

Defensa de Tesis para  
optar por el grado y  
título de Maestría  
Académica en  
Estadística

LA SOBREPAREMETRIZACIÓN EN EL ARIMA:  
UNA APLICACIÓN A DATOS COSTARRICENSES

César Gamboa Sanabria

# Agenda

 Introducción

 Fundamentos teóricos

 Metodología

 Principales resultados

 Conclusiones

# Progreso

Introducción

Fundamentos teóricos

Metodología

Principales resultados

Conclusiones

# Antecedentes

- Uso de datos temporales.

- Investigación académica.
- Finanzas.
- Investigación de mercados.
- Demografía.

- Proyecciones.

- Métodos de análisis

- Los modelos  $ARIMA(p, d, q)(P, D, Q)_s$
- Box-Jenkins.

# El problema

- La identificación.
  - Visual.
- Alternativas.
  - `auto.arima()`
  - Criterios de selección.
- Someter a prueba más modelos.

# Objetivos

## Objetivo general

Proponer un algoritmo alternativo más exhaustivo para la selección de modelos ARIMA mediante la sobreparametrización de los términos de la ecuación del ARIMA.

## Objetivos específicos

1. Generar los escenarios de estimación de los distintos modelos ARIMA mediante permutaciones de los términos  $(p, d, q)$  y  $(P, D, Q)$  para la estimación de los posibles procesos que gobiernan una determinada serie temporal.
2. Contrastar la precisión de la estimación así como la generación de pronósticos con otros métodos similares, aplicados en datos costarricenses.
3. Aplicar diversos métodos de validación en la estimación de procesos que gobiernan la serie cronológica.
4. Integrar el desarrollo de la metodología de análisis de series temporales en una librería del lenguaje estadístico R.



# Importancia y aportes del estudio

- ¿Campos que se benefician?
  - Demografía.
  - Actuarial.
  - Política pública.
- Comparación procesos que gobiernan la serie cronológica.

# Progreso

Introducción

Fundamentos teóricos

Metodología

Principales resultados

Conclusiones



# Fundamentos teóricos

- Qué es una serie de tiempo (Hipel & McLeod (1994)).
- Dos grandes formas de estudio en las series de tiempo:
  - Componentes (Hernández (2011)).
    - Tendencia-ciclo, estacionalidad, irregularidad.
  - Autocorrelaciones.
- Series aditivas o multiplicativas.
- Supuestos (Agrawal & Adhikari (2013)).
- Identificación del modelo (Hyndman & Athanasopoulos, 2018).

# Modelos ARIMA

- Modelos Autorregresivos (AR) (Box et al., 1994).
- Modelos de medias móviles (MA).
- Modelos Autorregresivos de Medias Móviles (ARMA).
- Modelos Autorregresivos Integrados de Medias Móviles (ARIMA).
- Funciones
  - Autocorrelación.
  - Autocorrelación parcial.
- Autocorrelogramas.
- La sobreparametrización y el análisis combinatorio.

# Progreso

Introducción

Fundamentos teóricos

Metodología

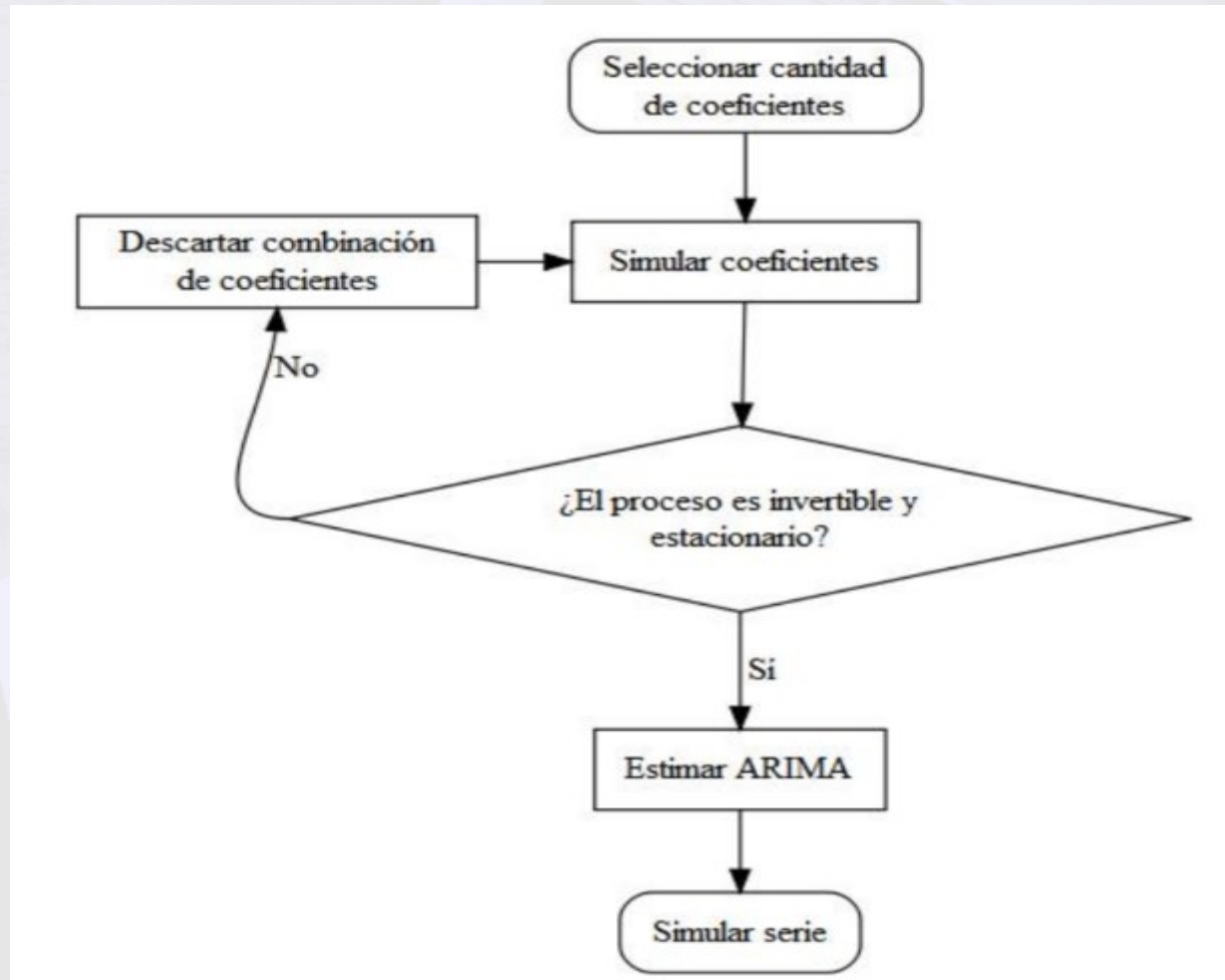
Principales resultados

Conclusiones

# Materiales

- Series reales.
  - Tasa de mortalidad infantil interanual.
  - Mortalidad por causa externa.
  - Incentivos salariales del sector público.
  - Intereses y comisiones del sector público.
- Herramientas analíticas y de procesamiento.
- Series simuladas.

# Simulación de series





# Métodos

- Análisis exploratorio.
- Partición de los datos.
- Estimación de modelos.
  - `auto.arima()`.
  - Sobreparametrización.
  - ARIMA estándar.
- Análisis de los errores.
- Pronósticos.
- Medidas de bondad de ajuste (Adhikari et al. (2013)).
  - AIC, AICc, BIC.
- Medidas de precisión.
  - MAE, MASE, RMSE.
- Tiempo de procesamiento.



# Progreso

Introducción

Fundamentos teóricos

Metodología

Principales resultados

Conclusiones

# Series simuladas: ajuste y precisión

**Cuadro 2**

Medidas de bondad de ajuste y de rendimiento según el método de estimación para los conjuntos de entrenamiento y validación simulados a partir de datos estacionales simulados

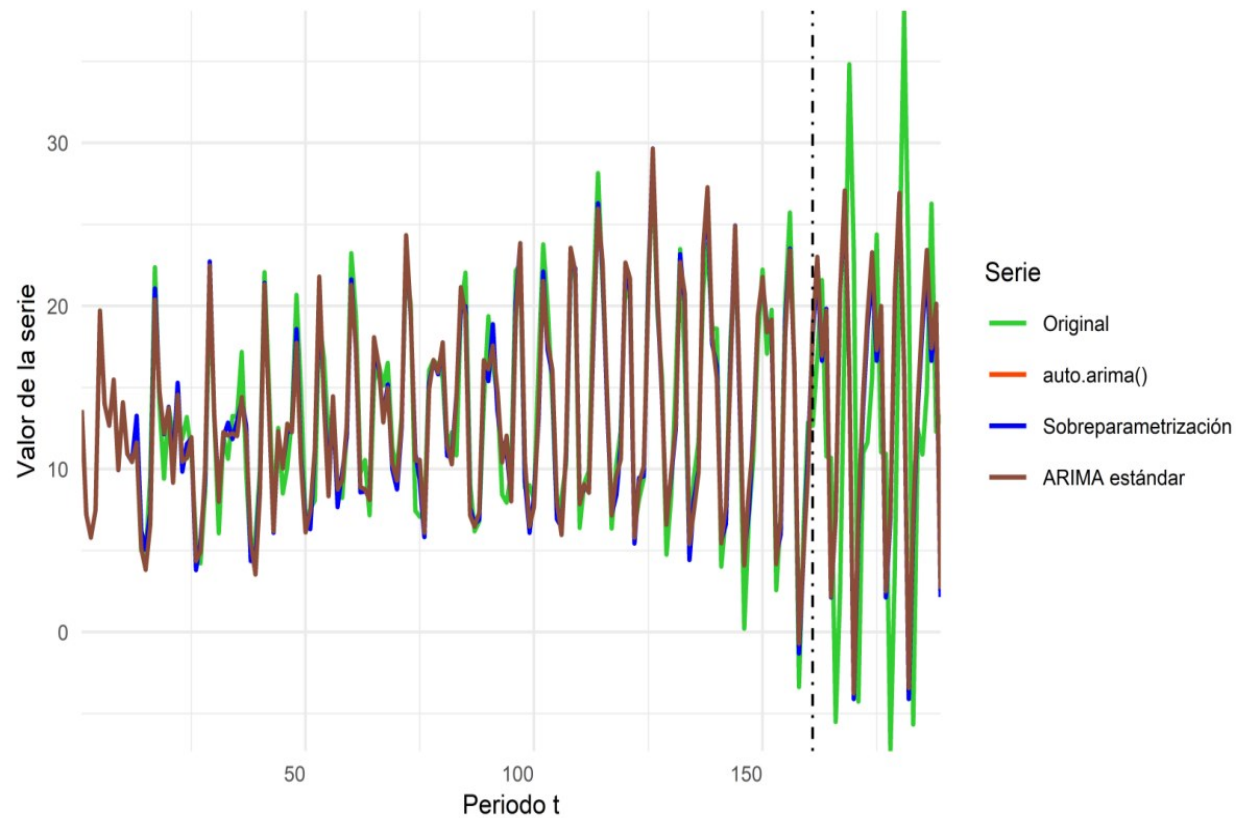
Proceso original	Datos	Estimación	AIC	AICc	BIC	RMSE	MAE	MAPE
ARIMA(0,0,1)(0,1,1)	Entrenamiento	auto.arima()	622,54	622,59	631,53	1,67	1,36	24,71
		ARIMA estándar	643,36	643,44	658,31	1,76	1,42	23,51
		Sobrep parametrización	622,54	622,59	631,53	1,67	1,36	24,71
	Validación	auto.arima()	119,17	119,45	123,27	5,6	4,68	49,72
		ARIMA estándar	124,81	125,28	131,64	5,82	4,97	53,61
		Sobrep parametrización	119,17	119,45	123,27	5,6	4,68	49,72
ARIMA(2,1,4)(3,0,3)	Entrenamiento	auto.arima()	632,18	632,28	653,11	1,68	1,34	6,69
		ARIMA estándar	678,92	679	693,87	1,97	1,57	7,84
		Sobrep parametrización	626,06	626,19	649,99	1,64	1,29	6,35
	Validación	auto.arima()	317,01	317,72	326,58	32,1	30,61	13,34
		ARIMA estándar	374,03	374,51	380,87	42,99	41,46	18,15
		Sobrep parametrización	272,98	273,83	283,92	24,72	23,24	10,08

Fuente: elaboración propia a partir de datos simulados

# Series simuladas: pronósticos

Figura 18

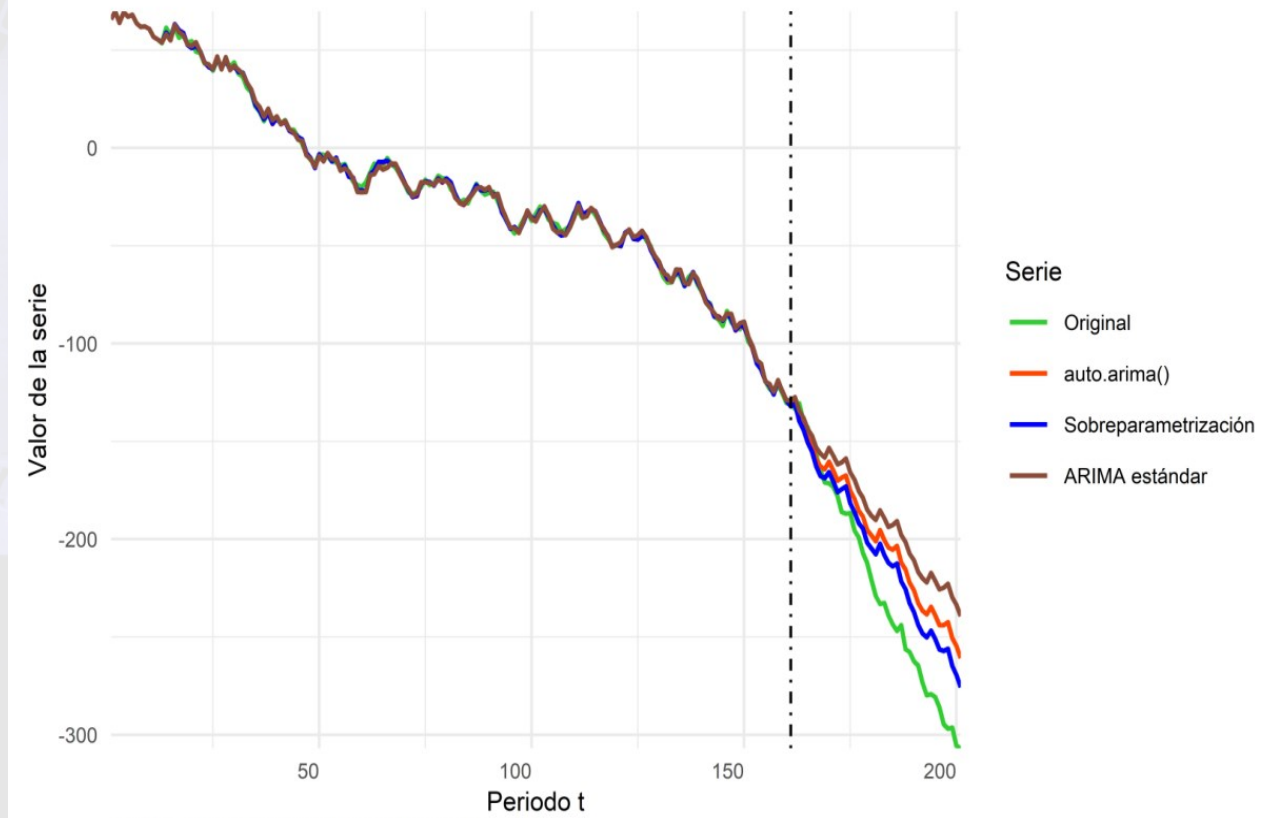
Pronóstico de los datos generados mediante un  $\text{ARIMA}(0,0,1)(0,1,1)$  según el método de estimación



Fuente: elaboración propia a partir de datos simulados  
auto.arima():  $\text{ARIMA}(0,0,1)(0,1,1)$   
Sobreparametrización:  $\text{ARIMA}(0,0,1)(0,1,1)$   
ARIMA estándar:  $\text{ARIMA}(1,1,1)(1,1,1)$

Figura 22

Pronóstico de los datos generados mediante un  $\text{ARIMA}(2,1,4)(3,0,3)$  según el método de estimación



Fuente: elaboración propia a partir de datos simulados  
auto.arima():  $\text{ARIMA}(1,1,2)(2,1,1)$   
Sobreparametrización:  $\text{ARIMA}(1,1,2)(2,1,2)$   
ARIMA estándar:  $\text{ARIMA}(1,1,1)(1,1,1)$

# Series reales: ajuste y precisión

**Cuadro 4**

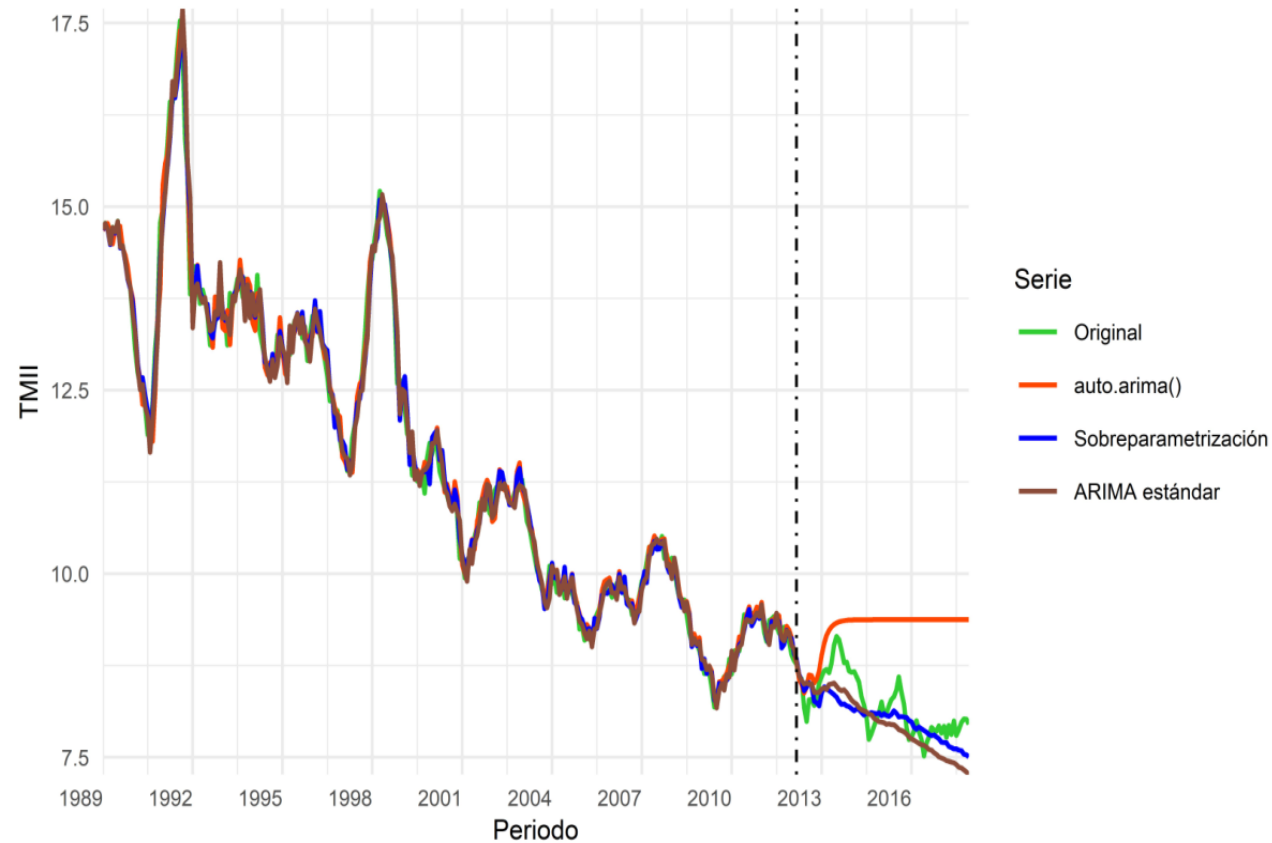
Medidas de bondad de ajuste y de rendimiento según el método de estimación para los conjuntos de entrenamiento y validación a partir de las series cronológicas reales

Proceso original	Datos	Estimación	AIC	AICc	BIC	RMSE	MAE	MAPE
TMII	Entrenamiento	auto.arima()	-88,34	-88,31	-73,84	0,21	0,16	1,29
		ARIMA estándar	-21,99	-21,94	-4,09	0,24	0,17	1,38
		Sobrep parametrización	-50,66	-50,52	10,26	0,22	0,16	1,33
	Validación	auto.arima()	369,38	369,52	378,37	1,16	1,03	12,69
		ARIMA estándar	78,11	78,28	89,35	0,37	0,32	3,7
		Sobrep parametrización	79,87	80,54	118,09	0,34	0,27	3,17
		auto.arima()	1615,68	1615,81	1624,67	4375,95	2349,42	6,84
		ARIMA estándar	1615,21	1615,39	1626,38	4310,66	2491,66	7,51
		Sobrep parametrización	1617,05	1617,22	1628,22	4359,18	2564,76	8,07
Incentivos Salariales	Validación	auto.arima()	407,12	407,72	411,11	5212,81	3701,08	4,51
		ARIMA estándar	397,69	398,48	402,67	3917,41	3011,28	3,94
		Sobrep parametrización	392,95	393,73	397,93	3476,92	2846,98	4,01

Fuente: elaboración propia a partir de datos simulados

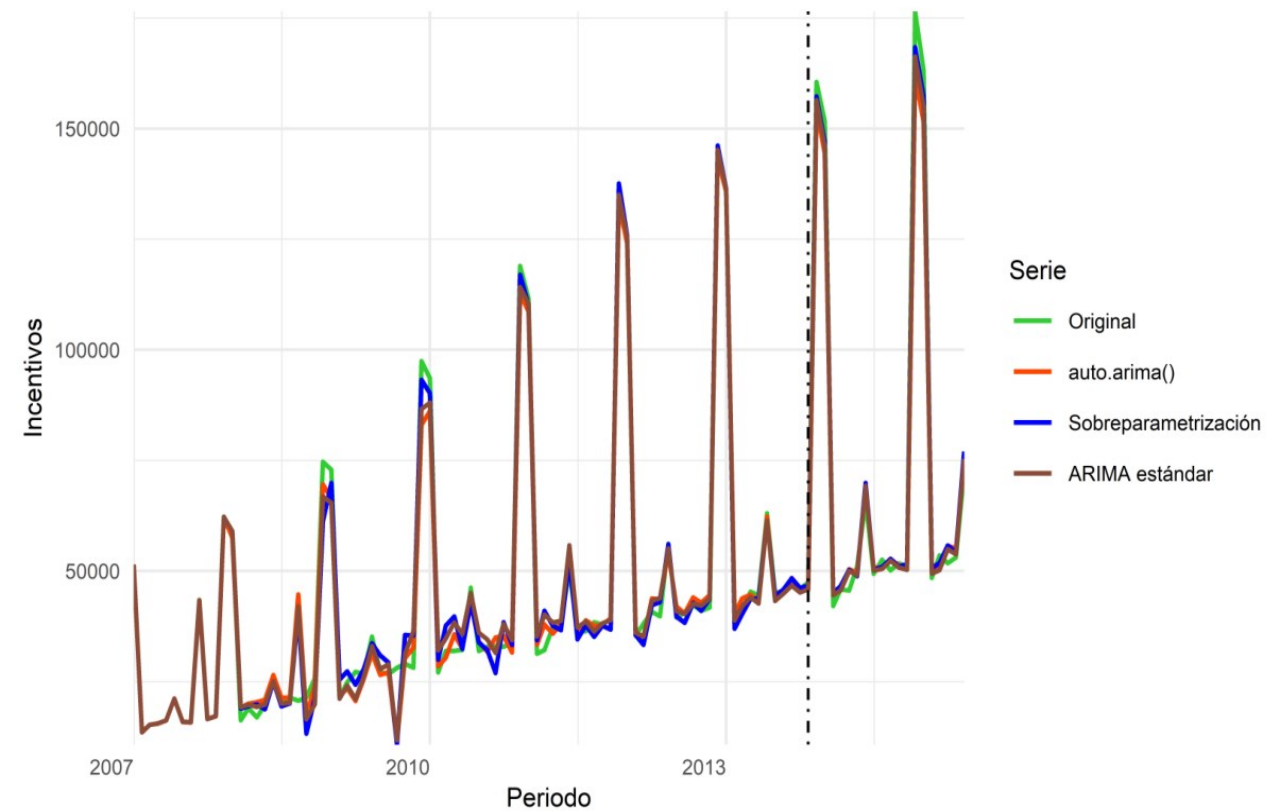
# Series reales: pronósticos

**Figura 39**  
Pronóstico de la TMII según el método de estimación



Fuente: elaboración propia.  
auto.arima(): ARIMA(2,1,0)(0,0,1).  
Sobreparametrización: ARIMA(4,1,0)(4,1,0).  
ARIMA estándar: ARIMA(1,1,1)(1,1,1).

**Figura 40**  
Pronóstico de la serie de incentivos salariales del sector público según el método de estimación



Fuente: elaboración propia.  
auto.arima(): ARIMA(0,0,1)(1,1,0).  
Sobreparametrización: ARIMA(2,1,0)(1,2,0).  
ARIMA estándar: ARIMA(1,1,1)(1,1,1).



# Resumen de resultados

- ARIMA(1,0,0)
- ARIMA(1,0,1)
- ARIMA(2,0,3)
- ARIMA(4,0,2)
- ARIMA(0,0,1)(0,1,1)\_12
- ARIMA(2,1,4)(3,0,3)\_12
- TMII
- Tasa de mortalidad por causa externa
- Incentivos salariales
- Intereses y comisiones



# Mejores resultados por método

## Cuadro 5

Distribución porcentual de los métodos de estimación que alcanzaron los mejores resultados según conjunto de datos y tipo de medición

Conjunto de datos	Medidas	auto.arima()	ARIMA estándar	Sobrep parametrización
Entrenamiento	Bondad de ajuste	33,34	8,34	58,34
	Precisión	45,46	9,1	45,46
Validación	Bondad de ajuste	25,01	25,01	50,01
	Precisión	33,34	0,01	66,68

Fuente: elaboración propia a partir de los resultados obtenidos

# Tiempos de estimación

## Cuadro 6

Tiempos de estimación en minutos para cada modelo según su tipo de estimación

Serie	autoarima	Sobrep parametrización	ARIMA estándar
ARIMA(1,0,0)	0,1056	8,236	0,0064
ARIMA(1,0,1)	0,0425	8,1266	0,0045
ARIMA(2,0,3)	0,0783	5,2904	0,0044
ARIMA(4,0,2)	0,1097	6,7233	0,0047
ARIMA(0,0,1)(0,1,1)[12]	0,1625	39,5444	0,2976
ARIMA(1,1,0)(1,1,0)[12]	0,0794	23,4079	0,3153
ARIMA(2,1,4)(4,1,4)[12]	0,1296	16,04	0,1951
ARIMA(2,1,4)(3,0,3)[12]	5,088	26,8193	0,1979
TMII	3,919	53,1779	0,3798
EXTERNA	0,1412	46,0911	0,2142
INCENTIVOS	2,8172	21,4405	0,1049
INTERESES	0,6145	32,542	0,108

Fuente: elaboración propia

# Progreso

Introducción

Fundamentos teóricos

Metodología

Principales resultados

Conclusiones

# Conclusiones

- Se evalúan más alternativas.
- Iguales o mejores resultados.
- Buen comportamiento en procesos de mayor orden.
- Procesamiento.
- Mejoras a futuro.
- Validación del método.
- Otras particiones.
- Disponibilidad en CRAN: [popstudy](#).

# Referencias

- 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>
- Rosero-Bixby, L. (2018). *Producto c para SUPEN. Proyección de la mortalidad de Costa Rica 2015-2150*. CCP-UCR. <http://srv-website.cloudapp.net/documents/10179/999061/Nota+t%C3%A9cnica+tablas+de+vida+segunda+parte>
- Hipel, K. W., & McLeod, A. I. (1994). *Time series modelling of water resources and environmental systems*. Elsevier Science. <https://books.google.co.cr/books?id=t1zG8OUbgdgC>
- Hernández, O. (2011). *Introducción a las series cronológicas* (1st ed.). Editorial Universidad de Costa Rica. <http://www.editorial.ucr.ac.cr/ciencias-naturales-y-exactas/item/1985-introduccion-a-las-series-cronologicas.html>
- Agrawal, R., & Adhikari, R. (2013). An introductory study on time series modeling and forecasting. *Nova York: CoRR*.
- Hyndman, R. J., & Athanasopoulos, G. (2018b). *Forecasting: Principles and practice*. OTexts. [https://books.google.co.cr/books?id=\\_bBhDwAAQBAJ](https://books.google.co.cr/books?id=_bBhDwAAQBAJ)
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control*. Prentice Hall. <https://books.google.co.cr/books?id=sRzvAAAAMAAJ>
- Adhikari, R., K, A. R., & Agrawal, R. K. (2013). *An introductory study on time series modeling and forecasting* (pp. 42–45). Lap Lambert Academic Publishing GmbH KG. <https://arxiv.org/ftp/arxiv/papers/1302/1302.6613.pdf>

Muchas gracias por su  
atención.

---