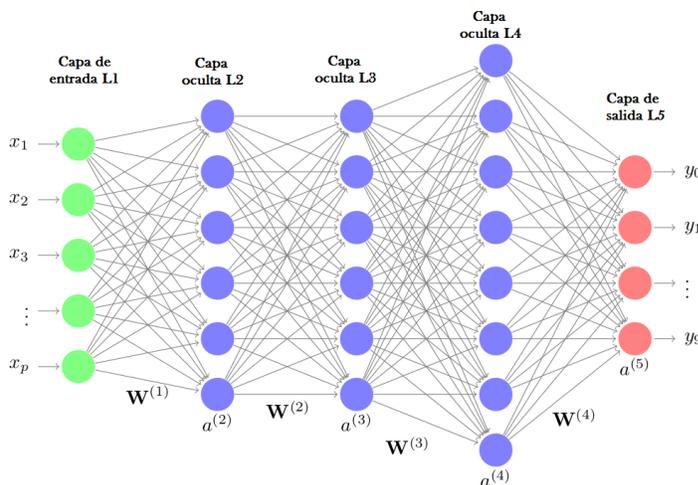


Figura 3.9: Representación de una red neuronal feed-forward (multi-layer perceptron).

La estructura de un Perceptrón Multi-Capa Figura 3.9 consta de varias capas de neuronas ocultas. cada neurona está conectada a todas las neuronas de la capa anterior y a las de la capa posterior. Aunque no es estrictamente necesario, todas las neuronas que forman parte de una misma capa suelen emplear la misma función de activación [14].

Combinando múltiples capas ocultas y funciones de activación no lineales, los modelos de redes pueden aprender prácticamente cualquier patrón. De hecho, está demostrado que, con suficientes neuronas, un Perceptron Multi-Capa (*MLP*) es un aproximador universal para cualquier función.

3.5. Entrenamiento

El proceso de entrenamiento de una red neuronal consiste en ajustar el valor de los pesos y sesgo de tal forma que, las predicciones que se generen, tengan el menor error posible. Gracias a esto, el modelo es capaz de identificar qué predictores tienen mayor influencia y de qué forma están relacionados entre ellos y con la variable respuesta.

La idea intuitiva de cómo entrenar una red neuronal es la siguiente:

1. Iniciar la red con valores aleatorios de los pesos y sesgo.
2. Para cada observación de entrenamiento (X, y) , calcular el error que comete la red al hacer sus predicciones. Promediar los errores de todas las observaciones.
3. Identificar la responsabilidad que ha tenido cada peso y sesgo en el error de la predicción.
4. Modificar ligeramente los pesos y sesgo de la red (de forma proporcional a su responsabilidad en el error) en la dirección correcta para que se reduzca el error.
5. Repetir los pasos 2, 3, 4 y 5 hasta que la red sea suficientemente buena.

Si bien la idea parece sencilla, alcanzar una forma de implementarla ha requerido la combinación de múltiples métodos matemáticos, en concreto, el algoritmo de retropropagación (backpropagation).

3.5.1. Retropropagación (Backpropagation)

Es el algoritmo que permite cuantificar la influencia que tiene cada peso y sesgo de la red en sus predicciones. Para conseguirlo, se hace uso de la regla de la cadena para calcular el gradiente, que no es más que el vector formado por la derivadas parciales de una función.

Cuando de redes neuronales se habla, la derivada parcial del error respecto al parámetro (peso o sesgo) mide cuanta responsabilidad ha tenido ese parámetro en el error cometido. Con lo que se puede identificar qué pesos de la red hay que modificar para mejorarla.

El método de Backpropagation, puede ser formulado de acuerdo con la siguiente ecuación.

$$V^{t+1} = V^t + \eta_i \nabla \mathbb{E}[V^t]. \quad (3.14)$$

Donde t es la i -ésima iteración, η_i es la razón de aprendizaje ($0 < \eta \leq 1$). Por otro lado, al pasar por la iteración t a la $t + 1$ el algoritmo aplica la corrección.

$$\nabla V = V^{t+1} - V^t = \eta_i \nabla \mathbb{E}[V^t].$$

La cual busca la minimización del error cuadrático medio (ECM) cuya interpretación matemática se muestra en la Ecuación En la Ecuación 3.15

$$ECM = \frac{1}{2} \sum_{i=1}^N (t_i - d_i)^2. \quad (3.15)$$

Donde t_i es la salida real de la red para los i -ésimos ejemplos y d_i es la salida esperada.

El algoritmo de Backpropagation se resume de la siguiente manera:

1. Inicializar aleatoriamente con valores pequeños los pesos de la red neuronal. Generalmente entre -0.5 y 0.5 .
2. Aleatoriamente elegir un patrón de entrada $X^{(a)}$.
3. Propagar la señal hacia adelante a través de la red.
4. Calcular δ_i^L para calcular la capa de salida.

$$\delta_i^L = f'(net_i^L)(d_i^L - o_i^L).$$

Donde net_i^L y o_i^L representa la entrada y salida (respectivamente) de la i -ésima unidad en la l -ésima capa y f' es la derivada de la función de activación $f(\cdot)$.

Para el Perceptron Multi-Capa en la función de activación se tendrá,

$$\delta_i^L = [z_j^q(1 - z_j^q)](t_j^q - z_j^q).$$

5. Calcular las deltas (δ) para las capas previas por propagación de error hacia atrás.

$$\delta_m^l = f'(net_m^l) \sum_k^K u_{mk}^{l+1} \delta_k^{l+1}.$$

Donde los u_{mk} representa los parámetros libres de la red o pesos. Para una red de tres capas se obtendría de la siguiente manera:

$$\delta_m^l = [y_m^q(1 - y_m^q)] \sum_{j=1}^J u_{mj}^r \delta_j^L.$$

Donde $r = (1, \dots, R)$ es el número de iteraciones o repeticiones del algoritmo.

6. Actualizar los pesos usando

$$\begin{aligned}\nabla u_{ki}^l &= \eta \delta_i^l o_k^{l-1}. \\ u_{mj}^{r+1} &= u_{mj}^r + \delta_j^L y_n^q.\end{aligned}$$

Y para la capa oculta

$$w_{nm}^{r+1} = w_{nm}^r + \eta \delta_m^l x_n^q.$$

7. Regresar al paso 2 y repetir el siguiente patrón hasta alcanzar el mínimo de error (fijado a priori) o hasta alcanzar el número máximo de iteraciones [17].

3.6. Preprocesado

El preprocesado engloba todas aquellas transformaciones realizadas sobre los datos con el objetivo que puedan ser interpretados por el algoritmo de machine learning lo más eficientemente posible. Todo preprocesado de datos debe aprenderse con las observaciones de entrenamiento y luego aplicarse al conjunto de entrenamiento y al de prueba. Esto es muy importante para no violar la condición de que ninguna información procedente de las observaciones de prueba participe o influya en el ajuste del modelo.

Este principio debe aplicarse también si se emplea validación cruzada (ver más adelante). En tal caso, el preprocesado debe realizarse dentro de cada iteración de validación, para que las estimaciones que se hacen con cada partición de validación no contengan información del resto de particiones.

Aunque no es posible crear un único listado, a continuación se resumen algunos de los pasos de preprocesado que más se suelen necesitar.

3.6.1. Imputación de valores ausentes

La gran mayoría de algoritmos no aceptan observaciones incompletas, por lo que, cuando el conjunto de datos contienen valores ausentes las cuales se pueden:

1. Eliminar aquellas observaciones que estén incompletas.
2. Eliminar aquellas variables que contengan valores ausentes.
3. Tratar de estimar los valores ausentes empleando el resto de información disponible.

Las primeras dos opciones, aunque sencillas, suponen perder información. La eliminación de observaciones solo puede aplicarse cuando se dispone de muchas y el porcentaje de registros incompletos es muy bajo. En el caso de eliminar variables, el impacto dependerá de cuánta información aporten dichas variables al modelo. Cuando se emplea la imputación, es muy importante tener en cuenta el riesgo que se corre al introducir valores en predictores que tengan mucha influencia en el modelo [13], [14].

3.6.2. Exclusión de variables con varianza próxima a cero

No se deben incluir en el modelo predictores que contengan un único valor (cero-varianza) ya que no aportan información. Tampoco es conveniente incluir predictores que tengan una varianza próxima a cero, es decir, predictores que toman solo unos pocos valores, de los cuales, algunos aparecen con muy poca frecuencia. El problema con estos últimos es que pueden convertirse en predictores con varianza cero cuando se dividen las observaciones por validación cruzada o bootstrap.

En el caso de variables cualitativas, cabe recordar que requiere que se binaricen (one hot encoding o dummy) para poder entrenar los modelos. Una variable booleana sigue una distribución de Bernoulli, por lo que su varianza puede ser calculada como:

$$\text{Var}[X] = p(1 - p).$$

Si bien la eliminación de predictores no informativos podría considerarse un paso propio del proceso de selección de predictores, dado que consiste en un filtrado por varianza, tiene que realizarse antes de estandarizar los datos, ya que después, todos los predictores tienen varianza 1 [13], [14].

3.6.3. Estandarización y escalado de variables numéricas

Cuando los predictores son numéricos, la escala en la que se miden, así como la magnitud de su varianza pueden influir en gran medida en el modelo. Muchos algoritmos son sensibles a esto, de forma que, si no se igualan de alguna forma los predictores, aquellos que se midan en una escala mayor o que tengan más varianza dominarán el modelo aunque no sean los que más relación tienen con la variable respuesta. Existen principalmente dos estrategias principales:

1. Centrado

Consiste en restarle a cada valor la media del predictor al que pertenece.

El centrado se consigue restándole a cada valor la media de la columna en la que se encuentra, como resultado de esta transformación, todos los predictores pasan a tener una media de cero, es decir, los valores se centran en torno al origen.

2. Normalización

Consiste en transformar los datos de forma que todos los predictores estén aproximadamente en la misma escala. Hay dos formas de lograrlo:

- Normalización Z-score

Divide cada predictor entre su desviación típica después de haber sido centrado,

$$z = \frac{x - \mu}{\sigma},$$

de esta forma, los datos pasan a tener una distribución normal.

- Estandarización max-min

Transforma los datos de forma que estén dentro del rango.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}.$$

3.6.4. Binarización de las variables cualitativas

La binarización (one-hot-encoding) consiste en crear nuevas variables dummy con cada uno de los niveles de las variables cualitativas. Por ejemplo, una variable llamada color que contenga los niveles rojo, verde y azul, se convertirá en tres nuevas variables (color rojo, color verde, color azul), todas con el valor 0 excepto la que coincide con la observación, que toma el valor 1 [13], [14].

3.7. Número y tamaño de capas

La arquitectura de una red, el número de capas y el número de neuronas que forman parte de cada capa, determinan en gran medida la complejidad del modelo y con ello su potencial capacidad de aprendizaje.

Las capas de entrada y salida son sencillas de establecer. La capa de entrada tiene tantas neuronas como predictores y la capa de salida tiene una neurona en problemas de regresión y tantas como clases en problemas de clasificación. En la mayoría de implementaciones, estos valores se establecen automáticamente en función del conjunto de entrenamiento. El usuario suele especificar únicamente el número de capas intermedias (ocultas) y el tamaño de las mismas.

Cuantas más neuronas y capas, mayor la complejidad de las relaciones que puede aprender el modelo. Sin embargo, dado que en cada neurona está conectada por pesos al resto de neuronas de las capas adyacentes, el número de parámetros a aprender aumenta y con ello el tiempo de entrenamiento.

3.8. Tasa de aprendizaje

La tasa de aprendizaje es un hiperparámetro que controla cuánto cambiar el modelo en respuesta al error estimado cada vez que se actualizan los pesos del modelo. Elegir la tasa de aprendizaje es un desafío, ya que un valor demasiado pequeño puede resultar en un proceso de entrenamiento largo que podría atascarse, mientras que un valor demasiado grande puede resultar en el aprendizaje de un conjunto de pesos subóptimo demasiado rápido o en un proceso de entrenamiento inestable.

La tasa de aprendizaje puede ser el hiperparámetro más importante al configurar la red neuronal. Por lo tanto, es vital saber cómo investigar los efectos de la tasa de aprendizaje en el rendimiento del modelo y construir una intuición sobre la dinámica de la tasa de aprendizaje en el comportamiento del modelo [13], [14].

3.9. Tipos de optimizadores de la función de costo

El objetivo del entrenamiento de redes neuronales es minimizar la función de costo encontrando los pesos adecuados para las aristas de la red.

Como nuestro objetivo es minimizar la función de costo al encontrar el valor optimizado para los pesos. También debemos asegurarnos de que el algoritmo se generalice de forma adecuada. Esto ayudará a hacer una mejor predicción de los datos que no se vieron antes.

3.9.1. Gradiente descendente

El descenso estocástico o gradiente de descenso incremental, es una aproximación estocástica del gradiente descendente usado para minimizar una función objetivo que se escribe como una suma de funciones diferenciables. Este optimizador trata de encontrar mínimos o máximos por iteración.

Al igual que en la función del gradiente descendente, el gradiente indica la dirección en la que la función tiene a la razón de aumento más pronunciada aunque no indica hasta donde se debe avanzar en esa dirección. Como este optimizador no nos dice cuanto avanzar, existe la razón de aprendizaje que no determina la distancia que recorrerá en cada iteración en la dirección del gradiente. Como queremos encontrar el punto mínimo, necesitamos ir en la dirección opuesta a la pendiente. Se actualizan los parámetros en la dirección del gradiente negativo para minimizar la pérdida.

Sea \mathbf{x}^t un punto actual. Entonces el vector será

$$\mathbf{p}^t = -\nabla f(\mathbf{x}^t)$$

es una dirección de descenso por lo que punto \mathbf{x}^{t-1} obtenido mediante la fórmula de actualización se tiene:

$$\mathbf{x}^{t-1} = \mathbf{x}^t - \alpha \nabla f(\mathbf{x}^t)$$

donde α es el tamaño de paso y con un α suficientemente pequeña se obtiene que.

$$f(\mathbf{x}^t - \alpha \nabla f(\mathbf{x}^t)) < f(\mathbf{x}^t) \text{ si } \nabla f(\mathbf{x}^t) = 0.$$

Dado que, calcular el error del modelo para todas las observaciones de entrenamiento, en cada iteración, puede ser computacionalmente muy costoso, existe una alternativa al método de descenso de gradiente llamada gradiente estocástico (stochastic gradient descent), (SGD). Este método consiste en dividir el conjunto de entrenamiento en lotes (minibatch o batch) y actualizar los parámetros de la red con cada uno. De esta forma, en lugar de esperar a evaluar todas las observaciones para actualizar los parámetros, se pueden ir actualizando de forma progresiva. Una ronda completa de iteraciones sobre todos los batch se llama época.

3.9.2. Descenso acelerado de Nesterov (NAG)

El algoritmo NAG puede considerarse del tipo de dos pasos Predictor-Corrector. En el paso predictor se extrapola linealmente la trayectoria actual. Luego, en el punto predicho, se evalúa el gradiente y se hace la corrección de la trayectoria. Con ello se logra una aproximación de segundo orden de la trayectoria con un costo computacional similar al de Descenso de Gradiente con Momento (MGD). Al mantener el efecto de inercia pero reducir los sobrepasos, NAG también incrementa la razón de convergencia.

Las formulas de actualización del algoritmo NAG están dadas por:

1. Sea $\mathbf{p}^t = -\nabla f(\mathbf{x}^t)$ y $\tilde{\mathbf{x}}^t = \mathbf{x}^t - \alpha \mathbf{p}^{t-1}$ el punto predicho en primer orden.
2. Calcular el gradiente en el punto predicho: $\mathbf{g}^t = \nabla f(\tilde{\mathbf{x}}^t)$.
3. Calcular la dirección de descenso con $\mathbf{p}^t = \mathbf{g}^t + \eta \mathbf{p}^{t-1}$.
4. Actualizar en punto $\mathbf{x}^{t+1} = \mathbf{x}^t - \alpha \mathbf{p}^t$.

En la Figura 3.10 se muestra gráficamente la actualización mediante NAG

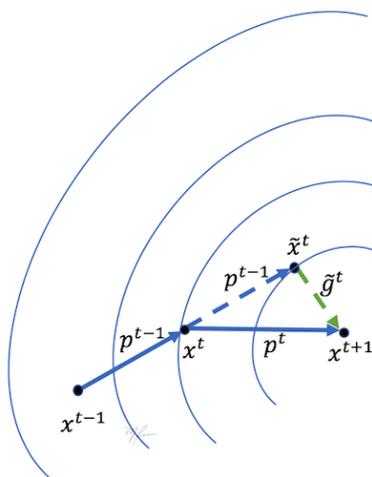


Figura 3.10: Representación del modelo NAG.

3.9.3. Descenso de Gradiente Adaptable (ADAGRAD)

Dado que la magnitud del gradiente es una cantidad a la que se le atribuyen todas las derivadas parciales, es posible que a pesar de tener una magnitud grande, existan coordenadas donde la función varía un poco.

Los algoritmos MGD como NAG se enfocan en estimar mejor la dirección de descenso. Sin embargo, el desempeño de dichos algoritmos se ve comprometido por la correcta selección del tamaño de paso α . Para ello, el algoritmo ADAGRAD busca

estimar el correcto tamaño de paso en cada iteración estimando un tamaño de paso para cada una de las variables en forma independiente.

Las formulas de actualización del algoritmo están dadas por:

1. Sea

$$g(\mathbf{x}_i^t) = \frac{\partial f(\mathbf{x}^t)}{\partial \mathbf{x}_i}.$$

La i -ésima derivada parcial de la función en el punto actual \mathbf{x}^t , esto es $g(\mathbf{x}_i^t) = [\nabla f(\mathbf{x}^t)]_i$.

2. Integrar los cuadrados de las parciales hasta la iteración actual:

$$G_i^t = \eta(g(\mathbf{x}_i^t))^2 + (1 - \eta)G_i^{t-1} \text{ con } \eta \text{ grande.}$$

3. Actualizar el punto, donde cada elemento tienen su propio caso.

$$\mathbf{x}_i^{t-1} = \mathbf{x}_i^t - \frac{\alpha}{\sqrt{G_i^t + \epsilon}} \mathbf{p}_i^t.$$

3.9.4. Momentun Adaptable (ADAM)

El algoritmo ADAM calcula la dirección de descenso usando momentum (similar a MGD) y utiliza una estrategia similar para calcular adaptar el tamaño de paso. Es decir, utiliza momentum para actualizar el paso, lo que evita cambios bruscos en el paso.

Esto lo hace muy estable para su uso en estrategia tipo Gradiente Estocástico (SGD) donde las muestras pueden provocar cambios grandes en la magnitud del gradiente, además calcula un paso global en vez de usar un paso para cada variable. También adecuado en estrategias de entrenamiento tipo estocásticas o por lotes, como en el caso de Redes Neuronales Profundas (Deep Learning). Una mejora importante es la corrección del sesgo en la estimación de los momentos.

Las formulas de actualización del algoritmo están dadas por.

Sea

$$g_t = \nabla f_t.$$

La i -ésima derivada parcial de la función en el punto actual \mathbf{x}^t .

1. Calcular la dirección de descenso con momentum

$$p_t = \eta_1 p_1 + (1 - \eta_1) g_t,$$

donde p_t conserva la escala de p_{t-1} y el gradiente g_t .

2. Actualizar, a la vez con momentum, el factor adaptativo del descenso \mathbf{v}^t

$$\mathbf{v}^t = \eta_2 \mathbf{v}^{t-1} + (1 - \eta_2) G_t,$$

donde $\bar{G} = [g_1^2, g_2^2, \dots, g_n^2]$ es el vector de elementos a cuadrado del gradiente.

3. Escalar la dirección de descenso y el momentum (Reducción del sesgo).

$$\hat{p}_t = \frac{1}{1 - (\eta_1)^t} p_t \quad \hat{v}_t = \frac{1}{1 - (\eta_2)^t} v_t.$$

4. Actualización del punto con la fórmula de paso adaptable

$$x_{t-1} = x_t = \frac{\alpha}{\sqrt{\hat{v}_t} + \epsilon} \hat{p}_t.$$

3.9.5. Importancia de la reducción del sesgo

El primer paso del algoritmo ADAM pretende mejorar el cálculo del gradiente, calculando, generalmente, sobre solo una muestra, promediándolo con los gradientes recientemente calculados:

$$\begin{aligned} p_t &= \eta_1 p_{t-1} + (1 - \eta_1) g_t \\ &= \eta_1^2 p_{t-2} + \eta_1 (1 - \eta_1) g_{t-1} + (1 - \eta_1) g_t \\ &= \eta_1^3 p_{t-3} + \eta_1^2 (1 - \eta_1) g_{t-2} + \eta_1 (1 - \eta_1) g_{t-1} + (1 - \eta_1) g_t \\ &\dots \end{aligned}$$

Sea $p_0 = 0$ como condición inicial de las iteraciones, se tiene.

$$p_t = (1 - \eta_1) \sum_{i=0}^t \eta_1^{t-i} g_i.$$

Tomando las esperanzas y usando la propiedad de $\mathbb{E}[xy] = \mathbb{E}[x]\mathbb{E}[y]$, si x y y son independientes.

$$\mathbb{E}[p_t] = \mathbb{E}[g_t] (1 - \eta_1) \sum_{i=0}^t \eta_1^{t-i}.$$

Utilizando una propiedad de la serie geométrica

$$\sum_{k=0}^{n-1} x^k = \frac{1 - x^n}{1 - x}.$$

Se tiene.

$$\hat{p}_t = \frac{p_t}{1 - \eta^t}.$$

3.10. Ejemplo

A continuación se muestra un ejemplo de una Red Neuronal. La siguiente serie registra el número de ventas diarias de un pequeño comercio de abarrotes. La base de datos van del 2 de Enero de 2017 al 30 de noviembre de 2018, un total de 604 observaciones.

Tabla 3.2: Ventas diarias [18].

Ventas diarias de una micro empresa.					
236	265	140	89	98	94
237	280	271	304	265	246
290	184	249	174	217	218
221	168	175	222	232	192
128	156	202	183	200	250
293	112	85	175	233	188
260	272	250	116	100	99
204	211	214	342	276	259
157	199	162	205	273	264
180	146	178	199	242	237
147	165	120	202	175	218
228	92	95	198	85	158
239	208	230	100	228	125
193	225	208	281	209	248
197	190	197	207	178	221
189	208	147	219	213	164
153	265	202	174	273	202
58	150	157	233	137	192
193	87	75	111	75	152
232	196	201	293	226	72
206	217	202	224	230	250
234	173	208	218	85	267
147	185	232	168	241	...
135	212	187	139	239	189

En la Figura 3.11 se muestra el gráfico de la base de datos y se puede observar que las observaciones no tienen tendencia y fluctúan alrededor de la media. También se puede ver cierta estacionalidad en el conjunto de datos que corresponden a cierto periodo.

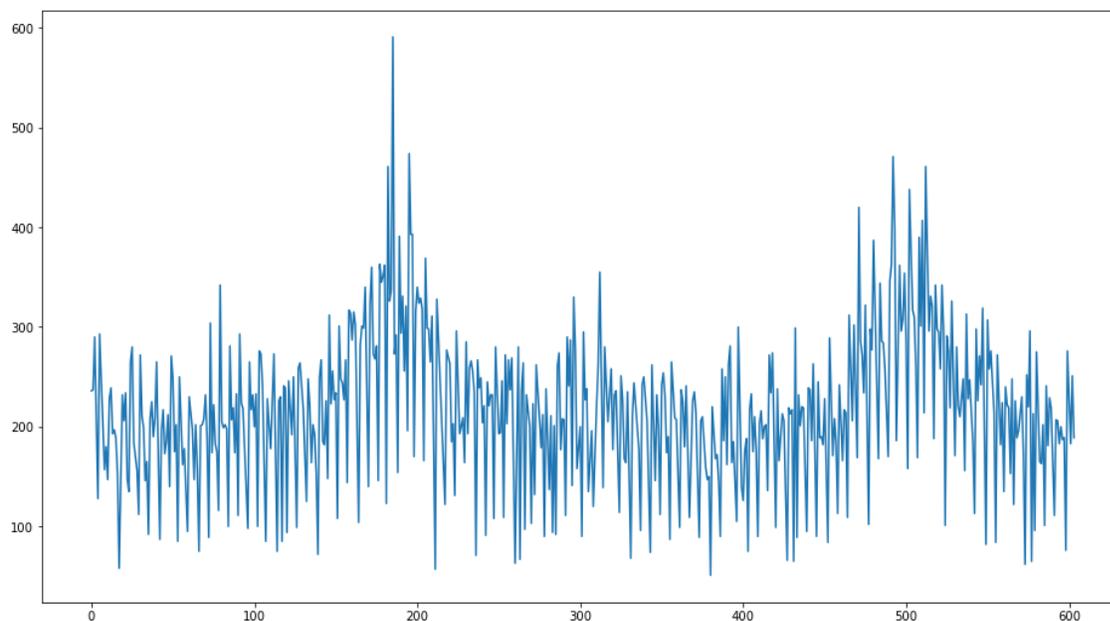


Figura 3.11: Gráfico de Datos

En la Tabla 3.3 se muestra la primer fecha y la última fecha de la base de datos, así como el número de datos de cada año.

Tabla 3.3: datos anuales.

Numero de Observaciones	
Año	Número de datos
2017	315
2018	289

El comercio cierra los domingos, por eso se puede observar que en el año 2017 el número de observaciones no es igual a 365 días como se podría presuponer. Y en el año 2018 el número de observaciones es menor ya que la base de datos no cuenta con las observaciones del mes de Diciembre. Ya que este será el mes que se pronosticará a lo largo de este ejemplo.

A continuación se muestran algunas gráficas de la base de datos pero antes en la Tabla 3.4 se muestra un resumen dando de algunas estadísticas de la base de datos.

Tabla 3.4: Estadísticas de ventas.

Estadísticas	
Numero de Observaciones	604
Media	215.93
Desviación Estándar	75.05
Mínimo	51
25 %	171
50 %	214
75 %	261.25
Máximo	591

Son 604 observaciones, la media de ventas es de 215 unidades diarias y un desvió de 75 que indica que por lo general los datos se encuentran entre 140 y 290 unidades.

Tabla 3.5: Medias mensuales de ventas.

Año	Medias mensuales de 2017	Medias mensuales de 2018
Enero	203.92	201.38
Febrero	184.66	190.62
Marzo	182.96	174.84
Abril	198.96	186.00
Mayo	201.18	190.66
Junio	209.51	196.03
Julio	278.92	289.50
Agosto	316.00	309.03
Septiembre	222.92	230.51
Octubre	207.85	209.44
Noviembre	185.92	184.48
Diciembre	213.20	212.15

En la Tabla anterior se puede observar que la variación entre cada año es pequeña por lo que se puede concluir que el comportamiento que tienen las ventas es estacionario.

En la Figura 3.12 se muestran las medias mensuales de los años 2017 y 2018.

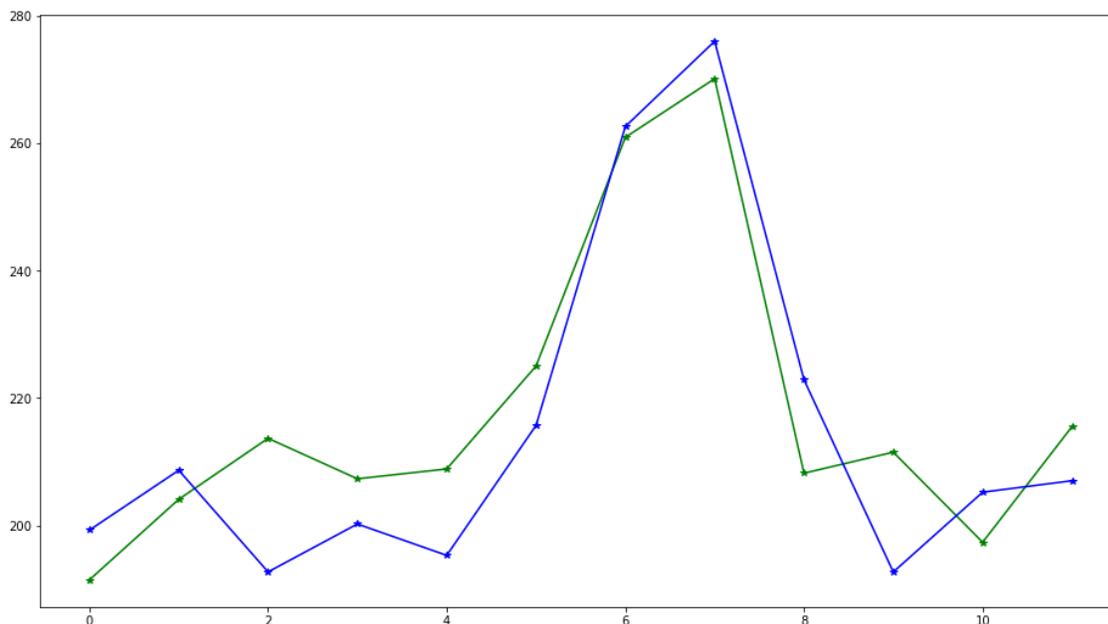


Figura 3.12: Gráfico de Medias Mensuales de los años 2017 y 2018.

Como se puede observar a inicios de año se tiene un descenso en la cantidad de unidades, luego comienza a subir hasta un poco más de medio año teniendo las ventas más altas en los meses de Junio y Julio para luego volver a disminuir y finalmente tener un pico en el mes de diciembre, por lo que podemos concluir que ambos años tiene un comportamiento semejante y por lo tanto la serie es estacionaria.

Una vez que se ha confirmado que la serie es estacionaria, se puede hacer pronóstico. Existen diversos métodos para hacer pronóstico. En nuestro caso, las ventas parecen comportarse bastante parecidas al año, con lo cual un método sencillo, si por ejemplo quisiéramos proveer el stock que necesitaría este comercio, sería decir “Si en 2017 en diciembre se vendió en promedio 213 unidades, pronóstico que en diciembre será similar”.

Para este ejemplo se ocupa una arquitectura sencilla de red neuronal FeedForward (también llamada MLP por sus siglas Multi-Layered Perceptron), con pocas neuronas y como método de activación tangente hiperbólica pues entregaremos valores transformados entre -1 y 1.

Para la preparación de los datos, lo que se aplica es la alteración del flujo de entrada de los datos de las unidades vendidas y se realiza una transformación para obtener una matriz, para así convertir nuestra base de datos en un problema de tipo supervisado para poder alimentar la red neuronal y poder entrenarla con *Backpropagation* para lograr esto se debe de tener entradas y salidas para el entrenamiento del modelo.

Se toman los 7 días previos para obtener el octavo. La red solo tendrá una salida por lo que queda como sigue:

- **Entradas:** Son 7 columnas que representan las ventas en unidades de los 7 días anteriores.
- **Salida:** El valor del octavo día. Es decir las unidades vendidas en ese día.

Luego se hace un escalamiento de la base de datos para transformar el rango de los valores de manera que queden entre -1 y 1 puesto que esto ayuda a la red para favorecer los cálculos.

Una vez transformados los datos se realiza la supervisión de dichos datos para poder explorar diferentes marcos de un problema de serie temporal con algoritmos de aprendizaje automático para investigar cuál podría dar como resultado modelos con mejor rendimiento, en la Tabla 3.6 se obtiene la matriz de entradas y salidas.

Tabla 3.6: Matriz de entradas y salidas.

	$(t - 7)$	$(t - 6)$	$(t - 5)$	$(t - 4)$	$(t - 3)$	$(t - 2)$	$(t - 1)$	(t)
7	-0.31	-0.31	-0.11	-0.37	-0.71	-0.10	-0.22	-0.43
8	-0.31	-0.11	-0.37	-0.71	-0.10	-0.22	-0.43	-0.60
9	-0.11	-0.37	-0.71	-0.10	-0.22	-0.43	-0.60	-0.52
10	-0.37	-0.71	-0.10	-0.22	-0.43	-0.60	-0.52	-0.64
11	-0.71	-0.10	-0.22	-0.44	-0.60	-0.52	-0.64	-0.34

Se utilizan las columnas de encabezados $(t-7)$ a $(t-1)$ y la salida es (t) que es la última columna.

Para poder crear la red neuronal, subdividiremos el conjunto de observaciones en datos de entrenamiento y prueba (*train and test*), algo de suma importancia de este procedimiento es que nos importa mantener el orden en el que alimentaremos a la red. Por lo que se realiza una subdivisión de los primeros 567 días consecutivos para entrenamiento de la red y los siguientes 30 para su validación.

Se realizó una transformación en un arreglo con forma $(567,1,7)$ esto significa que son 567 entradas con vectores de 1×7 .

La arquitectura de la red neuronal será:

- Entrada de 7 ventas diarias.
- Capa 1 oculta con 7 neuronas.
- La salida será 1 sola neurona.
- Función de activación tangente hiperbólica puesto que los valores se encuentran entre -1 y 1.
- Se utiliza el optimizador Adam.
- Para la métrica de pérdida se utiliza el Error Absoluto Medio.
- La predicción será un valor continuo.
- Para el cálculo de la precisión se utiliza el Error Cuadrático Medio.
- Finalmente para el entrenamiento utilizaremos un total de 40 iteraciones.

En la Tabla 3.7 se muestran los resultados de las 40 observaciones obtenidas después del entrenamiento de la red.

Tabla 3.7: Resultados del Entrenamiento.

Pérdida	0.1692
Error cuadrático medio	0.0551
Valor de Perdida	0.1383
Valor del error cuadrático medio	0.03

La visualización al conjunto de validación se muestra en la Figura 3.13

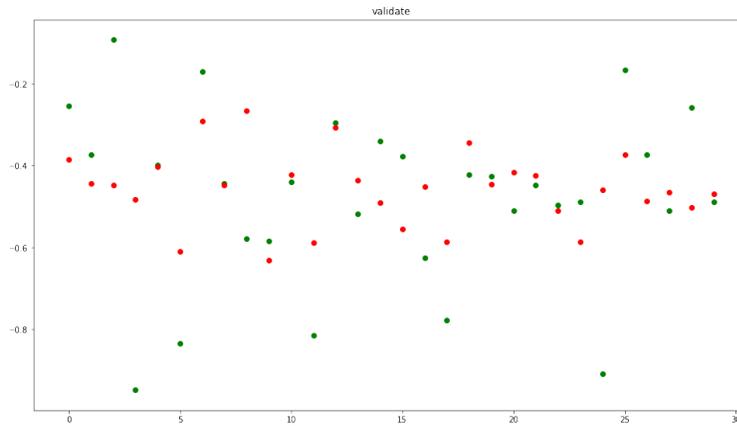


Figura 3.13: Gráfico de Datos del Comportamiento de la Red Neuronal contra los Datos Originales.

En la Figura 3.13 se puede observar que los punto verdes intentan aproximarse a los rojos. Cuanto mas cercanos estén la aproximación es mejor.

Finalmente se observa cómo disminuye la función de perdida (Loss) tanto en el conjunto de entrenamiento como el de validación, en la Figura 3.14 se muestra dicha comparación, y se puede decir que el modelo está aprendiendo y el modelo no cuenta con sobre ajuste (Overfitting) pues ambas curvas son distintas.

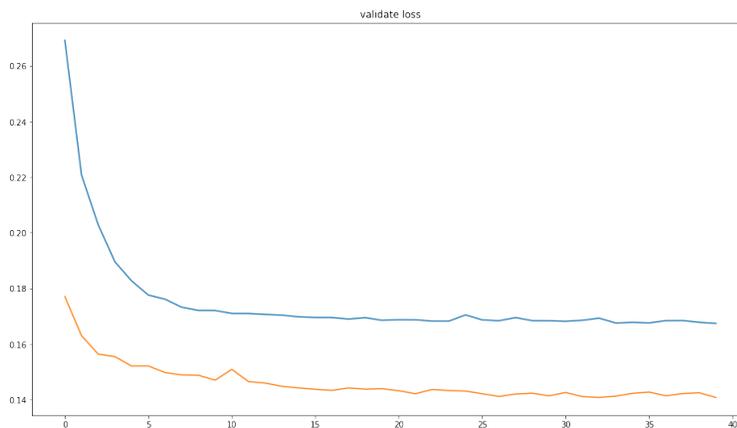


Figura 3.14: Curva de pérdida.

En la Tabla ??, se procede a realizar una nueva predicción, en este caso se usarán los últimos días de noviembre de 2018 para calcular la primer semana de diciembre.

Tabla 3.8: Datos de la mitad de los días del mes de noviembre del 2018.

Fecha	Valor
2018-11-16	203.92
2018-11-17	184.66
2018-11-18	182.96
2018-11-19	198.96
2018-11-20	201.18
2018-11-21	209.51
2018-11-22	278.92
2018-11-23	316.00
2018-11-24	222.92
2018-11-25	207.85
2018-11-26	185.92
2018-11-27	213.20
2018-11-28	213.20
2018-11-29	213.20
2018-11-30	213.20

Bajo la metodología anterior y el mismo procesado de datos que se realizo para el entrenamiento: escalado de valores sin incluir la columna de salida (t) puesto que es el valor a encontrar en la matriz que se muestra en la Tabla 3.9.

Tabla 3.9: Matriz de entradas y salidas.

	$(t - 7)$	$(t - 6)$	$(t - 5)$	$(t - 4)$	$(t - 3)$	$(t - 2)$	$(t - 1)$
7	-0.24	-0.65	0.31	0.30	0.07	0.24	0.11
8	-0.65	0.31	0.30	0.07	0.24	0.11	0.13
9	0.31	0.30	0.07	0.24	0.11	0.13	-1.00
10	0.30	0.07	0.24	0.11	0.13	-1.00	1.00
11	0.07	0.24	0.11	0.13	-1.00	1.00	0.44
12	0.24	0.11	0.13	-1.00	1.00	0.44	0.07
13	0.11	0.13	-1.00	1.00	0.44	0.07	0.75

De la tabla anterior se toma sólo la última fila, pues es la que correspondería a la última semana de noviembre.

Luego se crea una función de auto rellenado del desplazamiento que se realiza para cada predicción.

Entonces para el 1 de diciembre, se tiene el conjunto de datos con los últimos 7 días de noviembre. Pero para pronosticar el 2 de diciembre necesitamos los 7 días anteriores que incluyen al 1 de diciembre y ese valor, lo obtenemos en nuestra predicción anterior. Y así hasta el 7 de diciembre.

Finalmente, aplicada la función que realiza el pronóstico, los resultados se muestran en la Tabla 3.10.

Tabla 3.10: Pronóstico primer semana de Diciembre.

Fecha	Valor Pronosticado
2018-12-01	174.48
2018-12-02	147.26
2018-12-03	225.49
2018-12-04	203.73
2018-11-05	177.30
2018-11-06	208.15
2018-11-07	175.23

Finalmente, la gráfica del pronóstico se representa en la Figura 3.15

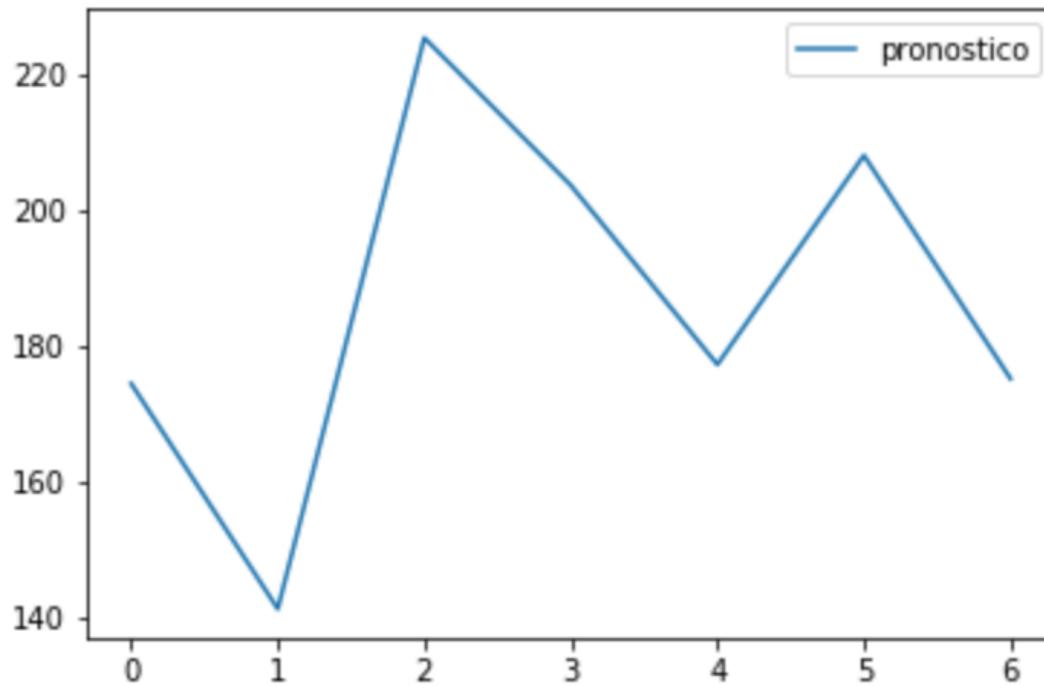


Figura 3.15: Pronóstico de ventas en los próximos 7 días.

Capítulo 4

Caso de estudio

Pocas cosas pueden impactar más a la población entera de un país que una crisis económica. Un claro ejemplo es la crisis de 1995 en México que hizo que una gran parte de la población perdiera su trabajo, su casa o coche entre otras cosas.

Con esta crisis, aumento en más de 15 millones el número de personas que se encontraban en situación de pobreza extrema y causó una pérdida enorme en el poder adquisitivo de casi todos los mexicanos.

La crisis del 2008 causó grandes estragos en los mercados del país, volvió a generar desempleo y trajo el cierre de muchos negocios. Esta vez el origen fue externo, producto de desequilibrios financieros y burbujas en los precios de la vivienda en Estados Unidos. En ambos casos atropellaron los planes de retiro y desarticulaban los ahorros de muchas familias; hubo un aumento descomunal del desempleo, bajó el consumo y se afectó el bienestar de casi toda la población [11].

La marcha de la economía afecta a todos los mexicanos. En tiempo de auge hay más oportunidades de empleo, las familias pueden gastar más y las empresas ampliar sus negocios; cuando hay recesión, todos tienen que ser cautelosos y estar más conscientes de sus gastos.

La economía del país se rige a través de una serie de indicadores económicos dirigidos y monitoriados a través de una institución llamada Instituto Nacional de Estadística, Geografía e Informática (**INEGI**).

El INEGI tiene como principal objetivo informar a la sociedad y dar a conocer mediante indicadores el estado y evolución de la actividad económica del país.

Como un indicador a corto plazo se encuentra el Índice Global de la Actividad Económica (**IGAE**).

El IGAE es un indicador que mide la evolución de las empresas productoras de bienes y servicios, así como el grupo dedicado a la comercialización, en el corto plazo de forma mensual.

El IGAE ofrece una visión oportuna, completa y coherente de la evolución de la actividad económica del país, proporcionando información oportuna, para apoyar en la toma de decisiones. Se elabora utilizando el mismo marco conceptual y metodológico empleado en el cálculo de las Cuentas de Bienes y Servicios del Sistema de Cuentas Nacionales de México (SCNM), complementándose con el Manual de Cuentas Nacionales Trimestrales, conceptos, fuentes de datos y compilación (CNT) del Fondo Monetario Internacional. Es importante señalar que el cálculo del Producto Interno Bruto Trimestral (PIBT) y el cálculo mensual del Indicador Mensual de la Actividad Industrial (IMAI), utilizan los mismos métodos, la misma clasificación por actividades económicas y las mismas fuentes de información que cuentan con una gran oportunidad mensual. Garantizando, con ello, la compatibilidad entre los productos de corto plazo [10].

En el año 2008 se publicaron las series mensuales del IGAE con el año base 2003, abarcando el período 2003 a 2008. La actualización del año base cada diez años de todos los productos del Sistema de Cuentas Nacionales de México (SCNM) dió la oportunidad de adoptar el nuevo Sistema de Clasificación Industrial de América del Norte (SCIAN), que en su versión 2002 permitía llevar a cabo comparaciones con las estadísticas de nuestros socios en el Tratado de Libre Comercio de América del Norte (TLCAN). Posteriormente, en una segunda etapa de los trabajos de modernización de los productos del Sistema de Cuentas Nacionales de Corto Plazo, se dieron a conocer series retropoladas del IGAE abarcando hasta el año 1993, las cuales incrementaron la disponibilidad de información de coyuntura, atendiendo con ello la demanda de los usuarios de esta información.

En agosto de 2013 se actualizó el Sistema de Cuentas Nacionales de México al año base 2008, esta se realizó conforme a las recomendaciones internacionales en la materia. Asimismo, se adoptó el SCIAN 2007. Para los cálculos del IGAE base 2008 se contó con una serie de datos de enero de 1993 al mes de junio de agosto de 2017, atendiendo las recomendaciones de los usuarios nacionales e internacionales en el marco de las Cuentas Nacionales, entre las que sobresalió la asignación de los servicios de intermediación financiera medidos indirectamente (SIFMI) por clase de actividad económica [10].

La cifras mensuales del IGAE se encuentran en la página del INEGI, en el apartado del banco de información económica (BIE), la serie de datos con la que se trabaja es la serie original total, la cual toma al 2013 como año base, teniéndose información desde el mes de enero de 1993, y ésta se expresa en índices de volumen físico, donde se considera a un año base como referencia para comparar los siguientes años, actualmente se tiene como año base al 2013 con un índice de volumen físico de 100 y a partir de éste valor se tiene un aumento del IGAE o lo contrario, con índices mensuales, acumulados y sus respectivas variaciones anuales.

A partir del 2013 el cálculo del IGAE utiliza el esquema conceptual y metodológico de las cuentas de bienes y servicios del Sistema de Cuentas Nacionales de México (SCNM), donde el índice de volumen físico de las actividades económicas, primarias, secundarias y terciarias, contienen un volumen físico para las clases más representativas.



Figura 4.1: Actividades económicas dominantes en México [6].

El análisis se hace para la República Mexicana, tomando en cuenta las actividades primarias, secundarias y terciarias. La forman las actividades económicas con mayor peso y con información. La clasificación de las actividades económicas son muy diversas y son clasificadas de acuerdo al INEGI [11]. En la Figura 4.1 se clasifican las actividades dominantes en México de acuerdo al INEGI.

Para la elaboración de éste indicador se dispone de datos estadísticos provenientes de las actividades primarias, secundarias y terciarias, en la Tabla 4.1 se presenta la clasificación de las actividades que incorporan al IGAE [6].

Tabla 4.1: Clasificación de actividades económicas.

Actividades Primarias	Agricultura, Ganadería, Explotación forestal, Pesca.
Actividades Secundarias	Minería, manufacturas, electricidad, construcción, agua, suministro de gas producto al consumidor final.
Actividades Terciarias	Comercio, transporte, correo y almacenamiento, servicio financieros y de seguros, servicios de alojamiento temporal y de preparación de alimentos y bebidas, actividades gubernamentales, información en medios masivos, servicios inmobiliarios y de alquiler de bienes muebles e intangibles, servicios profesionales, científicos y técnicos, servicio de apoyo al negocio, servicio educativo, servicio de salud, servicio de esparcimiento, deportivo y otros servicios recreativos y servicio de intermediación financiera medido indirectamente.

Es importante destacar que la información básica que incorpora el IGAE contiene datos muy preliminares y está sujeta a revisión por parte de las empresas y organismos públicos y privados; adicionalmente, no incluye la totalidad de las actividades como lo hace el PIB. Sin embargo el IGAE tiene la ventaja de ser un indicador a corto plazo que muestra la dirección que va tomando la economía del país.

4.1. Metodología del cálculo

El método utilizado para calcular el Indicador Global de la Actividad Económica es el de la producción, la base de cálculo consiste en elaborar índices mensuales de volumen físico de la producción para cada una de las clases seleccionadas, con base fija en el año 2013, con estos índices se realiza la extrapolación del valor agregado bruto del año base, asumiendo el supuesto de que el valor agregado bruto tiene el mismo comportamiento que el valor bruto de la producción.

La producción típica se sustenta en un amplio y homogéneo detalle de los productos que elaboran las empresas.

Para lograr resultados regulares y consistentes se emplea un método estadístico donde las cantidades se cuentan en precios constantes. Se utiliza la fórmula de Laspeyres, de agregación ponderada, con base en los precios del año 2013. De esta forma se suma para cada periodo, el valor de la producción de cada clase en términos reales de 2013, y dispone de un índice de volumen físico directo y ponderado para cada clase. La medición mensual real de cada clase utiliza la fórmula de Laspeyres para cada producto o clasificación de productos, lo que permite obtener cada periodo de forma agregada

en términos reales del 2013, además de contar con el volumen físico para cada clase de las actividades económicas.

4.2. Análisis de los datos del IGAE usando la metodología de Box-Jenkins (Modelo *SARIMA*).

Los datos analizados se toman mensualmente, a partir de Enero de 1993 hasta Diciembre de 2019, por temas de una pandemia suscitada en el año 2020 las observaciones de ese año no serán tomadas en cuenta por la variación que estas pudieran tener. La información recaudada es la de 324 meses es decir 27 años.

En la figura 4.2 se presenta de manera gráfica nuestra data de información.

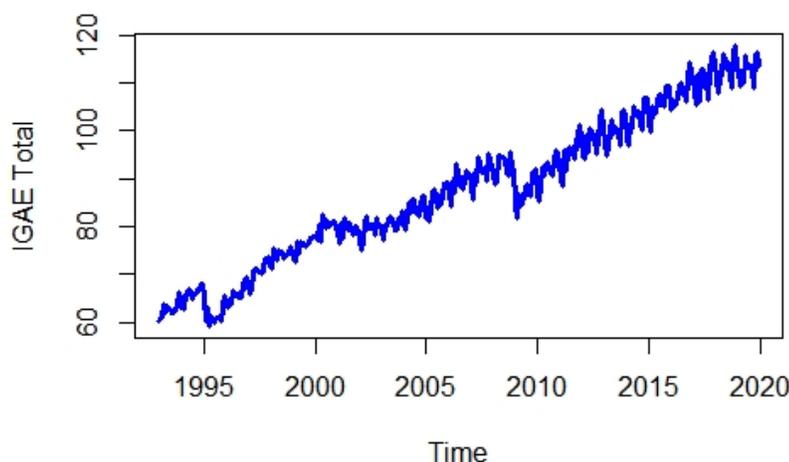


Figura 4.2: Serie de tiempo obtenida para el IGAE.

En la Figura 4.2 se muestra claramente que la serie no tiene una media y varianza constante por lo que se tiene que realizar una transformación para poder convertirla a estacionaria.

Se observa que la serie muestra dos caídas aproximadas en los años 1995 y 2009, dadas a las respectivas crisis económicas que sufrió el país en esos años, por eso se consideran los modelos de intervención dentro de los modelos *ARIMA* o *SARIMA*.

A continuación se presenta el análisis de la serie sin tomar en cuenta los modelos de intervención y después agregándolos, para así contrastar su eficacia y determinar el mejor modelo.

4.2.1. Análisis del modelo sin intervención

Anteriormente se presentó que la serie de tiempo no es estacionaria respecto a su varianza, se realiza una pre-diferenciación con el objeto de estabilizar dicha varianza, a continuación se presenta en la Figura 4.3 tres tipos de pre-diferenciación y se concluye que la mejor es aquella que utiliza raíces cuartas del IGAE ya que esta genera una serie con variación estacional constante.

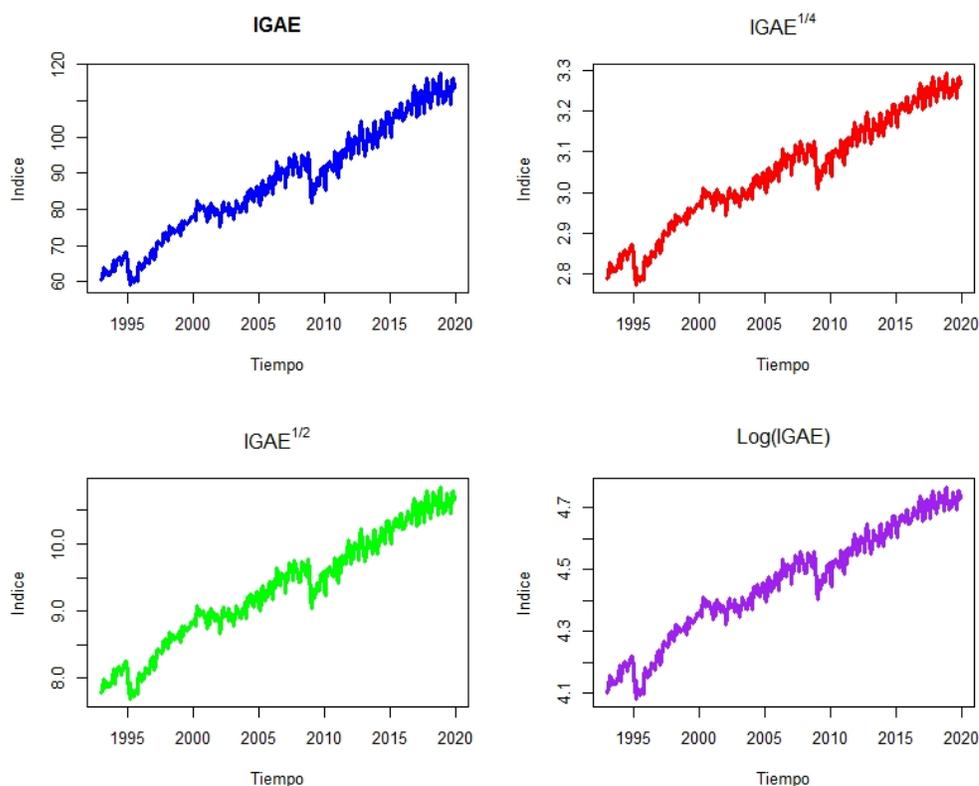


Figura 4.3: Representación de las series después de su transformación.

Los datos originales del IGAE se representan como X_t , los datos transformados por la pre-diferenciación de *raíz cuarta* por X_t^* , sin embargo la serie sigue siendo no estacionaria ya que no tiene una media constante, así que se aplica una diferencia a la serie X_t^* obteniéndose una nueva, cuyos valores serán representados por X_t^1 .

En la Figura 4.4 se muestra el gráfico de la serie de tiempo con una diferenciación, y enseguida se muestra en la Figura 4.5 su FAM y su $FAMP$ donde en su FAM hay evidencia de estacionalidad con un periodo de $s = 12$

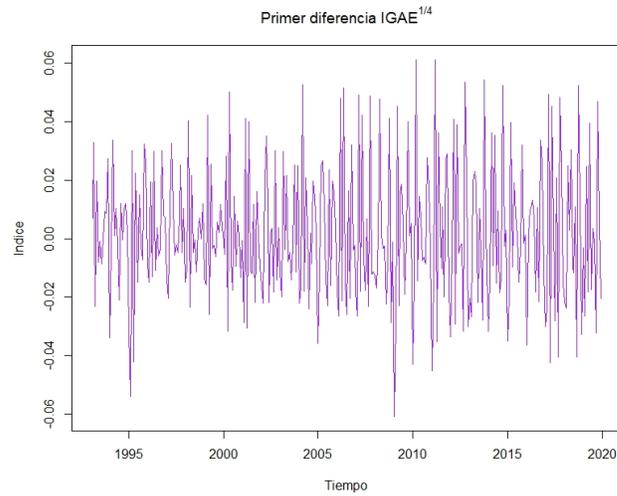


Figura 4.4: Serie X_t^1 con una diferencia.

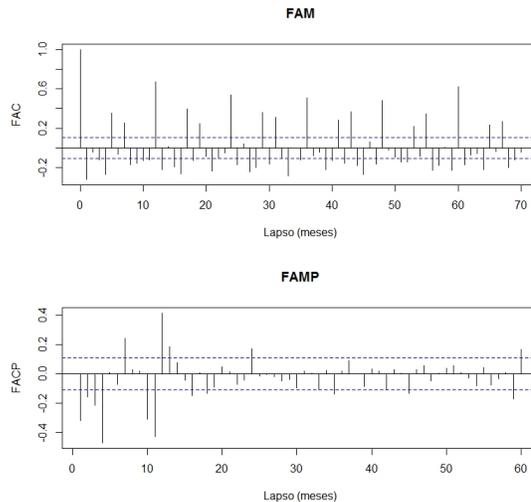


Figura 4.5: FAM y FAMP de la serie X_t^1 .

La estacionalidad también puede ser comprobada mediante el análisis de diagrama de cajas y bigotes, las cuales es una manera conveniente de mostrar de manera visual grupos de datos numéricos a través de sus cuartiles.

Los cuales, resumen la información de medidas estadísticas tales como: valor mínimo, valor máximo, primer cuartil, mediana y tercer cuartil.

Las líneas se extienden en paralelas a las cajas se les conoce como bigotes y se utilizan para indicar la variabilidad fuera de los cuartiles superior e inferior. Los valores atípicos se representan a veces como puntos individuales que están en línea con los bigotes. Los diagramas de cajas y bigotes se pueden dibujar vertical u horizontalmente.

Dicho diagrama es representado en la Figura 4.6 donde los valores tiene un leve incremento en los últimos periodos pero sin embargo no hay una notable estacionalidad con $s = 12$.

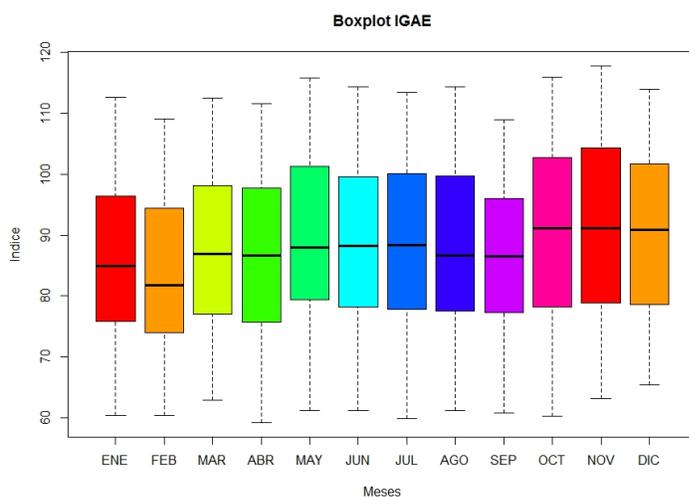


Figura 4.6: Diagrama de cajas y bigotes.

Como la metodología de Box-Jenkins lo indica, se procede a realizar un diagnóstico de estacionariedad en el nivel no estacional, donde el comportamiento parece ser senoidal amortiguado, mediante la prueba de Dickey-Fuller los resultados se muestran en la Tabla 4.2.

Tabla 4.2: Prueba Dickey-Fuller para la serie con una diferencia.

Test Dickey-Fuller	
Dickey-Fuller	-7.3026.
Lag order	6
p-value	0.01

Donde se puede observar que se obtiene un $p - valor = 0.01$ por lo que se puede concluir que la serie es estacionaria para cualquier nivel de significancia razonable ($\alpha > 0.1$). Posteriormente se deben de tomar medidas para eliminar la estacionalidad, se puede observar que los valores se extinguen con lentitud, por lo tanto se debe analizar la parte estacional, ya que se pone en duda la estacionariedad estacional.

En la Figura 4.7 se presenta la serie X_t^* a nivel estacional con un período de 12 meses ($s=12$), cuyo propósito es realizar el análisis a nivel estacional $s=12$, por otro lado se observa que su FAM se desvanece lentamente, por lo que se puede concluir que la serie no es estacionaria, y que su FAMP se desvanece muy rápidamente en su nivel no estacional y estacional, también se puede observar en la gráfica inferior que los datos no rondan sobre una misma media ni tienen una varianza constante.

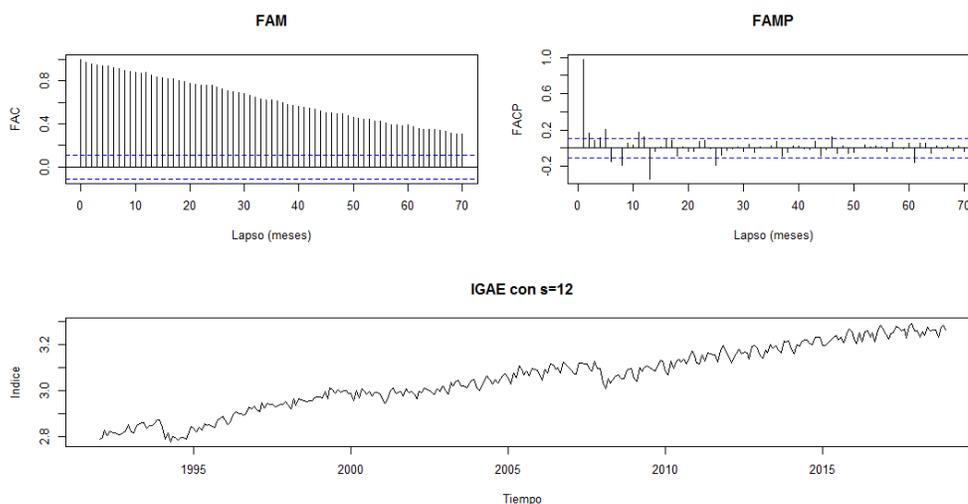


Figura 4.7: Gráfico FAM, FAMP y comportamiento del $IGAE^{\frac{1}{4}}$ nivel estacional.

Con ayuda del test de Dickey-Fuller mostrado en la Tabla 4.3 se puede observar que arroja un p - *valor* = 0.3399 de tal forma que si $\alpha = 0.01$ la conclusión es que la serie es no estacionaria.

Tabla 4.3: Prueba Dickey-Fuller para la serie transformada por la raíz cuarta.

Test Dickey-Fuller	
Dickey-Fuller	-2.5612
Lag order	6
p-value	0.3399

A continuación se verifican las siguientes series para mostrar cual seria la indicada para el modelo y así obtener el modelo adecuado.

Luego siguiendo con el análisis en la Figura 4.8 se muestra la serie X_t^1 a nivel estacional, donde se puede notar que la FAM va disminuyendo, sin embargo lo hace lentamente. Aplicando la prueba Dickey-Fuller mostrada en la Tabla 4.5 arroja un p -valor = 0.01 así la conclusión es que la serie es estacionaria para cualquier nivel de significancia razonable ($\alpha > 0.01$).

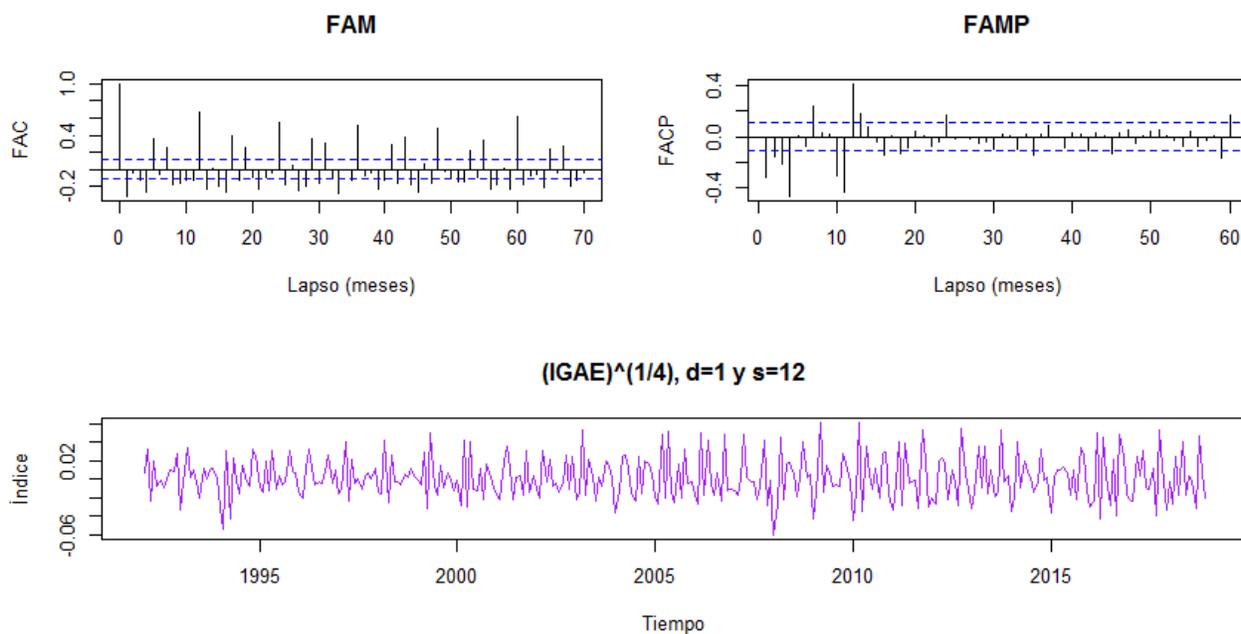


Figura 4.8: Gráfico FAM, FAMP y comportamiento del $IGAE^{\frac{1}{4}}$ con una diferencia no estacional.

Tabla 4.4: Prueba Dickey-Fuller para la serie con una diferencia.

Test Dickey-Fuller	
Dickey-Fuller	-7.3026
Lag order	6
p-value	0.01

Finalmente se toma en cuenta la serie X_t^1 a nivel estacional, con una diferencia en el nivel estacional, el modelo sigue poniendo en duda su estacionariedad ya que su FAM decae muy lentamente, tanto en el nivel no estacional como en el estacional, la prueba arroja un $p - valor = 0.01$ por lo tanto se puede afirmar que es estacionaria.

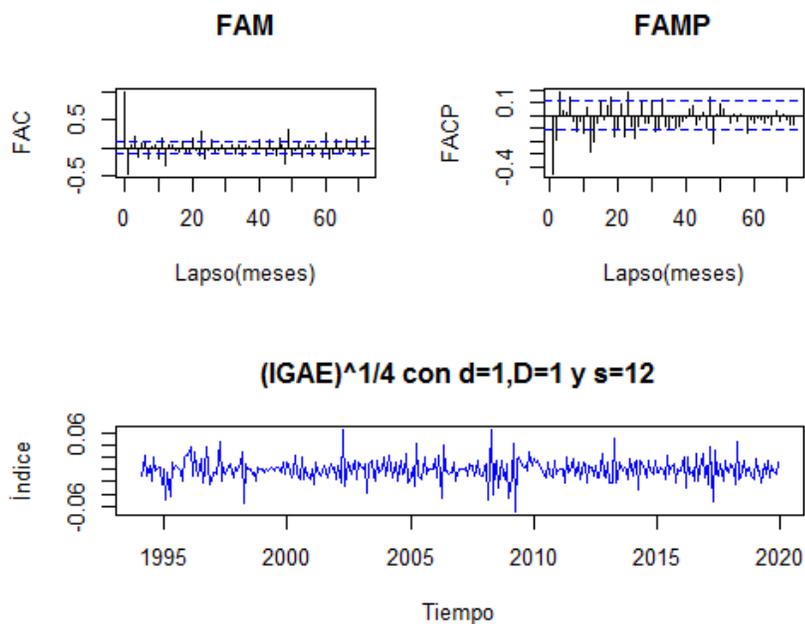


Figura 4.9: $IGAE^{\frac{1}{4}}$ con una diferencia en el nivel no estacional y en el estacional.

Tabla 4.5: Prueba Dickey-Fuller para la serie con una diferencia.

Test Dickey-Fuller	
Dickey-Fuller	-7.3026
Lag order	6
p-value	0.01

Con lo anterior un modelo estacional con período de $s=12$ en un principio sera adecuado para pronosticar los datos. Con ayuda del Software R, se comprobó que existen dos modelos adecuados con estacionalidad de período $s=12$.

Así bajo el criterio de Akaike, los resultados mostrados en la Tabla 4.6 se observa, que los modelos son adecuados, puesto que no se rechaza tal hipótesis con cualquier nivel de significancia razonable. Por lo que se concluya que los mas aceptables con estacionalidad de periodo $s=12$, son adecuados para predecir los datos del IGAE. Con ayuda del Software R, de acuerdo al Criterio de Akaike se selecciono al modelo

Tabla 4.6: Criterio de Akaike

Modelo	CIA
$SARIMA(4, 0, 1)(2, 1, 2)_{12}$	-1810.03
$SARIMA(4, 1, 1)(1, 1, 1)_{12}$	-1803.79
$SARIMA(2, 1, 0)(2, 1, 0)_{12}$	-1770.59
$SARIMA(2, 1, 2)(2, 1, 1)_{12}$	-1806.00

más adecuado, en la Tabla 4.6 anterior, donde arroja como mejor resultado al modelo **$SARIMA(4, 0, 1)(2, 1, 2)_{12}$** puesto que el estadístico del criterio de Akaike, toma este modelo.

Por lo tanto, el modelo propuesto queda:

$$\Psi_4(B^{12})\psi_2(B)\nabla_{12}^1 X_t^* = \Theta_1(B^{24})\theta_1(B)Z_t \quad (4.1)$$

donde $\nabla_{12}^1 = (1 - B)^1$ además de que $X_t^* = X_t^{\frac{1}{4}}$ es una observación del IGAE en el tiempo t . Otra forma de presentar el modelo es considerar que

$$X_t^1 = X_t^1 - X_{t-1}^1 - X_{t-1}^1 + X_{t-13}^1$$

Luego en la parte no estacionaria se tiene el modelo

$$X_t^1 = \psi_1 X_{t-1}^1 + \psi_2 X_{t-2}^1 + \psi_3 X_{t-3}^1 + \psi_4 X_{t-4}^1 + \theta_1 Z_{t-1}$$

En la parte estacional se propone

$$X_t^1 = \Psi_1 X_{t-1}^1 + \Psi_2 X_{t-2}^1 + Z_t + \Theta_1 Z_{t-1} + \Theta_2 Z_{t-2}$$

Por lo que el modelo integrado que de la siguiente forma,

$$(1 - B - B^2 - B^3 - B^4)(1 - B^{12} - B^{24})(1 - B^{12})X_t^1 = (1 - B^{24})(1 - B)Z_t \quad (4.2)$$

Desarrollando la Ecuación 4.2 tenemos el siguiente polinomio.

$$\begin{aligned}
X_t^1 = & X_{t-1}\psi_1 + X_{t-2}\psi_2 + X_{t-3}\psi_3 + X_{t-4}\psi_4 + X_{t-12} + X_{t-12}\psi_1 - X_{t-3}\psi_1 \\
& + X_{t-13}\psi_1\Psi_1 - X_{t-14}\psi_2 - X_{t-14}\psi_2\Psi_2 - X_{t-15}\psi_3 - X_{t-15}\psi_3\Psi_1 \\
& - X_{t-16}\psi_4 - X_{t-16}\psi_4\Psi_1 + X_{t-24}\Psi_2 - X_{t-24}\Psi_1 - X_{t-25}\psi_2\Psi_2 + X_{t-25}\psi_1\Psi_1 \\
& - X_{t-26}\psi_2\Psi_2 + X_{t-26}\psi_2\Psi_1 - X_{t-27}\psi_3\Psi_2 + X_{t-27}\psi_3\Psi_1 - X_{t-28}\psi_4\Psi_2 + X_{t-28}\psi_4\Psi_1 \\
& + X_{t-36}\Psi_2 + X_{t-37}\psi_1\Psi_2 + X_{t-38}\psi_2\Psi_2 + X_{t-39}\psi_3\Psi_2 + X_{t-40}\psi_4\Psi_2 \\
& - Z_t + \theta_1 Z_{t-1} + \Theta_1 Z_{t-24} + \theta_1 \Theta_1 Z_{t-25}.
\end{aligned}$$

Con ayuda del software R se realiza la estimación de parámetros, utilizando el método de máxima verosimilitud, los resultados se presentan en la Tabla 4.7

Tabla 4.7: Valores obtenido del modelo $ARIMA(4, 0, 1)(2, 1, 2)_{12}$

Parámetro	ar1	ar2	ar3	ar4	ma1	sar1	sar2	sma1	sma2
Valor	0.99	0.10	0.12	-0.28	-0.54	-0.46	-0.03	-0.13	-0.47
S.e.	0.16	0.11	0.09	0.05	0.17	0.14	0.09	0.14	0.11

Los resultados obtenidos deben ser elevados a la cuarta potencia ya que recordemos que $X_t^* = X_t^{\frac{1}{4}}$ por lo tanto $\hat{X}_t^{\frac{1}{4}}$ se sustituye por $(\hat{X}_t^{\frac{1}{4}})^4 = \hat{X}_t$. Una vez realizado esto se usa el modelo propuesto $SARIMA(4, 0, 1)(2, 1, 2)_{12}$ para pronosticar el IGAE teniendo como referencia a los datos de IGAE hasta Diciembre de 2018, con el propósito de pronosticar los últimos 12 datos del IGAE ya propuestos por el INEGI, para así poder observa la eficacia del modelo propuesto, mediante los errores totales y relativos.

Una vez estimados los parámetros el modelo queda como,

$$\begin{aligned}
X_t = & X_{t-1}^* - X_{t-12}^* + X_{t-13}^* + (0.90)(X_{t-1}^* - X_{t-12}^* + X_{t-13}^*) \\
& + (0.10)(X_{t-1}^* + X_{t-2}^* - X_{t-13}^* + X_{t-14}^*) \\
& + (0.12)(X_{t-3}^* + X_{t-4}^* - X_{t-14}^* + X_{t-15}^*) \\
& - (0.28)(X_{t-4}^* + X_{t-5}^* - X_{t-15}^* + X_{t-16}^*) \\
& + (0.90)(X_{t-1}^* - X_{t-12}^* + X_{t-13}^*) \\
& - (0, 90)(X_{t-1}^* - X_{t-12}^* + X_{t-13}^*) \\
& \vdots \\
& - Z_t - (0.54)Z_{t-1} + -(0.13)Z_{t-24} - (0.54)(0.13)Z_{t-25}
\end{aligned}$$

A continuación en la Tabla 4.8 se muestran los resultados.

Tabla 4.8: Pronóstico de Enero a Diciembre del 2019

Periodo	X	\hat{X}	Error	Relativo	Lo 80	Hi 80	Lo 95	Hi 95
01/2019	112.71	110.82	1.89	0.01	108.57	113.11	107.39	114.34
02/2019	109.09	108.43	0.65	0.00	105.94	110.97	104.64	112.32
03/2019	112.46	111.27	1.19	0.01	108.38	114.22	106.88	115.80
04/2019	109.98	111.84	-1.85	-0.01	108.41	115.35	106.63	117.24
05/2019	115.45	115.01	0.43	0.00	111.24	118.88	109.28	120.97
06/2019	113.04	114.30	-1.26	-0.01	110.23	118.48	108.12	120.74
07/2019	113.54	113.18	0.35	0.00	108.84	117.66	106.59	120.08
08/2019	113.42	113.51	-0.085	-0.00	108.92	118.24	106.54	120.81
09/2019	109.01	109.78	-0.77	-0.00	105.06	114.66	102.63	117.31
10/2019	115.46	115.91	-0.44	-0.00	110.76	121.23	108.11	124.12
11/2019	116.32	117.68	-1.36	-0.01	112.27	123.29	109.48	126.34
12/2019	113.47	115.18	-1.71	-0.01	109.65	120.92	106.80	124.04

En la Figura 4.10 se muestran los datos pronosticados.

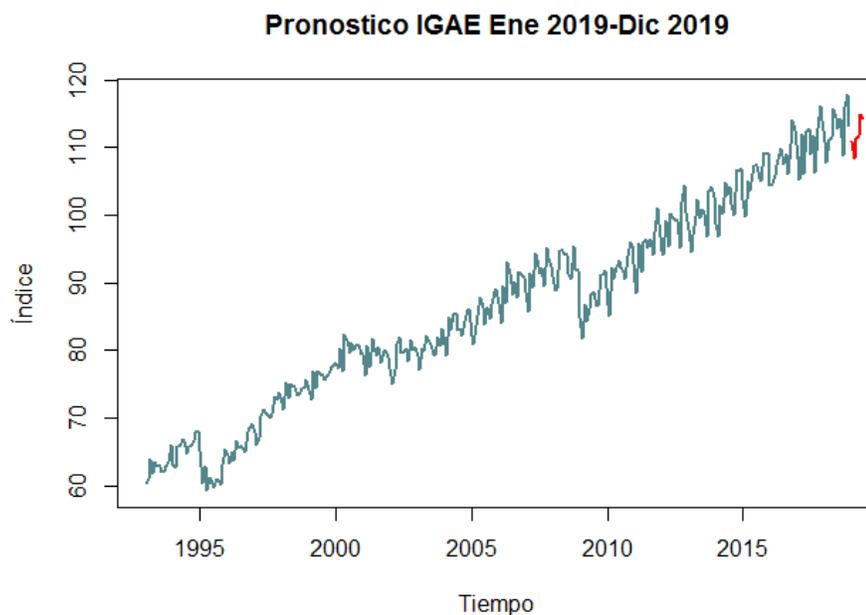


Figura 4.10: Pronóstico de Enero a Diciembre 2019.

En la figura 4.11 se muestra la serie de tiempo original sobre la ajustada.

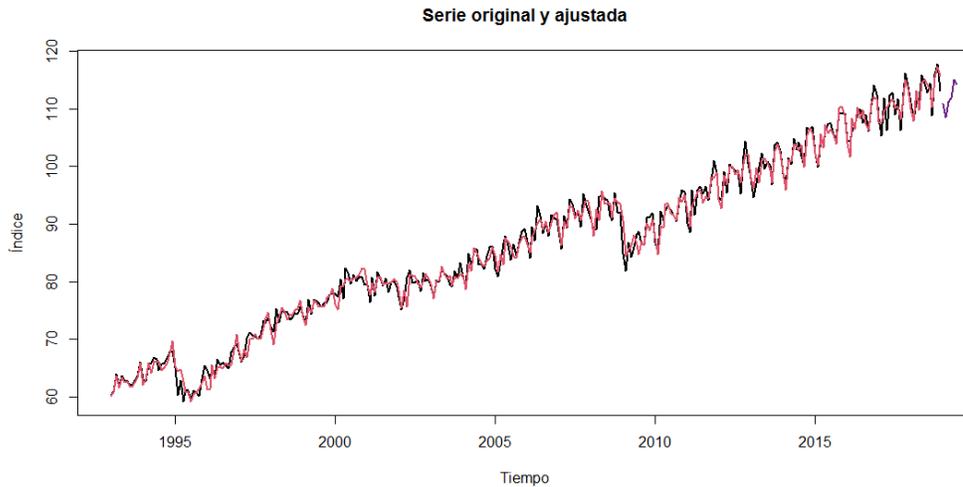


Figura 4.11: Comparación de los datos originales con la propuesta obtenida por el modelo.

Ahora bien, se obtiene el pronóstico de interés, es decir el IGAE de Enero de 2020 a Diciembre del 2020.

Los resultados, así como sus intervalos de confianza se muestran en la Tabla 4.9

Tabla 4.9: Pronostico Enero a Diciembre del 2020

Periodo	\widehat{X}	Lo 80	Hi 80	Lo 95	Hi 95
01/2020	110.9	108.60	113.11	107.59	114.40
02/2020	107.60	109.34	115.17	103.94	111.42
03/2020	112.17	109.3	115.1	107.8	116.6
04/2020	109.7	106.4	113.1	104.7	114.9
05/2020	114.3	110.6	118.1	108.7	120.1
06/2020	113.4	109.5	117.5	107.4	119.7
07/2020	112.5	108.3	116.9	106.1	119.2
08/2020	112.8	108.3	117.4	106.0	119.9
09/2020	109.0	104.4	113.7	102.0	116.3
10/2020	115.0	110.0	120.2	107.4	123.0
11/2020	116.5	111.2	122.0	108.5	124.9
12/2020	114.4	109.0	120.0	106.3	123.1

En la Figura 4.12 Se muestran los datos pronosticados para el año 2020.

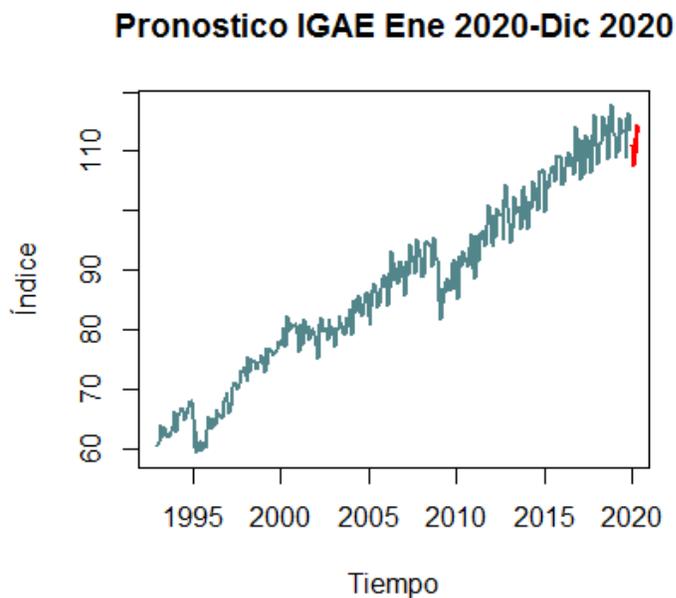


Figura 4.12: Pronostico IGAE para Enero a Diciembre del 2020.

En la figura 4.13 se muestra la serie de tiempo original sobre la ajustada.

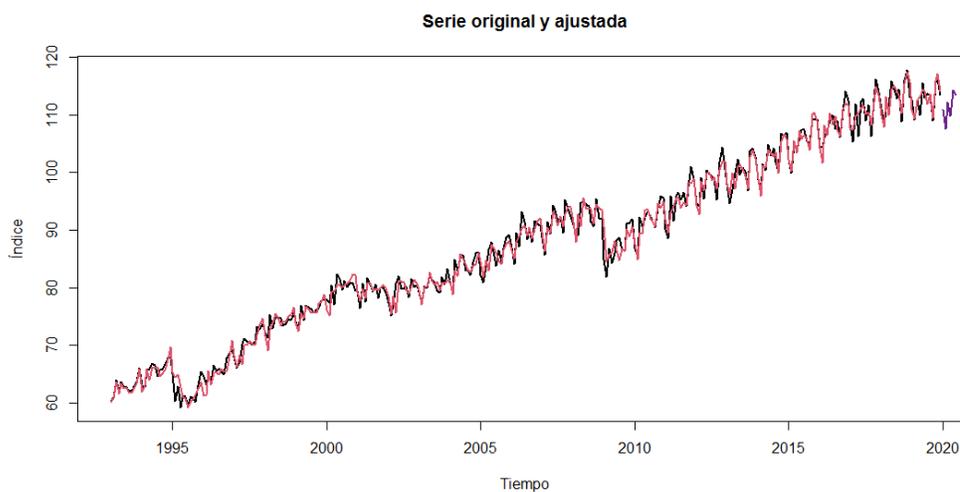


Figura 4.13: Serie original y propuesta del 2020.

Los residuales también son una buena estrategia de diagnóstico importante para examinar la dependencia de los residuales, las cuales muestran en éste caso que el componente estocástico de la serie es muy parecido a un ruido blanco el análisis se muestra en la Figura 4.14 en donde se puede concluir que se apoya la hipótesis

de que los residuales tienden a estar normalmente distribuidos. Por lo tanto se determina que el mejor modelo para representar de forma adecuada a los datos es un $SARIMA(4, 0.1)(2; 1; 2)_{12}$.

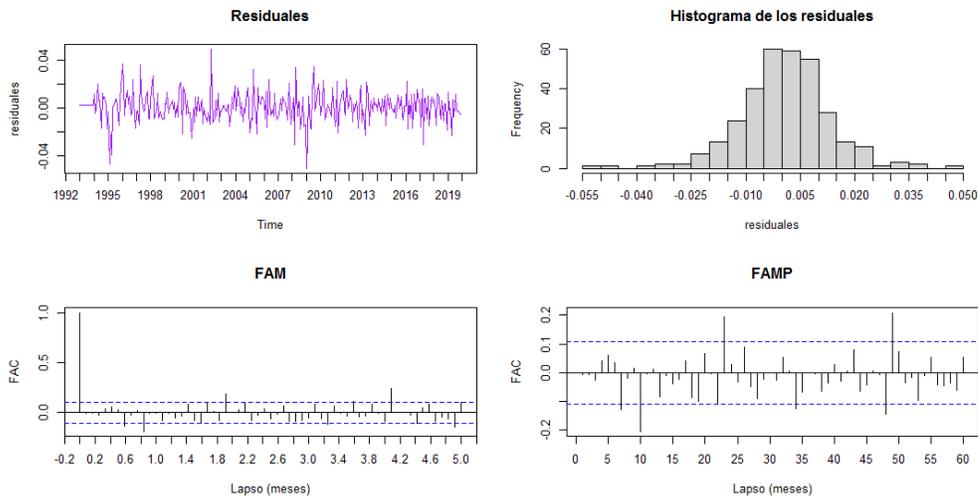


Figura 4.14: Gráfico de residuales del modelo $SARIMA(4,0,1)(2,1,2)_{12}$.@

4.3. Análisis de la serie de datos del IGAE con Redes Neuronales Artificiales.

Continuando con el análisis de datos del IGAE, se presenta un análisis utilizando la metodología de Redes Neuronales Artificiales con el único objetivo de poder crear una red neuronal que sea capaz de pronosticar el indicador IGAE a base de entrenamiento.

A partir de la información obtenida en la página del Instituto Nacional de Estadística, Geografía e Informática (INEGI) se puede realizar la extracción de un archivo de entradas con los índices a través del tiempo del IGAE en México.

Una vez extraída la información con ayuda del Software Python se procede a crear una red neuronal que sea capaz de pronosticar dicho indicador a través del tiempo. Existen diversos métodos para hacer pronóstico. En nuestro caso, y como se mostró en la sección anterior parecen comportarse bastante parecidas cada año. Un método muy utilizado en estadística es el llamado ARIMA, el cual se explico en la sección anterior.

En esta sección se aplicara un modelo Machine Learning: una red neuronal para hacer el pronóstico. Curiosamente crear esta red es completamente computacional, y en aquí se mostrara un modelo de lo más moderno para hacer el pronóstico. Se usara una arquitectura sencilla de red neuronal FeedForward (también llamada MLP por sus siglas Multi-Layered Perceptron), con pocas neuronas y como método de activación la tangente hiperbólica pues entregaremos valores transformados entre -1 y 1.

4.3.1. Análisis de las muestras para la creación de la Red.

En la Figura 4.15 se muestra el comportamiento de los datos tomando las fechas de Enero de 1993 a Diciembre de 2019 con la intención de pronosticar el año 2020.

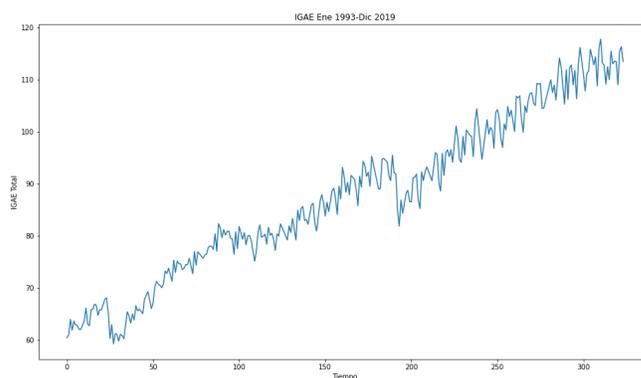


Figura 4.15: Comportamiento de los datos de Enero de 1993 a Diciembre de 2019.

En la Tabla 4.10 se muestra como se encuentra dividida la base de datos a trabajar, en donde podemos observar que del lado derecho se encuentran las fechas de registro y del lado izquierdo los valores obtenidos en dichas fechas.

Tabla 4.10: Primeras observaciones.

Año	Valor
1993-01-01	60.41
1993-01-02	61.02
1993-01-03	63.94
1993-01-04	61.87

Continuando con el análisis de información en la Tabla 4.11 se muestra la cantidad de datos obtenidos en cada año, con lo que se observa que cada año cuenta con 12 registros y se puede concluir que la información a trabajar es mensual. Finalmente en

Tabla 4.11: Cantidad de observaciones por año.

Año	Obs.
1993	12
1994	12
1995	12
1996	12
...	...
2017	12
2018	12
2019	12

la Tabla 4.12 se presenta un resumen estadístico de la información de los datos del IGAE

Tabla 4.12: Resumen estadístico de los datos observados del IGAE.

conteo	324
mean	87.56
std	15.48
min	59.24
25 %	76.53
50 %	86.86
75 %	99.88
max	117.77

se observan 324 datos con una media total de 87.56, una desviación estándar de 15.48, como dato mínimo se tiene 59.24 y como máximo dato se tiene 117.77.

4.3.2. Algunas gráficas descriptivas de los datos.

En la Figura 4.16 se muestra el comportamiento de las medias en todos los años.

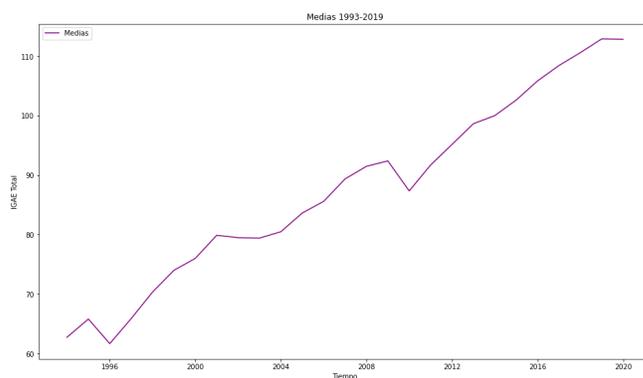


Figura 4.16: Gráfica de la medias a través del tiempo.

Se observa que la media de cada año ha ido en aumento a través de estos últimos 27 años por lo que se puede sugerir que el IGAE ha ido en aumento.

Ahora en la Figura 4.17 se comparan las medias de los primero dos años y de los dos últimos años observados y en donde se puede confirmar que el valor de los datos ha ido en aumento.

También se puede concluir que el comportamiento de las medias tiene casi el mismo movimiento.

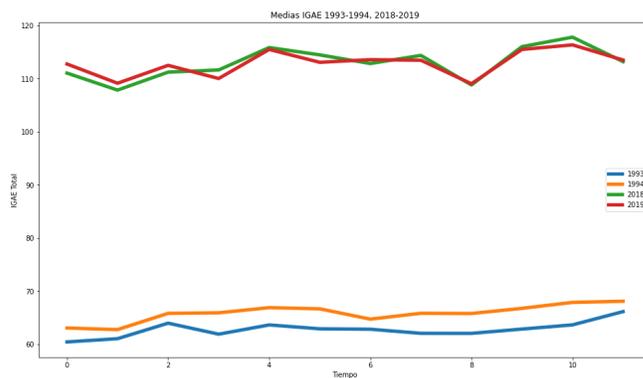


Figura 4.17: Gráfica de la medias a través del tiempo.

4.3.3. Preparación de los datos

Para la preparación de los datos se realiza la alteración del flujo de entrada de los datos que contiene una columna con los indicadores, y se convertirán en varias columnas.

Lo que se puede traducir como tomar la serie temporal y la convertirla en un “problema de tipo supervisado“ para poder alimentar la red neuronal y poder entrenarla con backpropagation. Para hacerlo, se deben tener unas entradas y unas salidas para entrenar al modelo.

Para el caso de estudio se toman los 12 últimos meses de un año para predecir el primero del siguiente año, por lo que se tiene.

- **Entradas:** 12 columnas que representan los 12 meses del año.
- **Salida:** El valor número 13 es decir el primer mes del siguiente año.

Hecho lo anterior, se procede a realizar el escalamiento de la información para transformar los valores en un intervalo de -1 a 1 para que la red neuronal tenga un mejor calculo. Este estimador de escala traduce cada característica individualmente de modo que esté en el rango dado en el conjunto de entrenamiento, por ejemplo entre cero y uno, esta transformación se utiliza a menudo como alternativa al escalado de varianza unitaria de media cero.

Una vez escalados los datos se procede a crear una función que toma una serie de tiempo univariante o multivariante y la enmarca como un conjunto de datos de aprendizaje supervisado.

La función toma los siguientes argumentos:

- **Datos:** Secuencia de observaciones como una lista o matriz.
- n_{in} : Número de observaciones de retraso como entrada (X).
- n_{out} : Número de observaciones como salida (Y).
- **dropnan:** Elimina los valores NAN.
- **return:** La función devuelve un solo valor.

El nuevo conjunto de datos se construye como un DataFrame, con cada columna nombrada adecuadamente tanto por número de variable como por paso de tiempo. Esto le permite diseñar una variedad de diferentes problemas de pronóstico del tipo de secuencia de pasos de tiempo a partir de una serie de tiempo univariante o multivariante determinada.

Una vez que se devuelve el DataFrame, se puede decidir cómo dividir las filas del DataFrame devuelto en componentes X e Y para el aprendizaje supervisado de la forma que desee.

La función se define con parámetros predeterminados, de modo que si la llama sólo con sus datos, construirá un marco de datos con $t - 1$ como X y t como y .

Hecho lo anterior se aplica la función a la base de datos quedando como sigue:

- **Datos:** Datos escalados.
- n_{in} : 12 Entradas.
- n_{out} : 1 salida.

En la Tabla 4.13 se muestra la matriz resultante.

Tabla 4.13: Matriz de entradas y salidas.

	$(t - 12)$	$(t - 11)$	$(t - 10)$...	$(t - 2)$	$(t - 1)$	(t)
12	-0.960021	-0.939177	-0.839399	...	-0.850333	-0.764907	-0.870494
13	-0.939177	-0.839399	-0.910132	...	-0.764907	-0.870494	-0.880745
14	-0.839399	-0.910132	-0.850675	...	-0.870494	-0.880745	-0.776183
15	-0.910132	-0.850675	-0.875619	...	-0.880745	-0.776183	-0.772766
16	-0.850675	-0.875619	-0.878353	...	-0.776183	-0.772766	-0.739621

Se usarán como entradas las columnas con encabezado $(t - 12)$ hasta $(t - 1)$ y la salida sera (t) la última columna.

4.3.4. Creación de la Red Neuronal Artificial.

Antes de crear la Red Neuronal, se subdividen los conjuntos de datos en una para el entrenamiento y el otro para su validación, algo muy importante a tomar en cuenta es que en este procedimiento, a diferencia de otros problemas es que en este caso es de suma importancia mantener el orden en la alimentación de la red. Por lo tanto se realiza una subdivisión de los primeros 288 meses consecutivos para el entrenamiento de la red y los siguientes 37 meses para su validación.

Con ayuda del Software Python se realiza dicho procedimiento obteniendo un arreglo con forma $(288,12,1)$ es decir 288 entradas con vectores de 12×1 .

Por lo que la arquitectura de la red neuronal queda de la siguiente manera:

- Entradas de 12 inputs.
- 1 capa oculta con 12 neuronas.
- La salida sera de solo 1 neurona.
- Como función de activación se utilizan **Tangente hiperbólica** puesto que los valores se encuentran entre -1 y 1.
- Utilizaremos como optimizador el **Método Adam**.

- La métrica de pérdida será el **Error Absoluto Medio**.
- Como la predicción será un valor continuo y no discreto, para calcular la precisión se utiliza **Error Cuadrático Medio**.
- Y finalmente para su entrenamiento utilizaremos 100, 000 iteraciones.

En la Figura 4.18 se muestra la estructura de la Red Neuronal implementada para la pronóstico del IGAE.

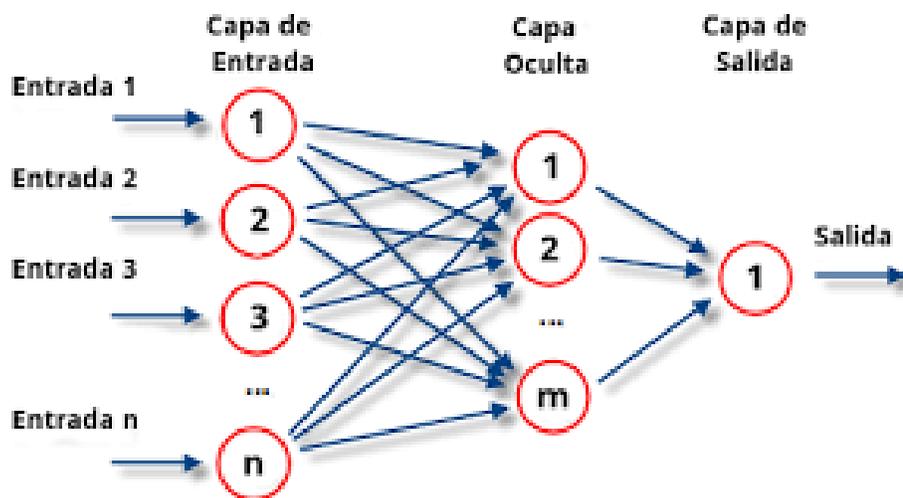


Figura 4.18: Estructura de la Red Neuronal con $n = 12$ y $m = 12$.

En la Tabla 4.14 se muestran los resultados después de 100,000 iteraciones de entrenamiento .

Tabla 4.14: Primeras observaciones.

Iteraciones	loss	ECM	Val loss	Val ECM
100,000	0.022	0.001	0.05	0.004

En la Tabla anterior se puede observar que los errores son muy cercanos a cero por lo que se puede concluir que el modelo es adecuado para el pronóstico.

Finalmente en la Figura 4.19 se muestra la comparación de los datos reales con los datos pronosticados por la Red Neuronal artificial.

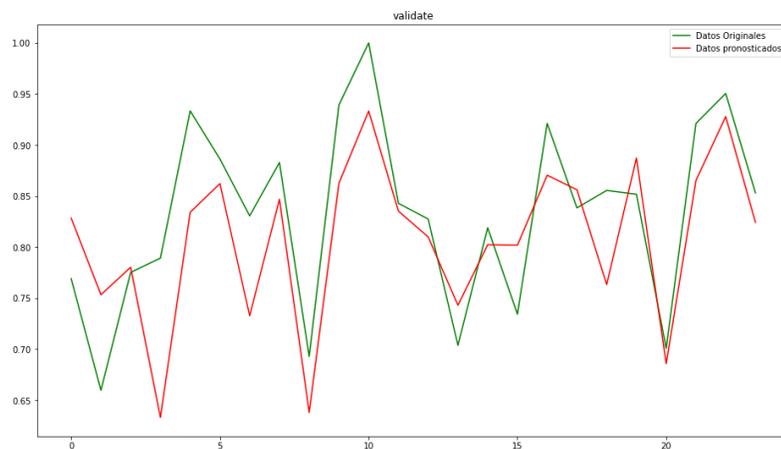
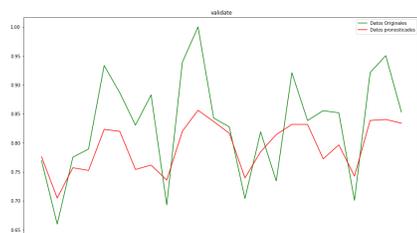
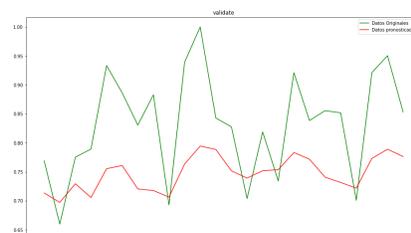


Figura 4.19: Comparación de información.

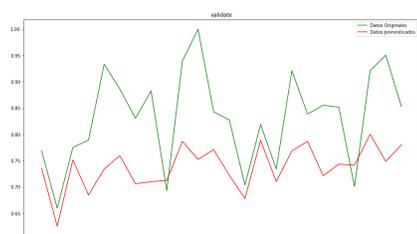
Adicional al modelo anterior, 4 modelos mas fueron considerados como posibles pronósticos, sin embargo fue la Figura 4.19 la mas adecuada. A continuación la Figura 4.20 se muestran los 4 modelos considerados.



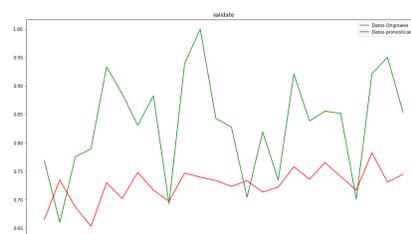
(a) Modelo de entrenamiento (A).



(b) Modelo de entrenamiento (B).



(c) Modelo de entrenamiento (C).



(d) Modelo de entrenamiento (D).

Figura 4.20: Resultados de otros modelos de entrenamiento.

A pesar de que el modelo no se adapta de la mejor manera a los datos originales, este modelo se dará por bueno ya que fue el que tiene mayor tasa de aprendizaje.

Ahora que la Red Neuronal esta dada por buena, se realiza una nueva predicción.

4.3.5. Pronósticos del IGAE

Finalmente, realizado lo anterior y con la Red Neuronal propuesta, se pronostica el IGAE teniendo como referencia los datos del IGAE hasta Diciembre de 2018, con el propósito de pronosticar 12 datos del IGAE y compararlos con los datos reales obtenidos por el INEGI, para así poder observar la eficacia de nuestra Red Neuronal, mediante los errores totales y relativos.

Los resultados obtenidos se presentan en la Tabla 5.3.

Tabla 4.15: Pronóstico Enero a Diciembre del 2019.

Periodo	X	\hat{X}	Error	Relativo
01/2019	112.72	112.02	0.70	0.01
02/2019	109.10	110.94	-1.84	-0.02
03/2019	112.47	112.04	0.43	0.00
04/2019	109.99	110.57	-0.58	-0.01
05/2019	115.46	115.91	-0.45	0.00
06/2019	113.04	113.68	-0.64	-0.01
07/2019	113.54	113.99	-0.45	0.00
08/2019	113.43	113.55	-0.12	0.00
09/2019	109.01	110.13	-1.12	-0.01
10/2019	115.46	114.91	0.55	0.00
11/2019	116.32	115.23	1.09	0.01
12/2019	113.47	113.04	0.43	0.00

En la Figura 4.21 se muestra las gráficas del IGAE en comparación al pronóstico obtenido mediante la Red Neuronal Artificial.

Ahora bien, se obtienen los pronósticos de interés, es decir el IGAE de Enero del 2020 a Diciembre de 2020, bajo la misma metodología y el mismo procesado de datos que se realizó para el entrenamiento, escalado de valores sin incluir la columna de salida (t) puesto que es el valor a encontrar se obtiene la matriz que se muestra en la Tabla 5.2.

De la tabla anterior se toma sólo la última fila, pues es la que correspondería al último mes de 2019.

Luego se crea una función de auto rellenado del desplazamiento que se realiza para cada predicción. Entonces para Enero de 2020, se tiene el conjunto con los últimos 12 meses de 2020. Para pronosticar todo el año del 2020. Al aplicar la función que realiza el pronostico, los resultados se muestran en la Tabla 4.17.

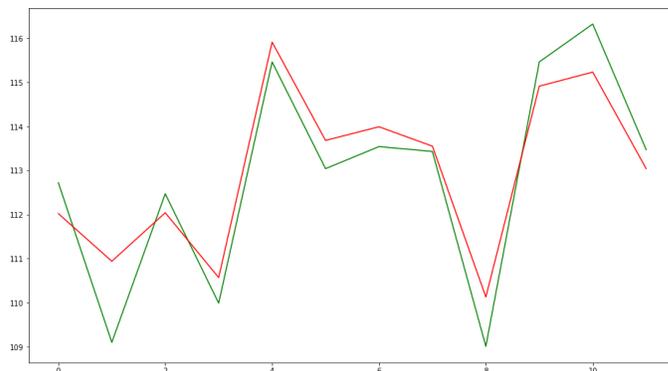


Figura 4.21: Comparación de los datos reales y pronosticados para el IGAE mensual del 2019.

Tabla 4.16: Matriz de entradas y salidas.

	$(t - 12)$	$(t - 11)$	$(t - 10)$...	$(t - 3)$	$(t - 2)$	$(t - 1)$
12	-0.357428	-1.000000	-0.321283	...	0.642570	1.000000	0.076305
13	-1.000000	-0.321283	-0.238955	...	1.000000	0.076305	-0.014055
14	-0.321283	-0.238955	0.608435	...	0.076305	-0.014055	-0.740963
15	-0.238955	0.608435	0.331327	...	-0.014055	-0.740963	-0.064255
16	0.608435	0.331327	0.004017	...	-0.740963	-0.064255	-0.562248
	⋮	⋮	⋮	...	⋮	⋮	⋮
21	0.642570	1.000000	0.076305	...	0.150604	0.128515	-0.759035
22	1.000000	0.076305	-0.014055	...	0.128515	-0.759035	0.536146
23	0.076305	-0.014055	-0.740963	...	-0.75903	0.536146	0.708836

Finalmente, en la Figura 4.22 se muestra el pronóstico arrojado por nuestra Red Neuronal Artificial

Fecha	Valor Pronosticado
01/2020	115.01
02/2020	114.20
03/2020	111.61
04/2020	112.71
05/2020	112.24
06/2020	115.00
07/2020	114.03
08/2020	113.94
09/2020	112.33
10/2020	109.79
11/2020	113.79
12/2020	114.38

Tabla 4.17: Datos obtenidos por la red neuronal para el pronóstico del año 2020.

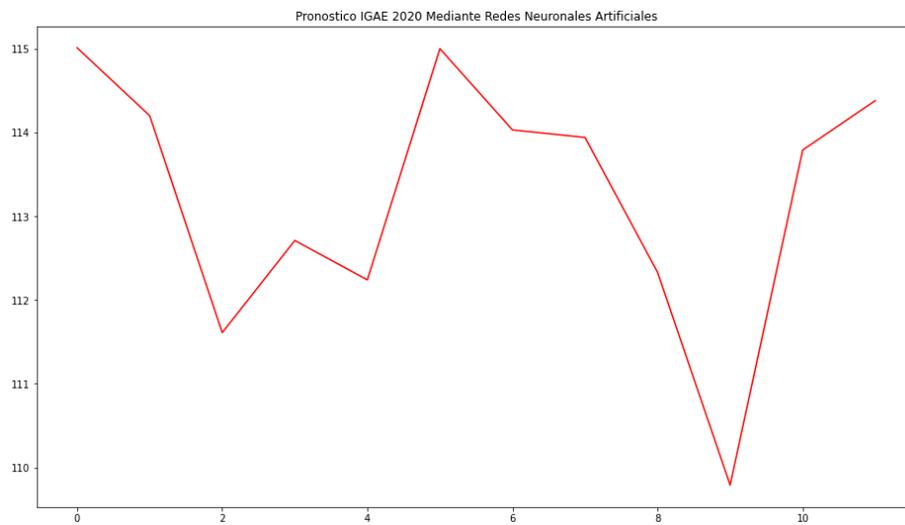


Figura 4.22: Pronóstico obtenido a través de redes Neuronales IGEA 2020.

Capítulo 5

Análisis y conclusiones

El propósito del trabajo expuesto en este documento es explicar y aplicar la teoría de series de tiempo, tratando de analizar la economía mexicana a través de un indicador siguiendo la metodología de Box-Jenkins y Redes Neuronales Artificiales, para así compáralas y tratar de concluir que metodología es más conveniente para este tipo de análisis.

En particular el análisis del crecimiento económico, a través de indicadores representativos y reportados en intervalos de tiempo tales como semestres o meses. Un ejemplo claro de estos indicadores son el PIB y el IGAE, siendo el PIB el indicador más acertado para medir el crecimiento económico, ya que representa una medida amplia de la actividad económica, sin embargo los datos son publicados trimestralmente y se reportan dos meses después de terminar dicho trimestre, lo cual limita bastante el análisis del movimiento económico en periodos mensuales. Por tal motivo, es importante contar con estimaciones en períodos más cortos, ya que actualmente el flujo constante de información en forma más frecuente es cada vez más importante para conocer el estado de la economía, esto lleva a elegir al IGAE como el indicador sobre el cual se realizan los análisis.

Una vez realizado el análisis, se decide realizar el pronóstico mediante un modelo *SARIMA* sin intervención, comparando los modelos que resultan más adecuados en ambos casos para, elegir al mejor.

Así al aplicar la metodología Box-Jenkis aplicada al IGAE, en un principio se realiza un prediferenciación para disminuir la variabilidad, logrando una mayor disminución en la heterogeneidad de la varianza al usar raíces cuartas.

Al llevar a cabo dicho análisis de la serie obtenida mediante la transformación, se concluye que dicha transformación no es estacionaria, por lo que es necesario realizar una diferencia. Se observa mediante el análisis de las FAM y FAMP de la primer diferencia, que la serie es estacionaria en el nivel no estacional, más aun, se nota de forma evidente la estacionalidad mensual, lo que es común observar en ciertos indicadores, ya que existe fuerte relación de estas actividades con los ciclos marcados en el calendario.

En seguida se analiza la serie a nivel estacional realizando una diferencia, concluyéndose la estacionariedad en este nivel, lo cual se confirma mediante la prueba

Dickey-Fuller, además de probar la eficiencia de los posibles modelos con estacionalidad de 12 meses con la prueba Ljung-Box.

Con ayuda del Criterio de Akaike se determina al mejor modelo con estacionalidad de 12 meses. De tal forma que el modelo propuesto resulta en un,

$$\mathbf{SARIMA(4, 0, 1)(2, 1, 2)_{12}}$$

el cual es adecuado para pronosticar los datos según la prueba de Ljung-Box y el análisis de residuales.

Una vez obtenido el mejor modelo para pronosticar el IGAE bajo la metodología Box-Jenkins se procede a la creación de un Red Neuronal Artificial para el pronóstico del IGAE y una vez obtenidos los resultados compararlos con los resultados obtenidos bajo la metodología explicada anteriormente y así tratar de determinar cual es el mejor modelo o cual modelo pronostica de la mejor manera.

El aprendizaje automático es una forma de inteligencia artificial, dicho aprendizaje es una categoría de la inteligencia artificial y automatiza eficientemente el proceso de creación de modelos analíticos y permite que las máquinas se adapten a nuevas situaciones de manera independiente. En la actualidad el uso de redes neuronales artificiales, son modelos matemáticos, que buscan replicar el comportamiento de las neuronas en la naturaleza, organizando su estructura de manera que emulen al cerebro. El objetivo de éstas redes es lograr dar lugar a un sistema inteligente que logre realizar con éxito tareas complejas. Para el análisis de este caso de estudio se creó un modelo que fuera capaz de replicar el comportamiento del IGAE a través de su información y entrenamiento para poder así realizar un pronóstico para los meses futuros.

Para realizar dicha Red Neuronal Artificial se realizó un pequeño análisis de información a través de gráficas para así observar su comportamiento ya que recordemos que para este entrenamiento no toma en cuenta que sea una serie de tiempo ya que para su entrenamiento solo se toma en cuenta el comportamiento de la información.

Una vez analizado el comportamiento de los datos se procede al escalamiento de la información para que a la Red le cueste menos trabajo aprender al momento de su entrenamiento.

Una vez escalada la información se procede a su subdivisión creando una matriz de entrenamiento y a la creación del modelo tentativo obteniendo que el modelo con mejor adaptación queda en:

- Entradas de vectores de orden 12x1.
- 1 capa oculta con 12 neuronas.
- La salida será de solo 1 neurona.
- Como función de activación se utilizó la **Tangente hiperbólica** puesto que los valores se encuentran entre -1 y 1.
- Como optimizador se usó el **Método Adam**.
- La métrica de pérdida fue el **Error Absoluto Medio**.
- Como la predicción será un valor continuo y no discreto, para calcular la precisión se utilizó **Error Cuadrático Medio**.
- Y finalmente para su entrenamiento se utilizaron 100, 000 iteraciones.

Una vez creada la Red Neuronal Artificial con la información descrita anteriormente se realizó la comparación con datos pasados y llegando a la conclusión que dicha neurona es la más adecuada para pronosticar el IGAE.

Es importante mencionar que en el presente trabajo, en ambas metodologías se realizó el pronóstico para el año 2020 ya que por la intervención de una pandemia que afectó a dicho año y que hasta la fecha sigue, así, las conclusiones pueden que no sean acertadas pero queda por analizar esta intervención, la cual puede incorporarse a los modelos anteriormente presentados y así arrojar mejores resultados, al mismo tiempo realizar otro tipo alternativo de análisis como la metodología de wavelets.

5.0.1. Análisis y Resultados

En este trabajo se diseñaron dos modelos para tratar de pronosticar el IGAE en ambas metodologías se obtuvieron dos posibles pronósticos uno para el año 2019 y el otro para 2020.

A continuación se presenta la información obtenida y su comparación con la información original al menos para el año 2019.

Tabla 5.1: Pronósticos de Enero a Diciembre del 2019, usando la metodología de Box-Jenkins y Redes Neuronales.

Resultados obtenidos 2019			
Periodo	Box-Jenkins	Red Neuronal	Diferencia
01/2019	110.82	112.02	-1.19
02/2019	108.44	110.94	-2.50
03/2019	111.28	112.04	-0.76
04/2019	111.84	110.57	1.27
05/2019	115.02	115.91	-0.89
06/2019	114.31	113.68	0.63
07/2019	113.19	113.99	-0.80
08/2019	113.51	113.55	-0.04
09/2019	109.79	110.13	-0.34
10/2019	115.91	114.91	1.00
11/2019	117.69	115.23	2.46
12/2019	115.19	113.04	2.15

A continuación en la Figura 5.1 se muestra el comportamiento de ambos pronósticos.

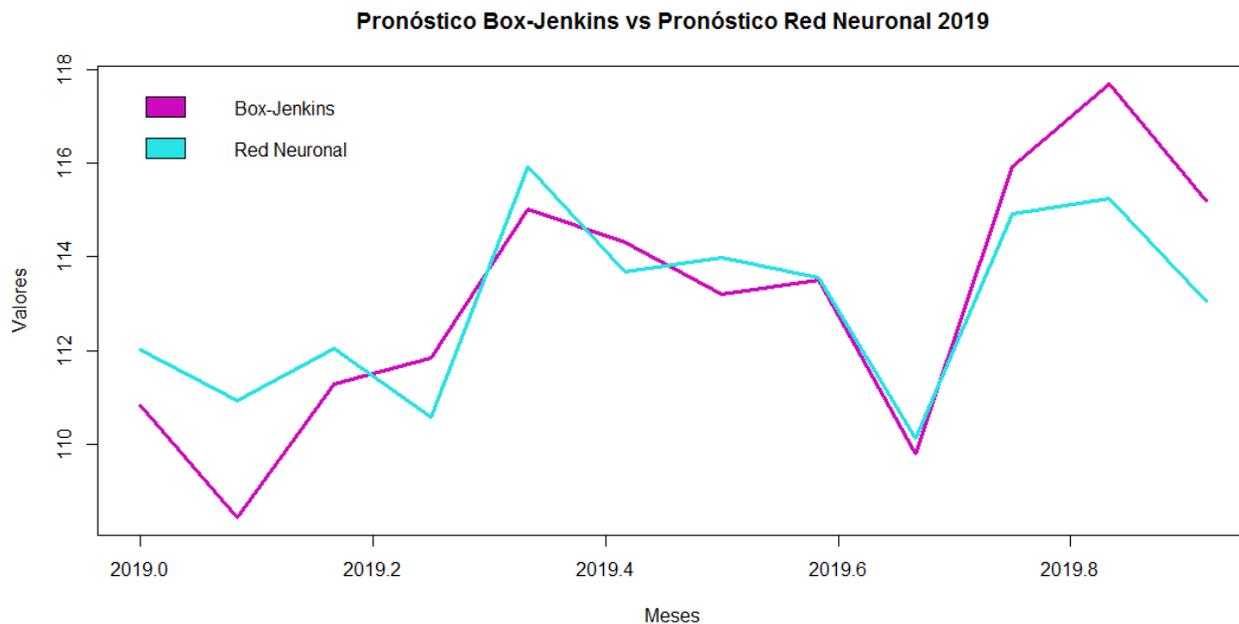


Figura 5.1: Comparación de los pronósticos de las metodologías Box-Jenkins y Redes Neuronales para 2019.

Se observa que ambas gráficas tienen un comportamiento parecido con un punto máximo en el mes de Mayo mostrándolo un poco más definido la Red Neuronal, para después tener una caída en el mes de Septiembre y repunte en el mes de Noviembre.

Al parecer ambos pronósticos cuentan con un comportamiento semejante con variabilidad no tan alta.

Ahora se procede a realizar la comparación de ambos pronósticos con los datos originales obtenidos por el INEGI.

En la Tabla 5.2 se muestran los pronósticos obtenidos y la información arrojada por el INEGI para el IGAE.

Tabla 5.2: Información arrojada por el INEGI y resultados obtenidos de la metodología Box-Jenkins y Redes Neuronales Artificiales para el año 2019.

Resultados obtenidos 2019			
Periodo	INEGI	Box-jenkins	Red Neuronal
01/2019	112.72	110.82	112.02
02/2019	109.10	108.44	110.94
03/2019	112.47	111.28	112.04
04/2019	110.99	111.84	110.57
05/2019	115.46	115.02	115.91
06/2019	113.04	114.31	113.68
07/2019	113.54	113.19	113.99
08/2019	113.43	113.51	113.55
09/2019	110.01	109.79	110.13
10/2019	115.46	115.91	114.91
11/2019	116.32	117.69	115.23
12/2019	113.47	115.19	113.04

En la Figura 5.2 Se muestra el comportamiento de los resultados obtenidos en ambas metodologías comparados con el comportamiento de los datos originales arrojados por el INEGI.

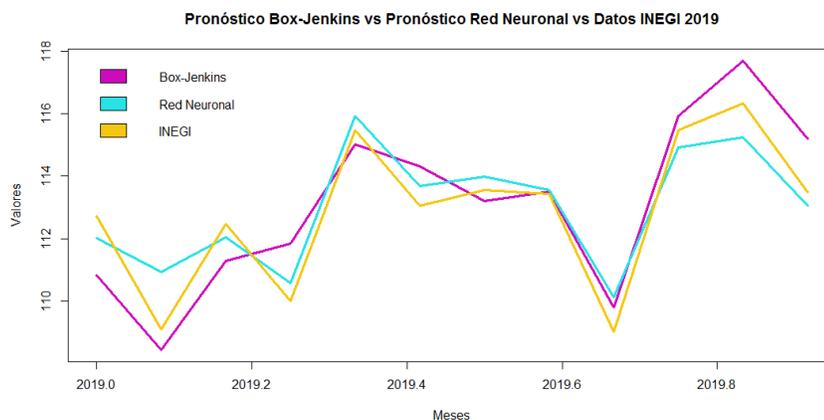


Figura 5.2: Comparación de resultados obtenidos con resultados INEGI 2019.

En la Figura 5.2 se puede apreciar que el comportamiento de ambas metodologías comparados con los datos reales tienen un comportamiento parecido acercándose más a la realidad el pronóstico de la Red Neuronal.

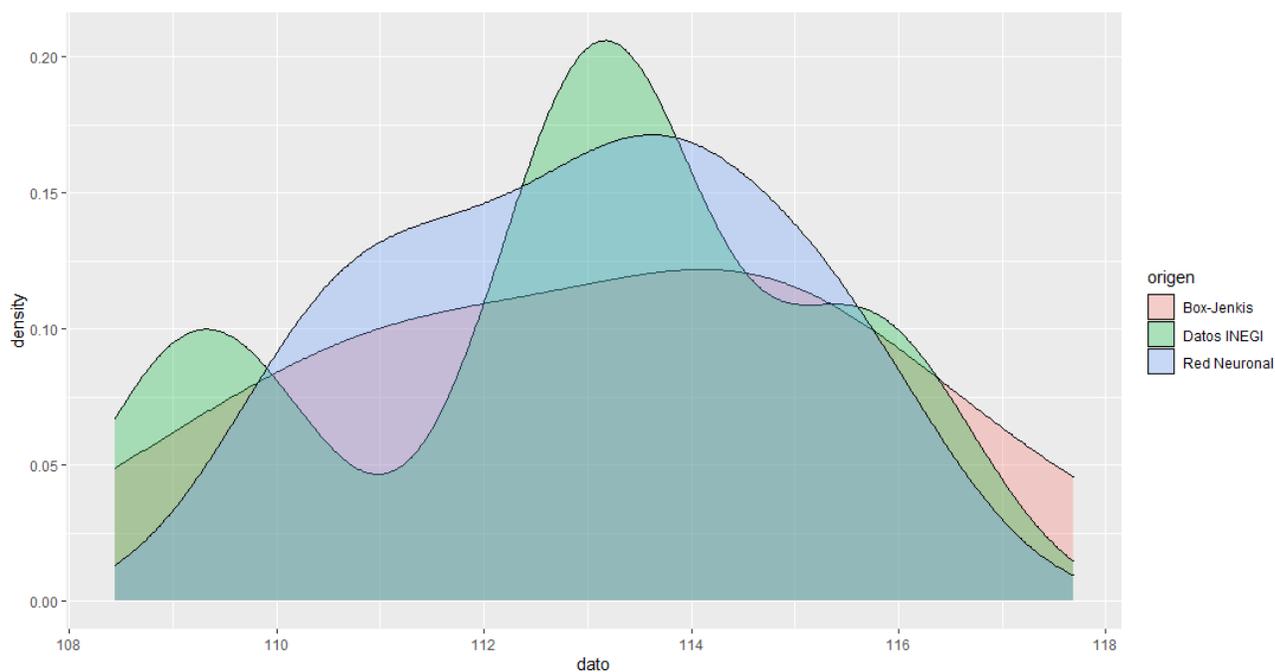


Figura 5.3: Diagramas de densidad generadas por el método Kernel.

En la Figura 5.3 se muestra el diagrama de densidad y se puede observar normalidad en las tres figuras, sin embargo se puede decir que la metodología que se asemeja más a la densidad de los datos reales es la de la Red Neuronal Artificial, por lo que podemos tener como primera conclusión que para el caso estudio del pronóstico del Indicador Global de la Actividad Económica en México el mejor pronosticador es el Modelo de Red Neuronal Artificial ya que presenta un comportamiento más semejante a la realidad.

Finalmente procedemos a comparar los pronósticos de interés obtenidos por ambas metodologías y los resultados se muestran a continuación.

Tabla 5.3: Comparación de resultados obtenidos por el pronóstico de Enero a Diciembre del 2020.

Pronósticos obtenidos 2020					
Periodo	INEGI	Box-Jenkins	Redes Neuronales	Error Relativo Box-Jenkins	Error Relativo Redes Neuronales
01/2020	111.92	110.87	115.01	0.01	0.03
02/2020	108.40	107.59	114.20	0.01	0.05
03/2020	108.91	112.14	111.61	0.02	0.02
04/2020	88.00	109.73	112.71	0.20	0.28
05/2020	89.47	114.31	112.24	0.22	0.25
06/2020	98.01	113.43	115.00	0.14	0.17
07/2020	102.22	112.51	114.03	0.09	0.12
08/2020	102.93	112.83	113.94	0.09	0.11
09/2020	103.13	108.97	112.33	0.05	0.09
10/2020	109.33	115.04	109.79	0.05	0.00
11/2020	111.49	116.50	113.79	0.04	0.02
12/2020	110.43	114.44	114.38	0.04	0.04

En la Tabla anterior se puede observar que las diferencias ahora son mayores y el comportamiento del modelo $SARIMA(4, 0, 1)(2, 1, 2)_{12}$ sigue teniendo un comportamiento más estable, mientras que la Red Neuronal a pesar de tener un buen entrenamiento al momento de pronosticar los resultados obtenidos están un poco elevados pero ambos conservan un comportamiento similar.

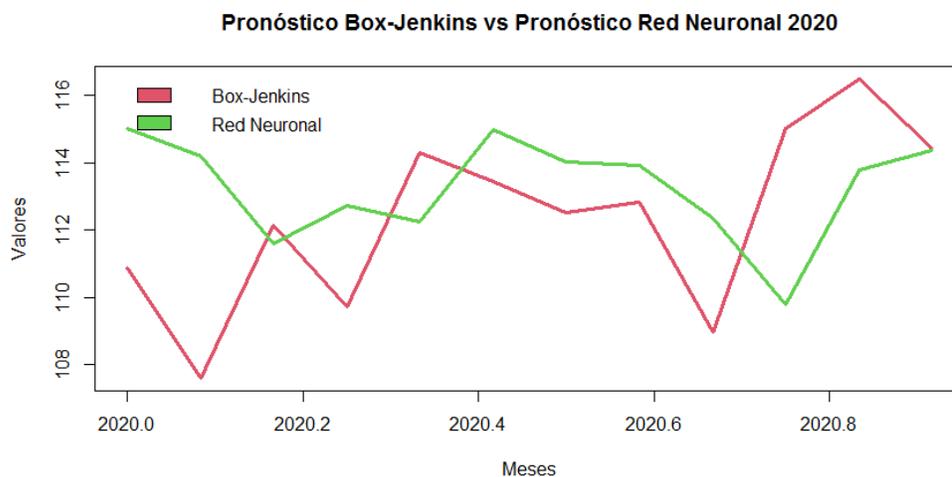


Figura 5.4: Gráficos de pronósticos obtenidos a través de la metodología Box-Jenkins y Redes Neuronales para el año 2020.

Para concluir este trabajo compararemos los resultados de los pronósticos obtenidos en el año 2020 con los resultados preliminares arrojados por el INEGI de ese año. Recordemos que a finales de 2019 el mundo fue atacado por una enfermedad llamada COVID-19 que se extendió en todo mundo y que atacó de manera muy fuerte a la economía del país, sin embargo aun no se sabe con certeza cuanto influyó esta enfermedad al IGAE por lo que el INEGI arrojó información preliminar de dicho año pero aun no es la información verdadera.

A continuación en la Figura 5.5 se muestran la comparación de los pronósticos obtenidos contra la información preliminar del INEGI.

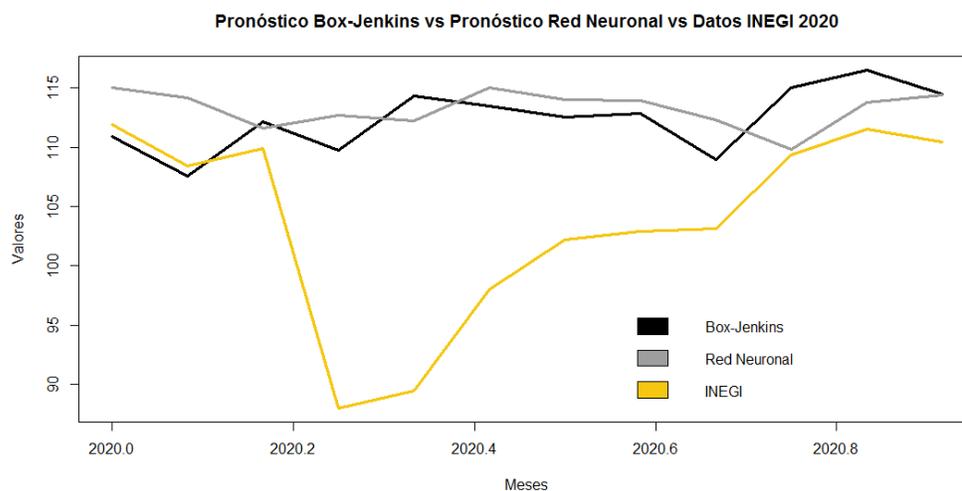


Figura 5.5: Comparación de ambas metodologías con datos preliminares 2020.

Como se puede observar en la información arrojada por el INEGI se tiene una caída justo por el mes de Marzo, mes en el que ataco de manera muy fuerte dicha enfermedad a México, como se mencionó anteriormente en este trabajo, queda por analizar los datos y obtener pronósticos usando un modelo de intervención, para incorporarse a los modelos ya descritos para así poder obtener mejores resultados.

5.0.2. Conclusiones.

Como conclusión a lo ya expuesto anteriormente, se puede decir que ambas metodologías son buenas ya que generan un pronóstico bastante adecuado a la realidad, sin embargo no se puede dejar de lado la metodología que hay detrás de cada una de estas, ya que la metodología Box-Jenkins cuenta con respaldo más sólido desde el punto de vista matemático y las Redes Neuronales Artificiales son aún un método un tanto descriptivo ya que para poder llegar a un pronóstico adecuado, se tuvieron que hacer varios modelos para así determinar el mejor.

Es importante mencionar que el presente trabajo queda abierto para un posible análisis complementario que ayude a mejorar los pronósticos vistos anteriormente, esto es, podrían incorporarse modelos de intervención completos y así obtener resultados más precisos y al mismo tiempo no descartar la idea de realizar otro tipo de análisis alternativo para complementarlo de la mejor manera.

Referencias

- [1] Brockwell J. P., Davis A. R. Introduction to Time series and Forecasting. Springer. Second Edition. ISBN: 978-0-387-95351-9. 2002.
- [2] Guerrero V. Modelos estadísticos para series de tiempo univariadas. V Coloquio del departamento de matemáticas. Centro de Investigación y Estudios Avanzados del IPN. 1987.
- [3] Cesar A. Hernández S. Luis Fernando Pedraza M. Andrés Escobar Diaz. Aplicaciones de las series de tiempo en modelos de tráfico para una red de datos. Scientia et Technica Año XIV, No 38, Junio de 2008. Universidad Tecnológica de Pereira. ISSN 0122-1701
- [4] Robert H. Shumway, David S. Stoer. Time Series Analysis and Its Applications, Third Edition. ISBN: 978-1-4419-7864-6. 2010..
- [5] William W.S. Wei. Time Series Analysis Univariate and Multivariate Methods, Pearson, Second Edition. ISBN: 0-321-32216-9. 2006.
- [6] Gisela Tlacuatl Sánchez. Tesis sobre Análisis y pronóstico del crecimiento económico mediante el Indicador Global de la Actividad Económica (IGAE). Benemérita Universidad Autónoma de Puebla 2016.
- [7] Liliana Hernández Izquierdo. Tesis sobre Pronóstico del número de accidentes de tránsito de la zona urbana de Puebla para empresas aseguradoras de automóviles. Benemérita Universidad Autónoma de Puebla 2016.
- [8] Bruce L. Bowerman, Richard T. Oconnell. Pronósticos, Series de Tiempo y Regresión: Un Enfoque Aplicado, ISBN: 9789706866066. 2007.
- [9] Jonathan D. Cryer Kung-Sik Chan. Time Series Analysis With Applications in R, Second Edition ISBN 978-0-387-75959-3, 2008.
- [10] Instituto Nacional de Estadística y Geografía. Sistema de Cuentas Nacionales de México, 2013.
- [11] INEGI. (25/03/2021). Indicador Global de la Actividad Económica. 28/03/2021, INEGI Sitio web: <https://www.inegi.org.mx/programas/igae/2013/>

- [12] Hastie, T., Tibshirani, R., Friedman, J. (2009). The elements of statistical learning (2.a ed.). Springer.
- [13] Kuhn, M., Johnson, K. (2013). Applied predictive modeling (1.a ed.). Springer.
- [14] Patterson, J., Gibson, A. (2017). Deep learning a practitioner´s approach (1.a ed.). OREILLY.
- [15] Daniela Urquizo Rojas, Fabián Mendivil Gomez. Aplicación de Redes Neuronales Artificiales para el Análisis de la Inflación en Bolivia. Encuentro economista de Bolivia 2011.
- [16] Matich, D. J. (2001). Redes Neuronales: Conceptos Básicos y Aplicaciones. (1.a ed.). Desconocido.
- [17] Monroy, J., Ramírez, A., Erika, L. (2016). Aspectos relevantes para mejorar el desempeño del algoritmo Backpropagation. Revista de Sistemas Computacionales y TIC´s, 1(1).
- [18] Na8. (2019). Pronóstico de Series Temporales con Redes Neuronales en Python. 2022, de Aprende Machine Learning Sitio web: <https://www.aprendemachinelearning.com/pronostico-de-series-temporales-con-redes-neuronales-en-python/>