RESUMEN

1

Resumen de comentarios: 20210522.1_Borrador_Tesis_OCM.pdf

Página:8

Número: 1 Autor: oscar Asunto: Nota adhesiva Fecha: 2021-05-15 12:25:31

Para la próxima ya debe de poner el resumen.

1 INTRODUCCIÓN

1.1 Antecedentes

Estimar los valores futuros en un determinado contexto ha producido un aumento en el análisis de los datos referidos en el tiempo, conocido también como series cronológicas. Este tipo de datos se encuentra en diferentes áreas, tanto en investigación académica como en el análisis de datos para la toma de decisiones. En el campo financiero es común hablar de la devaluación del colón con respecto al dólar, cantidad de exportaciones mensuales de un determinado producto o las ventas, entre otros (Hernández, 2011a). Las series cronológicas son particularmente importantes en la investigación de mercados o en las proyecciones demográficas; de manera conjunta apoyan la toma de decisiones para la aprobación presupuestaria en distintas áreas.

En la actualidad, la información temporal es muy relevante: El Banco Mundial[1] cuenta en su sitio web con datos para el análisis de series cronológicas de indicadores de desarrollo, capacidad estadística, indicadores educativos, estadísticas de género, nutrición y población. Kaggle[2], uno de los sitios más populares relacionados con el análisis de información, ofrece una gran cantidad de datos temporales para realizar competencias relacionadas con las series temporales y determinar los modelos ganadores para una determinada temática[3].

Asimismo, los pronósticos (estimación futura de una partícula en una serie temporal) son utilizados por instituciones públicas o del sector privado, centros nacionales o regionales de investigación y organizaciones no gubernamentales dedicadas al desarrollo social. Si las entidades previamente mencionadas cuentan con proyecciones de calidad, la puesta en marcha de sus respectivos planes tendrá un impacto más efectivo.

Los métodos existentes para llevar a cabo un análisis de series cronológicas son diversos, y responden al propio contexto y tipo de datos. Obtener buenos pronósticos o explicar el comportamiento de un fenómeno en el tiempo, siempre será un tema recurrente de investigación. Generar una adecuada estimación es fundamental para obtener un pronóstico de confianza, 2 además resulta importante 1 mencionar que los modelos ARIMA tienen como objetivo explicar las relaciones pasadas de la serie cronológica, para de esta manera conocer el posible comportamiento futuro de la misma (R. J. Hyndman & Athanasopoulos, 2018a).

Al trabajar con la metodología de Box-Jenkins, uno de las etapas a concretar es identificar los parámetros de estimación que gobiernan la serie temporal. Para indagar los términos en el proceso de investigación se ha utilizado 3 la identificación de parámetros mediante autocorrelogramas par-

[1] https://databank.worldbank.org/home.aspx
[2] Se trata de una subsidiaria de la compañía Google que sirve de centro de reunión para todos aquellos interesados en la ciencia de datos.
[3] Muchas de ellas incluyen recompensas económicas que van desde los \$500 hasta los \$100,000 para aquellos que logren obtener los mejor pronósticos.

Página:10

Número: 1 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 16:40:15

Esto ya lo habíamos discutido, pero acá la idea no es muy clara.

Faltaría un agregado o transición para poder hablar correctamente de los modelos ARIMA.

Número: 2 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 16:38:03

Poner "punto". Es importante resaltar que las técnicas de proyección ARIMA...

Número: 3 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 16:41:02

se suele utilizar

ciales y totales. Sin embargo, los autocorrelogramas formados no analizan de forma exhaustiva y óptima los posibles coeficientes que podrían contemplarse la ecuación de Wold. Según su definición matemática, esta posee infinitos coeficientes, por tanto, se debe buscar una alternativa distinta, que opte por aproximar de una mejor manera la identificación de los parámetros estimados, cubriendo un mayor número de posibilidades. Esto se podría obtener mediante un método analítico de sobreparametrización.

1.2 La problemática

La dificultad visual a la hora de identificar un modelo ARIMA radica en que los autocorrelogramas solo aportan una aproximación al proceso que gobierna la serie. De forma complementaria, es común caer en el problema de la subjetividad, pues a pesar de que alguien proponga un patrón que gobierne la serie, otro analista podría tener una interpretación visual diferente del mismo proceso, proponiendo así distintas identificaciones para un mismo proceso. Además, se posee el inconveniente de que algunos métodos de identificación automática del proceso que gobierna la serie subestiman el número de parámetros que se debería de contemplar.

Alternativas como la función `auto.arima()`, que ofrece el paquete `forecast` del lenguaje de programación R[4] (R. Hyndman & Khandakar, 2008), permite estimar un modelo ARIMA basado en pruebas de raíz unitaria y minimización del AICc (Burnham & Anderson, 2007). Así se obtiene un modelo temporal definiendo las diferenciaciones requeridas en la parte estacional d mediante las pruebas KPSS (Xiao, 2001) o ADF (Fuller, 1995), y la no estacional D utilizando las pruebas OCSB (Osborn, Chui, Smith, & Birchenhall, 2009) o la Canova-Hansen (Canova & Hansen, 1995), seleccionado el orden óptimo para los términos $ARIMA(p,d,q)(P,D,Q)_s$ para una serie cronológica determinada.

Sin embargo, estas pruebas suelen ignorar diversos términos que bien podrían ofrecer mejores pronósticos; no someten a prueba las posibles especificaciones de un modelo en un rango determinado, sino que realizan aproximaciones analíticas para definir el proceso que gobierna la serie cronológica, dejando así un vacío en el cual se corre el riesgo de no seleccionar un modelo que ofrezca mejores pronósticos. Poner a prueba un mayor número de posibilidades para la especificación de los modelos tiene la ventaja descartar ciertos modelos, y mantener otros con un criterio más científico y una evidencia numérica que despalde esa decisión.

1.3 Objetivos del estudio

El objetivo general de la presente investigación es proponer un algoritmo alternativo más exhaustivo para la selección de modelos ARIMA mediante la sobreparametrización de los términos de la ecuación del ARIMA.

[4]Descarga gratuita en https://cran.r-project.org/

Página:11

Número: 1 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 16:43:16

y la aproximación mediante los auto correlogramas no es una forma exacta de aproximar la serie que gobierno el proceso. Por lo tanto, se debe buscar....

Número: 2 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 16:44:48

Más adecuado decir:

Un alternativa al problema de aproximar los parámetros que Gobiernan la serie podría ser la sobre parametrización...

Número: 3 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 16:47:10

Sería mejor decir "Problemática..." quitar el La... ud que cree?

Número: 4 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 16:47:10

Número: 5 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 16:49:52

redacción "ventaja de descartar ciertos"...

1.5 Organización del estudio

El presente trabajo de investigación consta de cinco capítulos. El primer ofreció una contextualización del uso de las series de tiempo, así como la importancia de poder contar con pronósticos de calidad. Se presentó el objetivo del estudio, así como una breve descripción de la metodología empleada en la aplicación de series temporales, y cómo se planea modificar el método de estimación en los modelos ARIMA. Se concluye esta sección con hechos que justifican la importancia de esta investigación.

El siguiente capítulo consiste en el marco teórico, abarcando aspectos fundamentales 1 de la ecuación de Wold, la metodología Box-Jenkins, la selección de los procesos que gobiernan la serie, 2 la descripción del proceso iterativo, el análisis combinatorio que aborda los escenarios de estimación, 3 entre otros.

El tercer capítulo describe la metodología relacionada al estudio. Se inicia con una descripción global de los conceptos más fundamentales del análisis de series cronológicas, pasando por los componentes fundamentales de las mismas. Se discuten también los supuestos clásicos del análisis de series cronológicas, los distintos tipos de modelos, el análisis de intervención, los métodos de validación y las medidas de rendimiento; aspectos cruciales para obtener un modelo ARIMA vía sobreparametrización. La sección metodológica culmina con la descripción del proceso de simulación 4 que se utilizará, así como la discusión del método propuesto.

El capítulo cuatro consiste en la presentación de los resultados, tanto en los datos simulados como en la aplicación a datos costarricenses y se contrastarán contra los obtenidos por otros métodos como el de la función `auto.arima()`, 5 entre otros.

El último capítulo 6 busca discutir los principales resultados, así como señalar las conclusiones más importantes y ofrecer algunas recomendaciones que orienten futuros estudios relacionados.

Página:13

Número: 1 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 16:53:59

..."aspector fundamentales de las series temporales: ecuación de Wold, métodología de Box-Jenkins, etc...."

Número: 2 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 16:54:42

por esto hablamos de los autocorrelogramas y los procesos de estimación?

Número: 3 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 16:55:18

la selección por combinación, multiplicación y permutación....

Número: 4 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 16:58:31

me parece bien ,pero me hace falta ver la antes, la aplicación de las permutaciones en la selección de parámetros o estimadores de una ecuación ARIMA,

Número: 5 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 16:59:30

recordar que son 3:

-auto.arima()
-un ARIMA (1,1,1)(1,1,1)
-auto.cesar()

Número: 6 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 17:00:12

El capítulo de conclusión - discusión .... etc

2 MARCO TEÓRICO

Los modelos de series cronológicas han sido un importante tema de investigación durante décadas (De Gooijer & Hyndman, 2006). Su objetivo principal consiste en obtener simplificaciones de la realidad mediante el ajuste de diversos modelos, los cuales se ajustan a datos recolectados a lo largo del tiempo de forma regular. Estos modelos son luego utilizados para generar pronósticos sobre el comportamiento futuro del fenómeno de interés.

Sin embargo, encontrar un modelo que presente un buen comportamiento con respecto a los datos no es tarea fácil, pues deben considerarse diversos aspectos teóricos para obtener un modelo adecuado que logre generar pronósticos realistas y pertinentes para la toma de decisiones (Rezaee, Aliabadi, Dorestani, & Rezaee, 2020).

Una serie temporal se define como una secuencia de datos observados, cuyas mediciones ocurren de manera sucesiva durante un periodo de tiempo. Los registros de estos datos pueden referirse a una única variable en cuyo caso de dice que es una serie univariada; o bien, pueden registrarse distintas variables para el mismo periodo de tiempo, conocida como serie temporal multivariada. Según Hipel & McLeod (1994), cada observación puede ser continua o discreta, como la temperatura de una ciudad durante el día o las variaciones diarias del precio de un activo financiero, respectivamente; las observaciones continuas, además, pueden ser convertidas a su vez en observaciones discretas.

El presente capítulo consta de seis apartados: El primer apartado abarca los cuatro componentes de una serie cronológica, siendo estos la tendencia y los componentes estacionales, cíclicos e irregulares. Posteriormente, la segunda sección repasa los supuestos fundamentales en el análisis de series cronológicas. Con los elementos más básicos introducidos, el tercer apartado cubre el eje central de esta investigación: Los modelos Autorregresivos Integrados de Medias Móviles y sus componentes, los modelos autorregresivos y los modelos de medias móviles, así como la metodología Box-Jenkins y el proceso para la identificación de los modelos. En el cuarto apartado se introducen los métodos para la identificación de los modelos. El quinto apartado abarca los componentes relacionados a los autocorrelogramas, la forma más difundida para la selección de modelos y, finalmente, el sexto apartado introduce el principal aporte de este estudio, la sobreparametrización como método para la selección de modelos.

2.1 Componentes de una serie cronológica

En el análisis de series cronológicas existen dos grandes corrientes de estudio: Los componentes inherentes a la serie cronológica y el estudio de las autocorrelaciones. De acuerdo con Hernández (2011a), las series cronológicas poseen cuatro componentes principales: Tendencia, Ciclos, Estacionalidad e Irregularidad. Considerando estos cuatro elementos, las series cronológicas pueden ser

Página:14

Número: 1 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 17:01:28
eliminar...

Número: 2 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 17:01:15
forma periódica.

Número: 3 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 17:03:45
teóricos, prácticos, y de la temática de estudio para así obtener....

Número: 4 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 17:03:08
no es propio. Mejor decir "no es sencillo" pero tarea fácil es más del lenguaje oral.

Número: 5 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 17:04:51
punto

Número: 6 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 17:05:26
eliminar, dado que luego se va a explicar.

Número: 7 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 17:07:02
me parece que no... me parece mejor como métodos de selección de unidades como la multiplicación, combinatorio y la sobre parametrízación... se deberían explicar 4, pero se me va e nombre ahorita...

Número: 8 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 17:20:48
César para este apartado le pediría que implemente por favor de forma gráfico lo siguiente:

-Grafo para lo que es tendencia, ciclo estacionalidad e irregularidad
-Grafo para una serie aditiva
- Grafo para una serie multiplicativa.

¿Estamos claro que los componentes es la forma descriptiva de ver o entender una serie temporal?

Número: 9 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 17:07:50
Según el primer enfoque, de acuerdo con ....

aditivas, como se muestra en la ecuación 1, en cuyo caso se asume que los cuatro componentes son independientes entre sí; o *multiplicativa*, donde, por el contrario, los cuatro componentes son independientes, como muestra la ecuación 2.

$$Y(t) = T(t) + S(t) + C(t) + I(t) \tag{1}$$

$$Y(t) = T(t) \times S(t) \times C(t) \times I(t) \tag{2}$$

Donde Y es la serie cronológica, T es la tendencia, S es la parte estacional, C el componente cíclico, I la parte irregular o aleatoria, y t es el momento en el tiempo. Cada una de sus partes se definen a continuación. 1

2.1.1 La tendencia

A partir del texto de Calderón (2012), la tendencia general de una serie cronológica se refiere al crecimiento, decrecimiento o lateralización de sus movimientos a lo largo del periodo de estudio. Una tendencia bastante marcada es la del comportamiento poblacional, que con el tiempo su crecimiento suele comportarse de una forma muy similar a una exponencial. 2

2.1.2 Componentes estacionales

Calderón (2012) también se refiere a los cambios estacionales que se presentan en una serie de tiempo, los cuales se relacionan con las fluctuaciones naturales del fenómeno dentro de una temporada de observaciones. Ejemplos comunes de esto son las condiciones climáticas, consumo de alimentos en fechas festivas, entre otros. 3

2.1.3 Componente cíclico

Del informe elaborado también por Calderón (2012) se desprende que los periodos cíclicos, por su parte, se refieren a los cambios que se dan en una serie cronológica en el mediano plazo, que son causados por determinados eventos que suelen repetirse. Estos ciclos suelen tener una duración determinada, como es el caso de los índice bursátil S&P 500. Este indicador resume el estado de las 500 empresas más importantes de Estados Unidos, y sus ciclos suelen presentar un auge, seguido por un descenso que, posteriormente, se vuelve una depresión, y que finalmente se convierte en una recuperación a su estado inicial. 4 5

2.1.4 Componente irregular

Finalmente, la irregularidad de una serie cronológica, siguiendo a Calderón (2012), se refiere a las fluctuaciones propias de un fenómeno que no pueden ser predichas. Estos cambios no se dan de 6

Página:15

Número: 1 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 17:12:16

bien, pero por fa introducir los gráficos anter mencionadas.

Además, falta decir de que una perspectiva clásica, esta es una forma dde analizar de forma descriptiva una serie de tiempo

Número: 2 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 17:13:13

Por favor poner una gráfica de lo que es una tendencia... y si pusiera una ecuación mate mediante despede de los modelos ADI y MULTI mejos

Número: 3 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 17:13:35

Por favor poner una gráfica de lo que es una estacional... y si pusiera una ecuación mate mediante despede de los modelos ADI y MULTI mejos

Número: 4 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 17:14:08

Por favor poner una gráfica de lo que es un ciclo ... conste que los ciclos se ven en periodos muchos más largos, no de un año....

y si pusiera una ecuación mate mediante despede de los modelos ADI y MULTI mejos

Número: 5 Autor: oscar Asunto: Nota adhesiva Fecha: 2021-05-15 17:14:43

Mediano - largo plazo

Número: 6 Autor: oscar Asunto: Resaltado Fecha: 2021-05-15 17:15:25

Por favor poner una gráfica de lo que es un compoente irregular ... y si pusiera una ecuación mate mediante despede de los modelos ADI y MULTI mejos manera regular, es decir, no siguen un patrón determinado.

2.2 Supuestos en el análisis de series cronológicas

El análisis de series temporales, según Hipel & McLeod (1994), representa un método para comprender la naturaleza de la serie en cuestión y poder utilizarla para generar pronósticos. Es en este sentido que entran en escena las observaciones recolectadas de la serie, pues ellas son analizadas y sujetas a modelados matemáticos que logren capturar el proceso que gobierna a toda la serie cronológica (Zhang, 2003). Los pronósticos se generan a partir de este modelo, es decir, pronosticar 1 el futuro, se utilizan las correlaciones con las observaciones pasadas.

En un proceso determinístico, es posible predecir con certeza lo que ocurrirá en el futuro; las series cronológicas, sin embargo, carecen de esta condición. El análisis de series cronológicas asume 2 que las observaciones pueden ajustarse a un determinado modelo estadístico, esto se conoce como un proceso estocástico. Es de esta manera que Hipel & McLeod (1994) sugieren que una serie cronológica puede considerarse como una muestra aleatoria de una serie mucho más grande. 3

Como una serie de tiempo puede considerarse como un proceso estocástico, éstas se encuentran sujetas a múltiples supuestos. El más fundamental de ellos es que todas las observaciones son independientes e idénticamente distribuidas (i.i.d.) siguiendo una distribución aproximadamente Normal, con una media y variancia dadas. Lo anterior es contrario al uso de las observaciones 4 pasadas para pronosticar el futuro, por lo que este supuesto, según Cochrane (1997), no es exacto pues una una serie de tiempo no es exactamente, i.i.d., sino que siguen un patrón medianamente regular en el largo plazo.

Otro concepto de interés en las series cronológicas es el de estacionaridad. De acuerdo con Agrawal 5 & Adhikari (2013), una serie se considera estacionaria cuando su nivel medio y su variancia son aproximadamente las mismas durante todo el periodo, es decir, el tiempo no afecta a estos estadísticos de variabilidad. Este supuesto busca simplificar la identificación del proceso estocástico con el objetivo de obtener un modelo adecuado para generar los pronósticos. Sin embargo, y de una manera similar al supuesto de i.i.d., si una serie cronológica posee tendencias o patrones estacionales hace que esta sea no estacionaria. En la práctica, una serie puede volverse estacionaria al aplicarle transformaciones o diferenciaciones de distinto orden.

El último supuesto, y quizá el que más debate genera, es el criterio de parsimonia. 6 Como mencionan Zhang (2003) y Hipel & McLeod (1994), este principio sugiere que se prioricen modelos sencillos, con pocos parámetros, para representar una serie de datos. Mientras más grande y complicado sea el modelo, mayor será el riesgo de sobre ajuste, lo que implica que el ajuste sea muy bueno en el conjunto de datos con que se generó el modelo, pero que los pronósticos generados sean pobres ante nuevos conjuntos de datos. Este problema, sin embargo, se presenta al considerar un único modelo

Página:16

Número: 1 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:16:46

me hace falta la ecuación de una ST

Número: 2 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:16:23

tal vez decir simplemente que una serie de tiempo es un proceso estocástico.

Número: 3 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:17:32

igual me hace falta matemática de por medio que muestre que es un proceso estocástico - probabilístico.

Número: 4 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:19:22

me gustaría ver esto del iid en términos de ecuaciones

Número: 5 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:20:45

hay estacionaridad fuerte y debil, me gustaría que explico, con ecuaciones de fondo, cual es cada uno, y además que explique la diferencia.

Número: 6 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:21:53

vieras que tal vez en un modelo lineal es algo que concordaría, pero entiendo el punto de estos autores.

Claro está, la ecuación de Wald hace ver lo contrario a la parsimonia .... y claro, tenemos que llegar a aproximar el Box Jenkins con algo...

con muchos parámetros; pero si se consideran varios modelos y estos son sometidos a distintos criterios, puede obtenerse un modelo sobreparametrizado que ofrezca buenos pronósticos.

2.3 Modelos Autorregresivos Integrados de Medias Móviles 1

Hay dos grandes grupos de modelos lineales de series cronológicas: Los modelos Autorregresivos 2 (AR) (Lee, s. f.) y los modelos de Medias Móviles (MA) (Box, Jenkins, & Reinsel, 1994). La combinación de estos dos grandes grupos forman los Modelos Autorregresivos de Medias Móviles (ARMA) (Hipel & McLeod, 1994) y los modelos Autorregresivos Integrados de Medias Móviles (ARIMA), siendo este último de particular interés en esta investigación.

Los modelos ARIMA son los de uso más extendido en el análisis de series cronológicas. Se fundamentan en las autocorrelaciones pasadas, y contempla un proceso iterativo para identificar un posible proceso óptimo a partir de una clase general de modelos. El teorema de Wold (Surhone, Timpledon, & Marseken, 2010) sugiere que todo proceso estacionario puede ser determinado de una forma específica y cuya ecuación posee, en realidad, infinitos coeficientes, pero que debe ser reducido a una cantidad finita para luego evaluar su ajuste sometiéndolo a diferentes pruebas y medidas de rendimiento.

2.3.1 Ecuación de Wold

Según Sargent (1979), cualquier proceso estacionario puede ser representado mediante la ecuación 3:

$$x_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + \kappa_t \tag{3}$$

donde $\forall \psi_j \in \mathbb{R}, \psi_0 = 1, \sum_{j=0}^{\infty} \psi_j^2 < \infty$, y ε_t representa un ruido blanco i.i.d., es decir, $\varepsilon_t \sim N(0, \sigma^2)$; además, κ_t es el componente lineal determinístico de forma tal que $cov(\kappa_t, \varepsilon_{t-j} = 0)$, lo cual implica que este componente determinístico es independiente de la suma infinita de los choques pasados.

De lo anterior, si se omite la parte determinística κ_t de 3, el remanente es la suma ponderada infinita, lo cual implica que si se conocen los ponderadores ψ_j, y si además se conoce σ_ε^2, es posible obtener una representación para cualquier proceso estacionario; este concepto es conocido como *media móvil infinita.*

Sabiendo que $\varepsilon_t \sim N(0, \sigma^2)$, se tiene que ε_t tiene media 0, es decir, está centrado en este valor. De esta manera el ruido blanco es por definición un proceso centrado, lo cual implica que la suma ponderada infinita está centrada en sí misma. De esta manera, la representación de Wold de un proceso x_t supone que se suman los choques pasados más un componente determinístico que no es otro que el valor esperado del proceso: $\kappa_t = m$, donde m es una constante cualquiera. Así, la

Página:17

Número: 1 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:33:45

No está bien estructurado el 2.3

Ud debe hablar de

-Qué es una serie de tiempo
- Qué la hace un proceso estocástico
- De los supuestos de esta iid y estacionaro
- Su forma de aproximación mediante las autoccorelaciones,
- La forma de aproximarla mediante la ecuación de Wold
- La metodología de box y Jenkins y su resolución del problema mediante la ecuación es hacia el infinito de Wold
-Las aproximaciones AR, MA, ARMA, y ARIMA
- Los autocorrelogramas para determinar el modelo ARIMA más adecuado
- Los supuestos que se deben respetar para antes de realizar el proceso de estimación y luego pronóstico.

Número: 2 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:23:39

me parece que está algo mal ubicado.

hablaría antes de la ecuación de Wold, y luego la forma de aproximarla que es mediante el método de análisis de Box - Jenkins.

Acomodar un poco la cosa.

ecuación 3 puede sustuirse por:

$$x_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + m \tag{4}$$

y de 4 puede verificarse que,

$$E(x_t) = E\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + m\right) = \sum_{j=0}^{\infty} \psi_j E\left(\varepsilon_{t-j}\right) + m = m \tag{5}$$

La principal consecuencia del teorema de Wold es que, si se conocen los ponderadores ψ_j, y además σ_{ε}^2 es ruido blanco es posible conocer el proceso por medio del cual se rige la serie cronológica. Esto permite realizar cualquier previsión, denotada por $\hat{X}_{T+h}$ para el proceso de interés x_T en el momento $T+h$ para una muestra cualquiera de T observaciones de x_t. De acuerdo con Sargent (1979), basado en el teorema de Wold, la mejor previsión posible para un proceso x_t para el momento $T+h$, denotado por $\hat{x}_{T+h}$, la predicción está dada por:

$$\hat{x}_{T+h} = \sum_{j=1}^{\infty} \psi_j \varepsilon_{T-j+1} \tag{6}$$

De la ecuación 6 se desprende que el error de previsión asociado está dado por: 1

$$x_{T+h} - \hat{x}_{T+h} = \sum_{j=1}^{\infty} \psi_j \varepsilon_{T-h+1} \tag{7}$$

2.3.2 Modelos Autorregresivos 2

Un modelo autorregresivo de orden p, denotado como $AR(p)$, considera los valores futuros de una serie cronológica como una combinación lineal las p observaciones predecesoras, un componente aleatorio y un término constante. Hipel & McLeod (1994) y Lee (s. f.) emplean la notación de la ecuación 8.

$$y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t \tag{8}$$

Donde y_t y ε_t corresponden al valor de la serie y al componente aleatorio en el momento actual t, mientras que φ_i, con $i = 1, 2, \cdots, p$ son los parámetros del modelo, y c es su término constante, que en ciertas ocasiones se suele omitir para simplificar la notación. Los parámetros de esta clase de modelos suelen estimarse mediante la ecuación de Yule-Walker (Brockwell & Davis, 2009).

Página:18

Número: 1 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:25:24

bien, pero tal vez indicar de este 2.3.1 que la ecuación de Wold es una forma general de representar una serie de tiempo que está gobernada por un proceso, y que se debe de aproximar o estimar con algo,

Luego si pasar a la metodología Box - Jenkins

Número: 2 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:26:39

antes no el box - Jenkins ?

2.3.3 Modelos de Medias Móviles

De manera similar a como un $AR(p)$ utiliza los valores pasados para pronosticar los futuros, los modelos de medias móviles de orden q, denotados como $MA(q)$, utilizan los errores pasados de las variables independientes. Estos modelos se describen mediante la ecuación 9.

$$y_t = \mu + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{9}$$

Donde μ representa el valor medio de la serie cronológica y cada valor de $\theta_j (j = 1, 2, \cdots, q)$ son los parámetros del modelo. Como los $MA(q)$ utilizan los errores pasados de la serie cronológica, se asume que estos son i.i.d. centrados en cero y con una variancia constante, siguiendo una distribución aproximadamente Normal, con lo cual este tipo de modelos pueden considerarse como una regresión lineal entre una observación determinada y los términos de error que le preceden (Agrawal & Adhikari, 2013).

2.3.4 Metodología Box-Jenkins 1

La combinación de un $AR(p)$ y un $MA(q)$, descritos en las ecuaciones 8 y 9 respectivamente, como se mencionó al inicio de esta sección, generan los modelos autorregresivos de medias móviles, $ARMA(p, q)$, representados mediante la ecuación 10.

$$y_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{10}$$

Cochrane (1997) menciona que los modelos $ARMA(p, q)$ suelen manipularse mediante lo que se conoce como operador de rezagos, denotado como $Ly_t = y_{t-1}$. Esto significa que en un $AR(p)$ se tiene que $\varepsilon_t = \varphi(L) y_t$, mientras que en $MA(q)$ se tiene que $y_t = \theta(L)\varepsilon_t$, y por consiguiente en un $ARMA(p, q)$ se tiene $\varphi(L) y_t = \theta(L)\varepsilon_t$. Por lo tanto, de lo anterior se desprende que $\varphi(L) = 1 - \sum_{i=1}^{p} \varphi_i L^i$, y que $\theta(L) = 1 + \sum_{j=1}^{q} \theta_j L^j$.

Los modelos $ARMA$, sin embargo, solamente pueden ser utilizados en series cronológicas suyo proceso es estacionario. Esto, en la práctica, es poco común, pues una serie de tiempo a menudo posee tendencias y ciertos patrones estacionales y, además, como menciona Hamzaçebi (2008), presentan procesos no estacionarios por naturaleza. Esta condición hace necesaria la introducción de una generalización de los modelos $ARMA$, la cual se conoce como los modelos $ARIMA$ (Box, Jenkins, & Reinsel, 1994).

Página:19

Número: 1 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:29:09

que extraño....

para mi esto es antes:

-Primer box-Jenkins
-luego los AR - MA - ARMA

2.3.5 Modelos ARIMA

Partiendo de una serie con un proceso no estacionario, es posible aplicar transformaciones o diferenciaciones (d)a los datos con el objetivo de convertirlos en un proceso estacionario. Utilizar la notación de rezagos descrita anteriormente, según Flaherty & Lombardo (2000), permite plantear un modelo $ARIMA(p, d, q)$ como se describe en la ecuación 11.

$$\varphi(L)(1-L)^d y_t = \theta(L)\varepsilon_t \left(1 - \sum_{i=1}^{p} \varphi_i L^i\right)(1-L)^d y_t = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right)\varepsilon_t \tag{11}$$

Donde los términos p, d y q son positivos y mayores a cero y corresponden al modelo autorregresivo, a la diferenciación y al modelo de medias móviles, respectivamente. El componente d es el número de diferenciaciones, si $d = 0$ se tiene un modelo ARMA, y $d \geq 1$ representa el número de diferenciaciones; en la mayoría de casos $d = 1$ suele ser suficiente. Así, un $ARIMA(p, 0, 0) = AR(p)$, $ARIMA(0, 0, q) = MA(q)$, y un $ARIMA(0, 1, 0) = y_t = y_{t-1} + \varepsilon_t$, es decir, un modelo de caminata aleatoria.

Como sugieren Box, Jenkins, & Reinsel (1994), lo anterior puede generalizarse aún más al considerar los efectos estacionales de la serie cronológica. Si se considera una serie cronológica con observaciones mensuales, una diferenciación de primer orden es igual a la diferencia entre una observación y la observación correpondiente al mismo mes pero del año anterior; es decir, si el periodo estacional es de $s = 12$ meses, entonces esta diferencia estacional aplicada a un $ARIMA(p, d, q)(P, D, Q)_S$ es calculada mediante $z_t = y_t - y_{t-s}$.

De esta manera, el método de Box, Jenkins, & Reinsel (1994) inicia con el análisis exploratorio de la serie cronológica, teniendo un interés particular en identificar si hay presencia de factores no estacionarios en la misma. Si en efecto se cuenta con una serie no estacionaria, ésta debe volverse estacionaria mediante algún tipo de transformación, típicamente el logaritmo natural. Con la serie ya transformada, se busca identificar el proceso que gobierna la serie. La forma clásica de hacer esto es mediante los gráficos de autocorrelación y autocorrelación parcial. Cuando se logra identificar un proceso que se adecue más a la serie cronológica, se deben realizar los diagnósticos para evaluar la calidad del ajuste del modelo, así como las medidas de rendimiento referentes a los pronósticos que genera el modelo estimado hasta un horizonte determinado. 1

2.4 Identificación del modelo

Los métodos más clásicos para la identificación del proceso que gobierna a una serie cronológica son las funciones de autocorrelación y autocorrelación parcial, las cuales sirven de indicador acerca de qué tan relacionadas están las observaciones unas de otras. Estas funciones ofrecen indicios sobre el orden de los términos para los modelos $AR(p)$, $MA(q)$ y para la diferenciación y, por ende, para

Página:20

Número: 1 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:34:45

ok

la identificación de un modelo $ARIMA$ (R. J. Hyndman & Athanasopoulos, 2018b).

Para medir la relación lineal entre dos variables cuantitativas es común utilizar el coeficiente de correlación r de Pearson (Benesty & Chen, 2009), el cual se define para dos variables X e Y como se muestra en la ecuación 12.

$$r_{X,Y} = \frac{E(XY)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2}} \tag{12}$$

Este mismo concepto puede aplicarse a las series cronológicas para comparar el valor de la misma en el tiempo t, con su valor en el tiempo $t-1$, es decir, se comparan las observaciones consecutivas Y_t con Y_{t-1}. Esto también es aplicable a no solo una observación rezagada (Y_{t-1}), sino también con múltiples rezagos $(Y_{t-2}), (Y_{t-3}), \cdots, (Y_{t-n})$. Para esto se hace uso del coeficiente de autocorrelación.

El coeficiente de autocorrelación (ACF por sus siglas en inglés) recibe su nombre debido a que se utiliza el coeficiente de correlación para pares de observaciones $r_{Y_t,Y_{t-1}}$ de la serie cronológica. Al conjunto de todas las autocorrelaciones se le llama función de autocorrelación.

La función de autocorrelación parcial[5] 1, como menciona Hernández (2011b), busca medir la asociación lineal entre las observaciones Y_t y Y_{t-k}, descartando los efectos de los rezagos $1, 2, \cdots, k-1$. 2

Cuando se tiene el modelo ARIMA debidamente identificado, es importante realizar los pronósticos. Sin embargo, estos pronósticos no son imperativos, sino que se debe evaluar su calidad con las llamadas medidas de rendimiento. Estas mediciones son hechas comparando el pronóstico y su diferencia con el valor real. Existen múltiples medidas de rendimiento, Adhikari, K, & Agrawal (2013) menciona entre ellas el *MAE*, *MAPE*, *RMSE*, *MASE*, *AIC*, *AICc* y el *BIC*.

2.5 Los autocorrelogramas

El uso del ACF y el $PACF$ se suele aplicar de manera visual. Sin embargo, hacer usos de estos elementos implica considerar múltiples condiciones. En el caso de la identificación del orden de la diferenciación:

- Si la serie posee autocorrelaciones positivas en un amplio número de rezagos, entonces es posible que se requiera un orden más alto en el valor de d.
- Si la autocorrelación en $t-1$ es menor o igual a cero, o si las autocorrelaciones resultan ser muy bajas y sin seguir algún patrón en particular, entonces no se requiere un alto orden para la diferenciación.
- Una desviación estándar baja suele ser indicador de un orden adecuado de integración.

[5] $PACF$ por sus siglas en inglés

Página:21

Número: 1 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:35:58

necesito ver ecuación sobre eso, y hacer mención y ver fórmulas + matriz, de las ecuaciones de Yules - Walker

Número: 2 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:36:29

De igual forma, me gustaría que luego de cada autocorrelación, exponga los gráficos de esta.

- Si no se utiliza ninguna diferenciación, se asume que la serie cronológica es estacionaria. Aplicar una diferenciación asume que la serie cronológica posee una media constante, mientras que dos diferenciaciones sugiere que la tendencia varía en el tiempo. 1

Para la identificación de los términos p y q:

- Si la $PACF$ de la serie cronológica diferenciada muestra una diferencia marcada y si, además, la autocorrelación en $t-1$ es positiva, entonces debe considerarse aumentar el valor de p.
- Si la $PACF$ de la serie cronológica diferenciada muestra una diferencia marcada y si, además, y la autocorrelación en $t-1$ es negativa, entonces debe considerarse aumentar el valor de q.
- Los términos p y q pueden cancelar sus efectos entre sí, por lo que si se cuenta con un modelo $ARMA$ más mixto que parece adaptarse bien a los datos, puede deberse también a que p o q deben ser menores.
- Si la suma de los coeficientes del modelo AR es muy cercana a la unidad, es necesario reducir la cantidad de términos en uno y aumentar el orden de la diferenciación en uno.
- Si la suma de los coeficientes del modelo MA es muy cercana a la unidad, es necesario reducir la cantidad de términos en uno y disminuir el orden de la diferenciación en uno. 2

Tener en consideración estos y otros posibles criterios para la identificación del proceso que gobierna la serie cronológica puede fácilmente volverse algo subjetivo, pues dos personas diferentes pueden llegar a dar distintas interpretaciones a las visualizaciones de los autocorrelogramas. Estas interpretaciones pueden sesgar la identificación de los modelos y, además, no considerar otros escenarios para los términos de un modelo $ARIMA$; para solventar esto es necesario considerar un abanico más amplio de opciones que a su vez elimine el criterio subjetivo del observador, lo cual se puede lograr al considerar múltiples permutaciones de términos, es decir, empleando la sobreparametrización.

2.6 La sobreparametrización y el análisis combinatorio

La identificación visual mediante los autocorrelogramas puede llevar a decisiones erradas acerca 3 del proceso que gobierna la serie cronológica. Una alternativa es considerar estimaciones procesos de ordenes bajos, como un $ARMA(1,1)$ y poco a poco ir incorporando términos, este proceso de revisión permite encontrar los puntos en que agregar un coeficiente más al modelo no aporta ninguna mejora en los resultados del pronóstico, y así considerar únicamente aquellos modelos que tengan coeficientes con un aporte estadísticamente significativo. Este procedimiento es conocido como sobreparametrización. Dependiendo de la cantidad de observaciones y del rango con que se trabajen los coeficientes, la comparación de los modelos puede volverse muy extensa y complicada, razón por la cual resulta imperativo generar un procedimiento sistemático que logre seleccionar el mejor modelo con base en sus medidas de ajuste y rendimiento del modelo.

Página:22

Número: 1 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:37:31

poner acá un ejemplo del autocorrlelograma para las auto, clásica.,

Número: 2 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:37:57

poner acá un ejemplo del autocorrlelograma para las auto, parcial.,

Número: 3 Autor: oscar Asunto: Resaltado Fecha: 2021-05-19 13:39:41

acá hay que hablar definitivamente de los procesos de elección...

-Multiplcación
-Combinación
- y Finalmente la permutaci´n

Luego explicar que una forma de seleccioar casos con repitencias sería la permutación.

Finalmente ver que los modelos ARIMA se podrían aproximar de dicha forma...

3 METODOLOGÍA

La aplicación de las series cronológicas tiene tres objetivos: 1) el análisis exploratorio de la serie en cuestión, 2) estimar modelos de proyección, y 3) generar pronósticos para los posibles valores futuros que tomará la serie cronológica. Asimismo, 2 existen múltiples formas de proceder mediante la etapa 1 de estimación, como lo son los métodos de suavizamiento exponencial (Brown, 1956), modelos de regresión para series temporales (Kedem & Fokianos, 2005), redes neuronales secuenciales aplicadas a datos longitudinales (Tadayon & Iwashita, 2020), estimaciones bayesianas (Jammalamadaka, Qiu, & Ning, 2018), y finalmente, los procesos Autorregresivos Integrados de Medias Móviles o ARIMA por sus siglas en inglés (Box, Jenkins, & Reinsel, 1994), siendo estos últimos el foco de interés en este estudio.

Esta sección aborda la metodología propuesta como método de estimación y pronóstico de series cronológicas. En la búsqueda de un modelo adecuado de entre varios candidatos, se cubren en un primer apartado las medidas de bondad de ajuste 3 y de precisión a utilizar.

El segundo apartado describe en detalle el uso de la sobreparametrización como herramienta para la generación de pronósticos de series cronológicas con temporalidades mensuales, bimensuales, trimestrales, cuatrimestrales o anuales mediante un proceso de selección fundamentada en las permutaciones de todos los parámetros de un modelo ARIMA hasta un rango determinado. Las medidas de precisión y de bondad de ajuste sirven de insumo para utilizar un método de consenso entre ellas y seleccionar el modelo más adecuado mediante la sobreparametrización: se comparan 4 todos los posibles en un intervalo específico de términos definiendo una diferenciación adecuada para la serie y permutando hasta un máximo definido para los términos autorregresivos y de medias móviles especificados para así seleccionar la especificación que ofrezca mejores resultados al momento de pronosticar valores futuros de la serie cronológica.

3.1 Medidas de bondad de ajuste y rendimiento

El objetivo último al estimar un modelo ARIMA es obtener los pronósticos de dicho modelo Sin embargo, estos pronósticos no son pueden asumirse como correctos, sino que se debe evaluar su calidad con las llamadas medidas de bondad de ajuste y de rendimiento. Existen múltiples medidas, Adhikari, K, & Agrawal (2013) menciona, entre otras, las siguientes:

3.1.1 AIC

Se calcula de la siguiente manera:

$$AIC = -2logL\left(\hat{\theta}\right) + 2k \qquad (13)$$

Página:23

Número: 1 Autor: oscar Asunto: Resaltado Fecha: 2021-05-22 14:47:13

esto está algo desubicado ... porque ud va a hablar de la metodología del presente trabajo, y no de forma general....

Número: 2 Autor: oscar Asunto: Resaltado Fecha: 2021-05-22 14:45:49

no creo que sea el conector adecuado.

Número: 3 Autor: oscar Asunto: Resaltado Fecha: 2021-05-22 14:47:55

me parece que el Goodness of fit, debería ser lo último....

Número: 4 Autor: oscar Asunto: Resaltado Fecha: 2021-05-22 14:54:07

Bien, pero creo que en etapas.

1. materiales
2. Métodos ARIMA y estimación
3. sobreparametrización.
4. medidas de bondad y de ajuste
5. Como va a desarrollar el análisis como tal

Importante, creo que debe hablar un poco de los procesos de simulación que piensa utilizar.

Otra cosa, creo que ud va a utilizar procesos de bootstrap u otros, y creo que debería comentar el proceso de partición de los datos y otros que se llevan a cabo en la aplicación de un procseo de análisis de series de tiempo.

Donde k es el número de parámetros y n el número de datos.

3.1.2 AICc

Su forma de cálculo se muestra en la ecuación 14

$$AICc = -2logL\left(\hat{\theta}\right) + 2k + \frac{2k+1}{n-k-1} \tag{14}$$

Donde k es el número de parámetros y n el número de datos.

3.1.3 BIC

El último estadístico de bondad de ajuste se calcula como se muestran en la ecuación 15.

$$BIC = -2logL\left(\hat{\theta}\right) + k \cdot log(n) \tag{15}$$

Donde k es el número de parámetros y n el número de datos.

3.1.4 MAE

El error absoluto medio se define mediante la ecuación 16

$$\frac{1}{n}\sum_{t=1}^{n}|e_t| \tag{16}$$

3.1.5 MASE

Esta medida de rendimiento tiene dos casos, uno para series cronológicas no estacionales y otro para series cronológicas estacionales, como se muestra en las ecuaciones 17 y 18.

$$\frac{\frac{1}{J}\sum_j |e_j|}{\frac{1}{T-1}\sum_{t=2}^{T}|Y_t - Y_{t-1}|} \tag{17}$$

$$\frac{\frac{1}{J}\sum_j |e_j|}{\frac{1}{T-m}\sum_{t=m+1}^{T}|Y_t - Y_{t-m}|} \tag{18}$$

Donde m es la temporalidad de la serie.

3.1.6 RMSE

Es la raíz del error cuadrático medio, como se define en la ecuación 19. 1

Página:24

Número: 1 Autor: oscar Asunto: Resaltado Fecha: 2021-05-22 14:52:40

esta parte bien!

$$\sqrt{\frac{1}{n}\sum_{t=1}^{n}|e_t^2|} \tag{19}$$

3.2 La sobreparametrización

La estimación de los modelos y posterior selección de los mismos vía sobreparametrización es un proceso que requiere de distintas etapas. El procedimiento completo fue programado utilizando el lenguaje R[6] y su código se muestra en el Código 1, la cuál fue construida haciendo uso de los paquetes de R `tidyr` (Wickham & Henry, 2019), `dplyr`(Wickham, François, Henry, & Müller, 2019) y `parallel`(R Core Team, 2019), los procesos internos de esta función son descritos a continuación.

A partir de una serie cronológica y_t, se realiza una partición de los datos para tener dos conjuntos distintos. Uno de ellos servirá para entrenar y estimar los distintos modelos; mientras que el segundo servirá para validar los pronósticos y posteriormente seleccionar el modelo más adecuado. De manera predeterminada, se utiliza una partición del 80 % de los datos para el conjunto de entrenamiento y un 20 % para los datos de validación, sin embargo, esto puede cambiar de acuerdo al interés propio del investigador(a). 1

Una vez que se define la partición que tendrá la serie cronológica, se prosigue con la selección de los escenarios para estimar los modelos de ARIMA. Es en esta instancia en donde se decide en valor máximo de los parámetros p, d, q, P, D, Q del modelo $ARIMA(p, d, q)(P, D, Q)_s$ que serán sujetos al análisis. Si se cuenta con una serie sin patrones estacionales y cuyo modelo con mayor cantidad de parámetros es un $ARIMA(3, 1, 4)$, la matriz de valores paramétricos es la que se muestra en 20:

$$\begin{array}{c} \overbrace{\hphantom{0\quad 0\quad 0}}^{p,d,q}\;\overbrace{\hphantom{0\quad 0\quad 0}}^{P,D,Q} \\ \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 3 & 1 & 4 & 0 & 0 & 0 \end{bmatrix} \end{array} \tag{20}$$

De manera análoga, al trabajar con un modelo con algún efecto estacional en una determinada periodicidad, como por ejemplo mensual, la matriz de valores paramétricos al definir el modelo con mayor número de parámetros como un $ARIMA(6, 1, 6)(6, 1, 6)_{12}$ es la mostrada en 21:

[6] https://cran.r-project.org/

Página:25

Número: 1 Autor: oscar Asunto: Resaltado Fecha: 2021-05-22 14:55:59

me gustaría esta explicación con un gráfico o dibujo ....

cuales pueden ser reales o simulados. Además, se especifican los valores p, d, q, P, D, Q del modelo ARIMA a partir del cual se desea obtener los datos, así como los valores de los coeficientes presentes en el modelo. Es partiendo de este modelo que se simulan los datos de las series cronológicas que serán insumo para la prueba la selección mediante sobreparametrización, este procedimiento se encuentra en el Código 2.

1

Página:27

Número: 1 Autor: oscar Asunto: Nota adhesiva Fecha: 2021-05-22 15:01:26

Creo que hacen falta bastante elementos

-No hay materiales
-la selección del entrenamiento y validación no es clara,
-Si en el marco teórico hablo de los modelo ARIMA, debe volver a recapitular la fórmula, pero ahora mostrar un poco lo que es sus métodos de estimación (OLS, cuasi OLS, u otros, poner fórmular),
y pues lo quye se tendría en términos de ecuación.
Luego con lo anterior, ahora si introduccir lo que es la sobre parametrización y como eso hace variar el procesod de estimación.
- Luego debe entonces hablar el proceso de estimación para el set de datos training, y finalmente como ese proceso afecta el validation data set
- Finalmente hablar sobre el proceso de validación para la toma en cuenta del mejor modelo de estimación.

A su vez, ud en la metodología debería hablar un poco sobre el proceso de simulación y sobre qué es lo que va a comparar...

Igualmente con que datos va a trabajar....

Terminado todo eso, creo firmemente que debe cerrar recapitulando un poco lo que es la forma de actuar o proceder en un análisis de series de tiempo bajo la sobre parametrización.