Determinación de modelos ARIMA vía sobre parametrización según la temporalidad de la serie cronológica con aplicaciones en datos costarricenses

Universidad de Costa Rica

César Gamboa Sanabria

www.cesargamboasanabria.com

info@cesargamboasanabria.com

01 enero, 2020

RESUMEN

La metodología de Box-Jenkins busca encontrar el mejor proceso autorregresivo integrado de medias móviles (ARIMA) que explique una serie temporal y_t de T periodos, para pronosticar hasta $T + h$. El paquete `forecast` de R permite hacer uso de la función `auto.arima()` para estimar un modelo ARIMA basado en pruebas de raíz unitaria, minimización del AICc y de la MLE. De esta forma se obtiene un modelo temporal definiendo las diferenciaciones requeridas en la parte estacional d mediante las pruebas KPSS o ADF, y la no estacional D utilizando las pruebas OCSB o la Canova-Hansen, seleccionado el orden óptimo para los términos $ARIMA(p, d, q)(P, D, Q)$ para una serie y_t. Se propone un método de selección fundamentada en las permutaciones de los parámetros de un modelo ARIMA, seleccionando la mejor especificación con base en medidas de rendimiento MAE, RMSE, MAPE y MASE: se comparan todos los posibles términos definiendo una diferenciación adecuada para la serie y permutando hasta un máximo determinado para los términos de especificación de un $ARIMA(p, d, q)(P, D, Q)$. El método propuesto se probó en 6 series cronológicas de distinta temporalidad: mortalidad infantil, mortalidad por causa externa, nacimientos, demanda eléctrica, intereses y comisiones del sector público, e incentivos salariales del sector público.

Palabras clave: ARIMA, R, automatización, selección, estadística

Índice

1 INTRODUCCIÓN

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 encuentras 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, 2011a).

En demografía, por ejemplo, el tema de las proyecciones de población tiene un alto impacto a nivel social, pues conocer con anticipación en 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 es finito. Durante una emergencia, que difícilmente se sabe cuándo ocurrirá, conocer la posible población que se tiene en una zona es clave para la rápida reacción de las autoridades para el envío de ayuda o para ejecutar planes de evacuación.

El campo actuarial también se ve beneficiado al mejorar sus métodos de pronóstico, pues uno de sus campos de estudio es la mortalidad pues representan 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 de estos datos 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 un simplificación de la realidad, y 3) generar pronósticos para los posibles valores futuros que tomará el problema en cuestión (Hernández, 2011b).

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 (ARIMA) (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 ONG 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, 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 comúnmente utilizados: los modelos de suavizamiento y los modelos ARIMA. Ambos representan enfoques complementarios a un problema, pues según Hyndman (R. J. Hyndman & Athanasopoulos, 2018a), 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. 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 y luego someter dicho modelo a diferentes pruebas y medidas de rendimiento para evaluar su ajuste. Al trabajar la metodología de Box-Jenkins, uno de los pasos es identificar el los parámetros del proceso ARIMA(p,d,q)(P,D,Q) que gobiernan la serie, siendo la manera clásica de trabajar este paso, el análisis visual de las funciones de autocorrelación parcial y total.

El gran obstáculo que presenta esta identificación visual es que en la actualidad contar con una gran cantidad de series cronológicas para analizar es algo muy común. Incluso con cantidades moderadas de series cronológicas a analizar, es difícil contar con personal capacitado para realizar este análisis visual y poder identificar los modelos, por lo que la generación de algoritmos que ayuden a dicha identificación se vuelven cada vez más necesarios (Hyndman & Khandakar, 2008).

Han sido 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 software **TRAMO** y **SEATS**; de manera similar se planteó una aplicación en los software **SCA-Expert** (Liu, 1989) y **TSE-AX** (Mélard & Pasteels, 2000). Otros algoritmos implementados en programas de cómputo de paga son **Forecast Pro** (Goodrich, 2000) y **Autobox** (Reilly, 2000). Uno de los métodos automatizados de estimación es el que ofrece el paquete `forecast` (Hyndman & Khandakar, 2008) del lenguaje de programación R[1] permite hacer uso de la función `auto.arima()` para estimar un modelo ARIMA basado en pruebas de raíz unitaria, minimización del AICc y de la MLE. De esta forma se obtiene un modelo temporal definiendo las diferenciaciones requeridas en la parte estacional d mediante las pruebas KPSS o ADF, y la no estacional D utilizando las pruebas OCSB o la Canova-Hansen, seleccionado el orden óptimo para los términos $ARIMA(p,d,q)(P,D,Q)_s$ para una serie cronológica determinada.

Es a partir de esta necesidad que se propone una metodología para la estimación del mejor modelo ARIMA para una serie cronológica determinada cuya temporalidad sea mensual, bimensual, trimestral o cuatrimestral mediante un proceso de selección fundamentada en las permutaciones de todos los pará-

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

metros de un modelo ARIMA hasta un cierto límite, considerando además la inclusión semi-automática 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 para seleccionar el modelo más adecuado: se comparan todos los posibles 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 6 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.

1.1 Contribución de la tesis a la Estadística como disciplina

El principal aporte de este estudio es, por medio de un estudio de simulación, aportar evidencia sobre cómo la sobreparametrización puede representar una herramienta para definir la especificación de un modelo ARIMA que genere pronósticos adecuados, contrastando la calidad de estos con respecto a otros métodos similares, como lo es la función auto.arima().

1.2 Objetivos

1.2.1 Objetivos general

- Evaluar la calidad de los pronósticos realizados con modelos ARIMA especificados vía sobre parametrización.

1.2.2 Objetivos específicos

- Diseñar un algoritmo para la selección del mejor modelo ARIMA según la temporalidad de la serie.
- Aplicar validación cruzada en distintos horizontes de pronóstico para identificar la mejor especificación de un modelo ARIMA.
- Comparar la precisión de los pronósticos con el método propuesto por Rob Hyndman.
- Integrar la metodología de análisis de series temporales en una librería del lenguaje estadístico R.

2 REFERENTES O ELEMENTOS TEÓRICOS QUE VA A UTILIZAR

2.1 Modelos Arima

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 *AutoRegresive 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, si no 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; el enfoque Box-Jenkins compara estas funciones con el objetivo de identificar el proceso que describa de manera adecuada el comportamiento de una serie cronológica (Hernández, 2011c).

El componente **AR** de los modelos ARIMA hace referencia al uso de modelos autorregresivos, en los cuales los pronósticos para la variable de interés utilizan una combinación lineal de las observaciones previas, llamándose así *autorregresivos* porque se aplica una regresión de dicha variable de interés con respecto a sí misma; caso contrario a la regresión múltiple, en donde los pronósticos se realizan con respecto a una combinación lineal de distintos predictores. Un modelo autorregresivo de orden p para una serie cronológica y_t puede expresarse de la siguiente manera

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t \tag{1}$$

Donde el término ϵ_t representa ruido blanco. El modelo anterior es muy similar a una regresión lineal múltiple, donde cada coeficiente ϕ va acompañado por su correspondiente rezago y_{t-p}. De manera muy similar, el término **MA** en los modelos ARIMA se refieren a los modelos de medias móviles, los cuales para pronosticar hacen uso de los errores; el modelo de medias móviles puede representarse de la siguiente manera:

$$y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \tag{2}$$

Donde el término ϵ_t representa nuevamente el ruido blanco. La ecuación anterior representa un modelo de medias móviles de orden q, en la cual cada término ϵ_{t-q} se entiende como una media móvil de los t

previos errores de predicción.

El componente **I** de los modelos ARIMA se refiere a "Itegrated", es decir, a la estacionariedad de la serie cronológica. Tradicionalmente, la metodología de Box-Jenkins consiste en visualizar la serie cronológica con el objetivo de, en caso de ser necesario, transformar los datos para estabilizar la variancia y generar así un proceso estacionario. Se dice que una serie posee un comportamiento estacionario si el comportamiento de esta no depende del tiempo, por lo que en principio no presentaría ningún patrón particular con respecto al tiempo; en otras palabras, la serie posee un movimiento bastante horizontal.

Cuando la serie cronológica muestre indicios de tendencia o patrones estacionales que resulten en un conjunto de datos que no es estacionario por naturaleza, es necesario realizar transformaciones sobre los datos para hacer que la serie se vuelva estacionaria (Adhikari et al., 2013a). Estas transformaciones hacen referencia al uso de logaritmos o alguna potencia que logre estabilizar la variabilidad de la serie. Los métodos más clásicos para identificar la no estacionariedad en una serie cronológica son las previamente mencionadas 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 previamente mencionados **AR** y **MA**.

2.2 Función de autocorrelación

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 sigue:

$$r_{X,Y} = \frac{E(XY)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2}} \tag{3}$$

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 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.

2.3 Función de autocorrelación parcial

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

2.4 Modelos ARIMA no estacionales

Como se mencionó anteriormente, los modelos ARIMA aplicados a una serie cronológica Y_t son una combinación de un modelo autorregresivo, uno de medias móviles, y alguna clase de diferenciación (logarítmica, exponenciación) para así obtener una serie diferenciada Y_t'. Si se juntan ambas se obtiene un modelo ARIMA(p,d,q) que no cubre los efectos estacionales, donde p es el orden del modelo autorregresivo, d e el grado de la diferenciación y q es el orden del modelo de medias móviles; y cuya estructura se muestra en la ecuación (4):

$$y_t' = c + \phi_1 y_{t-1}' + \phi_2 y_{t-2}' + \cdots + \phi_p y_{t-p}' + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \tag{4}$$

2.5 Modelos ARIMA estacionales

Los modelos ARIMA son también capaces de cubrir los efectos estacionales, es decir, particularidades de la serie cronológica que se repiten periódicamente con una cierta temporalidad (mensual, bimensual, etc.). Para ello se incorporan términos adicionales al modelo relacionados con la parte estacional de una manera similar a como se incorporan en el modelo ARIMA no estacional, pero ahora considerando retrocesos según sea la temporalidad estacional, pasando así de un $ARIMA(p, d, q)$ a un $ARIMA(p, d, q)(P, D, Q)_S$, donde P, D y Q se refieren a la parte estacional y S a la temporalidad presente en la serie.

2.6 Medidas de rendimiento

Cuando se tiene el modelo ARIMA estimado, 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 (Adhikari et al., 2013b) menciona las siguientes:

2.6.1 MAE

El error absoluto medio se define como $\frac{1}{n}\sum_{t=1}^{n}|e_t|$.

2.6.2 MAPE

El porcentaje promedio de error absoluto se define como $\frac{1}{n}\sum_{t=1}^{n}\left|\frac{e_t}{y_t}\right| \cdot 100$.

2.6.3 RMSE

Es la raíz del error cuadrático medio $\sqrt{\frac{1}{n}\sum_{t=1}^{n}|e_t^2|}$.

2.6.4 MASE

Para series no estacionales: $\frac{\frac{1}{J}\sum_{j}|e_j|}{\frac{1}{T-1}\sum_{t=2}^{T}|Y_t-Y_{t-1}|}$

Para series estacionales: $\frac{\frac{1}{J}\sum_{j}|e_j|}{\frac{1}{T-m}\sum_{t=m+1}^{T}|Y_t-Y_{t-m}|}$

Donde m es la temporalidad de la serie.

2.6.5 AIC

Se calcula de la siguiente manera: $AIC = -2logL\left(\hat{\theta}\right) + 2k$. Donde k es el número de parámetros y n el número de datos.

2.6.6 AICc

Se calcula de la siguiente manera: $AICc = -2logL\left(\hat{\theta}\right) + 2k + \frac{2k+1}{n-k-1}$. Donde k es el número de parámetros y n el número de datos.

2.6.7 BIC

Se calcula de la siguiente manera: $BIC = -2logL\left(\hat{\theta}\right) + k \cdot log(n)$.

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

3 FUENTE(S) DE DATOS (SI YA ESTÁN RECOLECTADOS) O DISEÑO DEL ESTUDIO (SI NO ESTÁN RECOLECTADOS)

Los datos reales que se van a utilizar pertenecen a la Unidad de Estadísticas demográficas del Instituto Nacional de Estadística y Censos. Corresponden a las series históricas de la tasa de mortalidad infantil desde el año 1970 hasta la actualidad; así como la serie histórica de nacimientos ocurridos en Costa Rica desde 1950 hasta la actualidad.

4 DEFINICIÓN DE VARIABLE(S) DE ESTUDIO

Mortalidad infantil: Se define como la muerte de seres humanos que nacieron vivos y cuya defunción se dió antes de cumplir el primer año de edad.

Nacimientos: Se define el acto de terminar el periodo de gestación en la madre y llegar con vida al mundo.

5 EVIDENCIAS DE CALIDAD DE LA MEDICIÓN PARA LA(S) VARIABLE(S) DEL ESTUDIO

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, esto porque 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).

6 MODELO(S) ESTADÍSTICOS O TÉCNICA(S) ESTADÍSTICA(S) DE ANÁLISIS A EMPLEAR

6.1 Selección de ARIMA vía sobreparametrización

El método consiste en una selección fundamentada en las permutaciones de los parámetros de un modelo ARIMA, seleccionando la mejor especificación con base en medidas de rendimiento MAE, RMSE, MAPE y MASE: se comparan todos los posibles términos definiendo una diferenciación adecuada para la serie y permutando hasta un máximo determinado para los términos de especificación de un $ARIMA(p, d, q)(P, D, Q)$

6.2 Estudio de simulación

A partir de datos reales o bien, valores aleatorios de una cierta distribución de probabilidad, se generarán n valores aleatorios que sigan un determinado proceso de series cronológicas.

Para generar un determinado proceso se deben fijar los valores de p, d, q en la parte no estacional y P, D, Q en la parte estacional de un modelo ARIMA, así como la temporalidad que se desea para la misma. Además, se ofrece la posibilidad de definir el valor de los coeficientes del modelo para cada orden del procese; por ejemplo, si se define un $ARIMA(2, 1, 1)(1, 1, 3)$, el 2 indica que se pueden fijar los valores de los coeficientes $AR(1)$ y $AR(2)$ en, digamos, ,2 y ,46 respectivamente; de forma análoga, pueden definirse los coeficientes $SMA(1)$, $SMA(2)$ y $SMA(3)$ en ,4, ,1 y ,3 respectivamente.

7 Software estadístico a utilizar

Se utilizará el lenguaje de programación R (R Core Team, 2019a) mediante su interfaz RStudio para todos los procesos relacionados con la estimación y la simulación. Se hará uso de funciones contenidas en los paquetes `tidyr` (Wickham & Henry, 2019), `dplyr` (Wickham et al., 2019) y `parallel` (R Core Team, 2019b).

8 REFERENCIAS

Adhikari, R., K, A. R., & Agrawal, R. K. (2013a). *An Introductory Study on Time Series Modeling and Forecasting* (p. 16). Recuperado de https://arxiv.org/ftp/arxiv/papers/1302/1302.6613.pdf

Adhikari, R., K, A. R., & Agrawal, R. K. (2013b). *An Introductory Study on Time Series Modeling and Forecasting* (pp. 42-45). Recuperado de https://arxiv.org/ftp/arxiv/papers/1302/1302.6613.pdf

Benesty, J., & Chen, Y. C., J.and Huang. (2009). Pearson Correlation Coefficient. En *Noise Reduction in Speech Processing* (pp. 37-38). https://doi.org/10.1007/978-3-642-00296-0_5

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. G. (1956). *Exponential Smoothing for Predicting Demand.* Recuperado de https://www.industrydocuments.ucsf.edu/docs/jzlc0130

Goodrich, R. (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

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. (2011a). *Introducción a las Series Cronológicas.* Recuperado de http://www.editorial.ucr.ac.cr/ciencias-naturales-y-exactas/item/1985-introduccion-a-las-series-cronologicas.html

Hernández, O. (2011b). *Introducción a las Series Cronológicas*. Recuperado de http://www.editorial.ucr.ac.cr/ciencias-naturales-y-exactas/item/1985-introduccion-a-las-series-cronologicas.html

Hernández, O. (2011c). *Introducción a las Series Cronológicas*. Recuperado de http://www.editorial.ucr.ac.cr/ciencias-naturales-y-exactas/item/1985-introduccion-a-las-series-cronologicas.html

Hernández, O. (2011d). *Introducción a las Series Cronológicas*. 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/documetos-biblioteca-virtual/repoblacev2017_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

R Core Team. (2019a). *R: A Language and Environment for Statistical Computing*. Recuperado de https://www.R-project.org/

R Core Team. (2019b). *R: A Language and Environment for Statistical Computing*. Recuperado de https://www.R-project.org/

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

Wickham, H., François, R., Henry, L., & Müller, K. (2019). *dplyr: A Grammar of Data Manipulation*. Recuperado de https://CRAN.R-project.org/package=dplyr

Wickham, H., & Henry, L. (2019). *tidyr: Tidy Messy Data*. Recuperado de https://CRAN.R-project.org/package=tidyr