

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

LA SOBREPAREMETRIZACIÓN EN EL ARIMA: UNA APLICACIÓN A
DATOS COSTARRICENSES

Tesis sometida a la consideración de la Comisión del Programa de Estudios de
Posgrado en Estadística para optar por el grado y título de Maestría Académica en
Estadística

CÉSAR ANDRÉS GAMBOA SANABRIA B12672

Ciudad Universitaria Rodrigo Facio, Costa Rica

2020

DEDICATORIA

Pendiente

AGRADECIMIENTOS

También pendiente

“Esta tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Estadística de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Estadística”

Ph.D. Álvaro Morales Ramírez
Decano Sistema de Estudios de Posgrado

MSc. Óscar Centeno Mora
Director de Tesis

Ph.D. Gilbert Brenes Camacho
Lector

Ph.D. ShuWei Chou.
Lector

MSc. Johnny Madrigal Pana
Director Programa de Posgrado en Estadística

César Andrés Gamboa Sanabria
Candidato

Índice



DEDICATORIA	I
AGRADECIMIENTOS	II
RESUMEN	1
ABSTRACT	2
1 INTRODUCCIÓN	3
1.1 Antecedentes	3
1.2 El problema	4
1.3 Objetivos del estudio	5
1.4 Metodología del estudio	6
1.5 Justificación del estudio	7
1.6 Organización del estudio	8
2 MARCO TEÓRICO	9
2.1 Introducción	9
2.2 Investigaciones relacionadas	9
2.3 Observaciones finales sobre la revisión bibliográfica	9
3 METODOLOGÍA	11
3.1 Introducción	11
3.2 Conceptos y definiciones en el análisis de series cronológicas	11
3.2.1 Definición de una serie cronológica	11
3.2.2 Procedimiento al analizar series cronológicas	11
3.2.3 Estacionaridad	11
3.2.4 La parsimonia	11
3.3 Componentes de una serie cronológica	11
3.3.1 La tendencia	11
3.3.2 Componentes estacionales	11
3.3.3 Componente cíclico	11
3.3.4 Componente irregular	11
3.4 Supuestos en el análisis de series cronológicas	11
3.5 Modelos de series cronológicas	11
3.6 Modelos Autorregresivos Integrados de Medias Móviles	11
3.6.1 Modelos Autorregresivos	11
3.6.2 Modelos de Medias Móviles	11

3.6.3	Metodología Box-Jenkins	11
3.6.4	Etapa 1 - Identificación	11
3.6.5	Etapa 2 - Estimación y diagnóstico	11
3.6.6	Etapa 3 - Pronóstico	11
3.6.7	Notación de los modelos ARIMA	11
3.6.8	Diferenciación	11
3.7	Análisis de intervención	11
3.8	Validación cruzada	11
3.9	Medidas de rendimiento	11
3.9.1	MFE	11
3.9.2	MAE	11
3.9.3	MAPE	11
3.9.4	MPE	11
3.9.5	MSE	11
3.9.6	SSE	11
3.9.7	SMSE	11
3.9.8	RMSE	11
3.9.9	NMSE	11
3.9.10	AIC	11
3.9.11	AICc	11
3.9.12	BIC	11
3.10	La sobreparametrización	11
3.11	Simulación de series cronológicas	11
3.12	El método propuesto	11
4	RESULTADOS	12
4.1	Introducción	12
4.2	Datos simulados	12
4.2.1	Comparación en datos simulados - Sobreparametrización vs auto.arima	12
4.3	Estimaciones en datos costarricenses	12
4.3.1	Tasa de mortalidad infantil interanual	12
4.3.2	Tasa global de fecundidad	12
4.3.3	Mortalidad por causa externa	12
4.3.4	Incentivos salariales del sector público	12
4.3.5	Intereses y comisiones del sector público	12
4.3.6	Demanda eléctrica	12
4.3.7	Comparación en datos reales - Sobreparametrización vs auto.arima	12

4.4	Discusión de los resultados	12
5	CONCLUSIONES Y RECOMENDACIONES	13
5.1	Introducción	13
5.2	Conclusiones	13
5.3	Recomendaciones	13
6	ANEXOS	14
6.1	La función <code>funcion_1</code>	14
7	REFERENCIAS	15

Índice de cuadros

Índice de figuras

RESUMEN

ABSTRACT

1 INTRODUCCIÓN

1.1 Antecedentes



El manejo de información obtenida de manera secuencial, a lo largo del tiempo, hace referencia al uso de series cronológicas. Este tipo de datos se encuentra en diferentes áreas de investigación. En el campo financiero, por ejemplo, es común hablar de la devaluación del colón con respecto al dólar, cantidad de exportaciones mensuales de un determinado producto o las ventas de este (Hernández, 2011).




En la demografía, que es la rama de la estadística que estudia las poblaciones humanas según sus características o evolución a lo largo de su historia tiene, entre otras cosas, tres componentes que son de particular interés: fecundidad, mortalidad y migración. Uno de los principales temas de investigación en la demografía se trata de las proyecciones de población, el cual tiene un alto impacto a nivel social y económico, pues conocer con anticipación el posible comportamiento de la población en el futuro es clave para una adecuada planificación en diversos proyectos sobre los cuales se debe distribuir un presupuesto que, por norma general, es finito. Un claro ejemplo de esto se da durante una emergencia pues considerando el hecho de que difícilmente se sabe cuándo ocurrirá, conocer la posible cantidad de población que habita una zona es clave para la rápida reacción de las autoridades en el envío de ayuda o en la ejecución de planes de evacuación; recientemente en Costa Rica y en muchos lugares del mundo las autoridades gubernamentales debieron actuar con la mayor rapidez y eficacia posible ante la pandemia del covid-19. Gracias a las proyecciones de población es posible conocer, de manera aproximada, el posible tamaño de la población según sexo y grupos de edad; esto es de particular interés pues el covid-19 es particularmente peligroso para los adultos con 65 años y más, por cual conocer el tamaño poblacional del grupo de mayor riesgo, así como los adultos y los niños permiten conocer qué tanto podría propagarse el virus y cuántos podrían ser los infectados a lo largo del tiempo.

El campo actuarial también se ve beneficiado al mejorar sus métodos de pronóstico, pues uno de sus principales campos de estudio es la mortalidad, que representa un insumo de vital importancia para la planificación y sostenibilidad de los sistemas de pensiones, servicios de salud tanto pública como privada, seguros de vida y asuntos hipotecarios (Rosero-Bixby, 2018).

Sin embargo, las series cronológicas por sí solas representan solo un insumo para abordar, como mínimo, tres objetivos básicos: 1) realizar análisis exploratorios usando mediante métodos de visualización y medidas de posición y variabilidad, como ver su crecimiento o decrecimiento a lo largo del tiempo, detectar valores atípicos o cambios drásticos en el nivel o valor medio de la serie, 2) generar modelos estadísticos que sirvan como una simplificación de la realidad, y 3) generar pronósticos para los posibles valores futuros que tomará el problema en cuestión.



Los tres objetivos anteriores se trabajan de manera secuencial, pues es necesario realizar primero el análisis exploratorio de los datos para tener una noción global del panorama y así conocer la serie cronológica con la que se está trabajando. Una vez hecho esto, existen múltiples formas de generar modelos para estos datos, como por ejemplo los métodos de suavizamiento exponencial desarrollados en la década de 1950 (Brown, 1956), modelos de regresión para series temporales (Kedem & Fokianos, 2005) o los procesos autorregresivos integrados de medias móviles o ARIMA por si siglas en inglés (Box, Jenkins, & Reinsel, 1994). Cuando se ha establecido el modelo, los pronósticos son utilizados en instituciones públicas, gobiernos municipales, instituciones del sector privado, centros académicos, población civil, 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 mayor y más efectivo. 

De lo anterior debe rescatarse que generar un modelo adecuado es fundamental para obtener un pronóstico de calidad y es aquí donde resulta importante mencionar una diferencia clave entre los dos modelos clásicos más utilizados: los modelos de suavizamiento exponencial y los modelos ARIMA. Ambos representan enfoques complementarios a un problema, pues según los modelos de suavizamiento exponencial se fundamentan en un enfoque más descriptivo de los componentes de la serie cronológica en estudio, mientras que los modelos ARIMA tienen como objetivo explicar las relaciones pasadas de ésta (R. J. Hyndman & Athanasopoulos, 2018a). La importancia de la metodología de Box-Jenkins radica en que no supone ningún patrón en particular en la serie histórica que se busca pronosticar, sino que contempla un proceso iterativo para identificar un posible modelo a partir de una clase general de modelos, para luego evaluar su ajuste sometiénolo a diferentes pruebas y medidas de rendimiento. Al trabajar con la metodología de Box-Jenkins, uno de los pasos a concretar es identificar los parámetros del proceso $ARIMA(p, d, q)(P, D, Q)_s$ que gobiernan la serie, siendo la manera clásica de trabajar este paso, el análisis visual de las funciones de autocorrelación parcial y total.

1.2 El problema

El gran obstáculo que presenta esta identificación visual es que, actualmente, es común disponer de una gran cantidad de series cronológicas para analizar, lo que dificulta tener suficiente personal capacitado para realizar exitosamente este análisis visual e identificar adecuadamente los modelos. Aunado a esto, aún en el caso de que se dispusiera de los recursos para contar con el personal suficientemente capacitado para realizar esta identificación visual, siempre existirá el problema de la subjetividad, pues a pesar de que alguien proponga un patrón que gobierne la serie, otra persona podría tener otra interpretación del proceso, porponiendo así una especificación diferente. Esto sucede incluso con cantidades moderadas de series cronológicas a analizar, por lo que la generación

de algoritmos que ayuden a esta identificación se vuelve cada vez más necesaria (Hyndman & Khandakar, 2008).

Existiendo varias las aproximaciones a un método que genere de manera automática un modelo ARIMA, como por ejemplo los propuestos por Hannan y Rissanen (Hannan & Rissanen, 1982), la extensión de dicha propuesta realizada por Gómez (Gómez, 1998) y posteriormente aplicada (Gómez & Maraval, 1998) en los softwares TRAMO y SEATS; de manera similar se planteó una aplicación en los programas SCA-Expert (Liu, 1989) y TSE-AX (Mélard & Pasteels, 2000). Otros algoritmos desarrollados en programas de cómputo de paga son Forecast Pro (Goodrich, 2000) y Autobox (Reilly, 2000). Uno de los métodos automatizados de estimación más populares es el que ofrece el paquete `forecast` (Hyndman & Khandakar, 2008) del lenguaje de programación R¹ y permite hacer uso de la función `auto.arima()` para estimar un modelo ARIMA basado en pruebas de raíz unitaria y minimización del AICc (Burnham & Anderson, 2007). De esta manera 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, alternativas como la función `auto.arima()` no someten a prueba las posibles especificaciones de un modelo en un rango determinado, concepto conocido como sobreparametrización, dejando así un vacío en el cuál se corre el riesgo de no seleccionar un modelo que ofrezca mejores pronósticos.

1.3 Objetivos del estudio

Teniendo en consideración lo previamente mencionado, esta investigación tiene como objetivo general evaluar la calidad de los pronósticos realizados con modelos ARIMA especificados vía sobreparametrización para proponer un modelo adecuado en una serie cronológica. Para lograr esto, se busca de manera específica alcanzar lo siguiente:

1. Diseñar un algoritmo para la selección del mejor modelo ARIMA según la temporalidad de la serie.
2. Aplicar validación cruzada en distintos horizontes de pronóstico para identificar la mejor especificación de un modelo ARIMA.
3. Comparar la precisión de los pronósticos con métodos similares, como el propuesto por Rob Hyndman, de la Oficina de Censos de los Estados Unidos, entre otros.
4. Integrar el desarrollo de la metodología de análisis de series temporales en una librería del lenguaje estadístico R.

¹Descarga gratuita en <https://cran.r-project.org/>

1.4 Metodología del estudio

Los modelos ARIMA, junto con los de suavizamiento exponencial, son los de uso más extendido en el análisis de series cronológicas. El nombre ARIMA es la abreviatura inglesa para AutoRegressive Integrated Moving Average, y son aplicados mediante la metodología de Box-Jenkins. Como menciona Rob. Hyndman (R. J. Hyndman & Athanasopoulos, 2018b), la metodología de Box-Jenkins difiere a los demás métodos porque no supone un determinado patrón en la serie cronológica, sino que parte de un proceso iterativo para identificar el modelo de un gran grupo de estos para luego ponerlo a prueba según varias medidas de rendimiento. Un proceso ARIMA es caracterizado por dos funciones: la autocorrelación y la autocorrelación parcial; es mediante la comparación de dichas funciones que la metodología Box-Jenkins busca la identificación el proceso que describa de manera adecuada el comportamiento de una serie cronológica.

De esta manera, el método de Box-Jenkins inicia con el análisis exploratorio de la serie cronológica de interés, 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 adecúe 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.

Es a partir de la necesidad de obtener pronósticos lo más precisos posible que el presente trabajo propone una metodología para la estimación un modelo ARIMA de una serie cronológica determinada abarcando más posibilidades que los enfoques tradicionales, cuya temporalidad sea mensual, bimensual, trimestral o cuatrimestral, mediante un proceso de selección fundamentada en las permutaciones de todos los parámetros de un modelo ARIMA hasta en un rango determinado, considerando además la inclusión semiautomática de intervenciones en periodos específicos y la validación cruzada para evaluar la calidad de las particiones de la base de datos en conjuntos para entrenar y probar el rendimiento del modelo. Dichas pruebas involucran, entre otras medidas de rendimiento, el MAE, RMSE, MAPE y MASE, las cuales sirven de insumo para utilizar un método de consenso entre ellas y seleccionar el modelo más adecuado mediante la sobreparametrización: se comparan 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 de especificación de un $ARIMA(p, d, q)(P, D, Q)_s$ para así seleccionar la especificación que ofrezca mejores resultados al momento de pronosticar valores futuros de la serie cronológica. El método propuesto se probará comparándose con los resultados de seis series con distintas temporalidades: mortalidad

infantil, mortalidad por causa externa, nacimientos, demanda eléctrica, intereses y comisiones del sector público e incentivos salariales del sector público.

1.5 Justificación del estudio

El principal aporte de este estudio es, por medio de un proceso de simulación, aportar evidencia sobre cómo la sobreparametrización puede contribuir a definir la especificación de un modelo ARIMA que genere pronósticos adecuados, contrastando la calidad de estos con respecto a otros métodos similares, como lo son las funciones `auto.arima()` o `seas()`.

Una vez que el método ha sido desarrollado y probado mediante datos simulados, la aplicación real mostrará el verdadero alcance y potencial de la sobreparametrización en la estimación de modelos ARIMA al ser utilizado en datos reales para de este modo representar una herramienta para la toma de decisiones. En el campo demográfico, por ejemplo, las estadísticas vitales son sistematizadas y divulgadas año tras año, por tanto, revelan los cambios acontecidos durante este periodo. Esta información junto con la proveniente de los censos de población constituye la base para construir los diferentes índices, tasas y otros indicadores que revelan la situación demográfica del país, información de gran relevancia para la planificación nacional, regional y local en diversos campos. Uno de estos principales campos o áreas de acción es la salud pública, para la cual la tasa de mortalidad infantil se considera uno de los indicadores prioritarios dado que refleja no solo las condiciones de salud de la población infante, sino también los niveles de desarrollo del país, pues depende de la calidad de la atención de la salud, principalmente de la prenatal y perinatal, así como de las condiciones de saneamiento. Por tanto, su continuo monitoreo es fundamental para diseñar, implementar y evaluar políticas de salud pública orientadas a disminuir y erradicar aquellas que son prevenibles (INEC, 2017). Por su parte, la Contraloría General de la República dispone de información completa, pública, proveniente de fuentes primarias, actualizada en tiempo, de libre acceso, procesable tanto por humanos como por máquinas, en formatos abiertos, con licencia libre de uso, permanente y sin costos de utilización.

La validación de los datos se realiza en dos vías: un protocolo de detección semiautomático de inconsistencias el cual consiste en la tabulación de cuadros que sirvan para visualizar los cruces más importantes entre variables y que dejen en evidencia casos extraños, como que ocurran nacimientos en un asilo de ancianos, o verificar la edad de la madre con la cantidad de hijos tenidos. También se realiza una segunda verificación manual para revisar casos que por la forma en que fueron digitados luzcan sospechosos, como un peso al nacer inferior a 500 gramos o números de cédula repetidos en nacimientos que no son múltiples. En casos muy puntuales en los que se disponga de un medio de comunicación con el informante, este es contactado para corroborar datos que luzcan sospechosos o bien para completar información faltante.

1.6 Organización del estudio



Esta tesis consta de cinco capítulos, de los cuáles el primero ofrece una contextualización del uso de series de tiempo, así como la importancia de poder contar con pronósticos de calidad. Se presentaron además objetivos que busca alcanzar el estudio así como una breve descripción de la metodología que se empleará y que será discutida más detalladamente en los capítulos posteriores. Se concluye esta sección con hechos que justifican la importancia de esta investigación.

El siguiente capítulo consiste en el marco teórico y la revisión de la literatura asociada con el estudio.

En el tercer capítulo se describe en detalle toda la metodología relacionada al estudio, iniciando con una descripción global de los conceptos más fundamentales del análisis de series cronológicas, pasando por los componentes fundamentales de las mismas: tendencia, estacionalidad, ciclos e irregularidades. En este capítulo se discuten también los supuestos clásicos del análisis de series cronológicas, los distintos tipos de modelos, el análisis de intervención, la validación cruzada 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 que se utilizará, así como la discusión del método propuesto.

El siguiente capítulo consiste en la presentación de los resultados, tanto en los datos simulados como en la aplicación a datos costarricenses.

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

Finalmente se ofrece en la sección de anexos los códigos en lenguaje R más importantes utilizados en el estudio, así como una sección dedicada a las referencias bibliográficas obtenidas para esta investigación.

2 MARCO TEÓRICO

2.1 Introducción

2.2 Investigaciones relacionadas

2.3 Observaciones finales sobre la revisión bibliográfica

3 METODOLOGÍA

3.1 Introducción

3.2 Conceptos y definiciones en el análisis de series cronológicas

3.2.1 Definición de una serie cronológica

3.2.2 Procedimiento al analizar series cronológicas

3.2.3 Estacionaridad

3.2.4 La parsimonia

3.3 Componentes de una serie cronológica

3.3.1 La tendencia

3.3.2 Componentes estacionales

3.3.3 Componente cíclico

3.3.4 Componente irregular

3.4 Supuestos en el análisis de series cronológicas

3.5 Modelos de series cronológicas

3.6 Modelos Autorregresivos Integrados de Medias Móviles

3.6.1 Modelos Autorregresivos

3.6.2 Modelos de Medias Móviles

3.6.3 Metodología Box-Jenkins

3.6.4 Etapa 1 - Identificación

3.6.5 Etapa 2 - Estimación y diagnóstico

3.6.6 Etapa 3 - Pronóstico

3.6.7 Notación de los modelos ARIMA

3.6.8 Diferenciación

3.7 Análisis de intervención

3.8 Validación cruzada

3.9 Medidas de rendimiento

3.9.1 MFE

4 RESULTADOS

4.1 Introducción

4.2 Datos simulados

4.2.1 Comparación en datos simulados - Sobreparametrización vs auto.arima

4.3 Estimaciones en datos costarricenses

4.3.1 Tasa de mortalidad infantil interanual

4.3.2 Tasa global de fecundidad

4.3.3 Mortalidad por causa externa

4.3.4 Incentivos salariales del sector público

4.3.5 Intereses y comisiones del sector público

4.3.6 Demanda eléctrica

4.3.7 Comparación en datos reales - Sobreparametrización vs auto.arima

4.4 Discusión de los resultados

5 CONCLUSIONES Y RECOMENDACIONES

5.1 Introducción

5.2 Conclusiones

5.3 Recomendaciones

6 ANEXOS

6.1 La función `funcion__1`

Código 1: Una función

```
funcion_1 <- function(x,y){  
  x+y  
}
```

7 REFERENCIAS

- Box, G. E., Jenkins, G. M., & Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. Recuperado de <https://books.google.co.cr/books?id=sRzvAAAAMAAJ>
- Brown, R. (1956). *Exponential Smoothing for Predicting Demand*. Recuperado de <https://www.industrydocuments.ucsf.edu/docs/jzlc0130>
- Burnham, K. P., & Anderson, D. R. (2007). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Recuperado de <https://books.google.co.cr/books?id=IWUKBwAAQBAJ>
- Canova, F., & Hansen, B. E. (1995). Are Seasonal Patterns Constant over Time? A Test for Seasonal Stability. *Journal of Business & Economic Statistics*, 13(3), 237-252. Recuperado de <http://www.jstor.org/stable/1392184>
- Fuller, W. A. (1995). *Introduction to Statistical Time Series*. Recuperado de <https://books.google.co.cr/books?id=wyRhjmAPQIYC>
- Goodrich, R. L. (2000). The Forecast Pro Methodology. *International Journal of Forecasting*, 16(4), 533-535. Recuperado de [http://www.forecasting-competition.com/downloads/NN3/methods/Goodrich%20\(2000\)%20The%20Forecast%20Pro%20methodology%20science.pdf](http://www.forecasting-competition.com/downloads/NN3/methods/Goodrich%20(2000)%20The%20Forecast%20Pro%20methodology%20science.pdf)
- Gómez, V. (1998). *Automatic Model Identification in the Presence of Missing Observations and Outliers*. (D. G. de A. y P. P. Ministerio de Economía y Hacienda, Ed.). Working paper D-98009.
- Gómez, V., & Maraval, A. (1998). *Programs TRAMO and SEATS, Instructions for the Users*. (D. G. de A. y P. P. Ministerio de Economía y Hacienda, Ed.). Working paper 97001.
- Hannan, E. J., & Rissanen, J. (1982). Recursive Estimation of Mixed Autoregressive-Moving Average Order. *Biometrika*, 69(1), 81-94. Recuperado de <http://www.jstor.org/stable/2335856>
- Hernández, O. (2011). *Introducción a las Series Cronológicas* (1.^a ed., p. 1). Recuperado de <http://www.editorial.ucr.ac.cr/ciencias-naturales-y-exactas/item/1985-introduccion-a-las-series-cronologicas.html>
- Hyndman, R. J., & Athanasopoulos, G. (2018a). *Forecasting: principles and practice*. Recuperado de https://books.google.co.cr/books?id=_bBhDwAAQBAJ
- Hyndman, R. J., & Athanasopoulos, G. (2018b). *Forecasting: principles and practice*. Recuperado de https://books.google.co.cr/books?id=_bBhDwAAQBAJ
- Hyndman, R., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software, Articles*, 27(3), 1-22. <https://doi.org/10.18637/jss.v027.i03>
- INEC. (2017). *Población, nacimientos, defunciones y matrimonios*. Recuperado de http://inec.cr/sites/default/files/documentos-biblioteca-virtual/replac2017_0.pdf
- Kedem, B., & Fokianos, K. (2005). *Regression Models for Time Series Analysis*. Recuperado de <https://books.google.co.cr/books?id=8r0qE35wt44C>

-
- Liu, L.-M. (1989). Identification of seasonal arima models using a filtering method. *Communications in Statistics - Theory and Methods*, 18(6), 2279-2288. <https://doi.org/10.1080/03610928908830035>
- Mélard, G., & Pasteels, J.-M. (2000). Automatic ARIMA modeling including interventions, using time series expert software. *International Journal of Forecasting*, 16(4), 497-508. [https://doi.org/https://doi.org/10.1016/S0169-2070\(00\)00067-4](https://doi.org/https://doi.org/10.1016/S0169-2070(00)00067-4)
- Osborn, D. R., Chui, A. P. L., Smith, J., & Birchenhall, C. (2009). *Seasonality and the order of integration for consumption*. Recuperado de http://www.est.uc3m.es/esp/nueva_docencia/comp_col_get/lade/tecnicas_prediccion/OCSB_OxBull1988.pdf
- Reilly, D. (2000). The Autobox System. *International Journal of Forecasting*, 16(4), 531-533. Recuperado de <https://ideas.repec.org/a/eee/intfor/v16y2000i4p531-533.html>
- Rosero-Bixby, L. (2018). *Producto C para SUPEN. Proyección de la mortalidad de Costa Rica 2015-2150*. Recuperado de CCP-UCR website: <http://srv-website.cloudapp.net/documents/10179/999061/Nota+t%C3%A9cnica+tablas+de+vida+segunda+parte>
- Xiao, Z. (2001). Testing the Null Hypothesis of Stationarity Against an Autoregressive Unit Root Alternative. *Journal of Time Series Analysis*, 22(1), 87-105. <https://doi.org/10.1111/1467-9892.00213>