

# Modelos ARMA-SARMA para los residuos estructurales

Trabajo 2. Técnicas de Pronóstico.

Equipo 6

Santiago Franco Valencia  
Sergio Johan Giraldo  
Universidad Nacional de Colombia

1) 25/25  
2) 25/25  
3) 25/25  
4) 25/25

100x5/100=5.0

Nota: 5.0

## Introducción

Las series de tiempos definidas como una sucesión finita de variables aleatorias, ordenadas de acuerdo en una unidad de tiempo se constituyen como uno de los tipos de información más significativo para el análisis. El objeto de estudio de este trabajo es la serie de tiempo denominada Empleo Hombres Suiza, para la cual se aplicaron en un primer trabajo dos técnicas de pronósticos, que analizaron componentes de la tendencia y la estacionalidad: Método Cúbico+ Indicadoras y Modelo ETS-AAA. El objeto de este trabajo es extraer toda la información posible contenida en los datos para el desarrollo de los análisis y pronósticos sobre el futuro comportamiento de las series, pero esta vez se aprovecharán los residuos estructurales de los modelos de componentes con la aplicación de modelos especiales como los ARMA y SARMA, que tienen la capacidad de mejorar los pronósticos estructurales de los modelos iniciales.

La metodología para el estudio comprende dos fases. Una fase inicial de validación de las correlaciones de los residuos estructurales, a partir del diagnóstico gráfico con las bandas de Barlett y la aplicación de pruebas Ljung-Box y Durbin-Watson; validación que nos permitirá hallar evidencia para poder realizar la segunda fase de pronóstico con el uso de los residuos estructurales.

La segunda fase corresponde al tratamiento y modelación de los datos a través de los dos métodos elegidos con la aplicación software R.

Para el análisis, evaluación de los resultados finales y construcción de las conclusiones de los pronósticos de cada uno de los modelos se aplicarán los estadísticos de calidad (RMSE, MAPE y U-Theil).

## Descripción de la serie de Tiempo

Información tomada de la Oficina Federal de Estadísticas de Suiza, que comprende los **datos trimestrales de la tasa de empleo para los hombres en Suiza, en un espectro temporal desde julio 1991 y abril de 2020**. En la base de datos la variable o índice está expresado en términos de miles de personas,

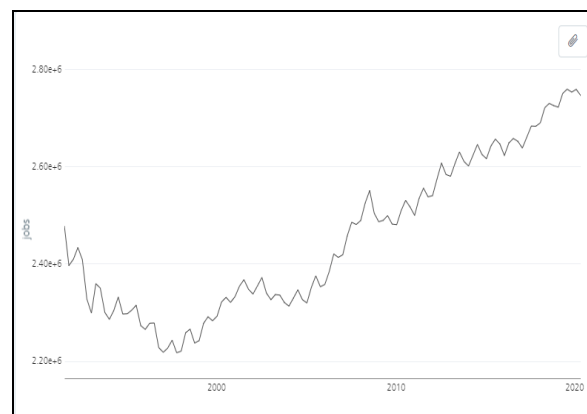
## Análisis gráfico de la Serie de Tiempo

**Tendencia:** En el último trimestre de 1997 observamos la ruptura de un comportamiento decreciente para iniciar una con comportamiento contrario: creciente a largo plazo, por lo que se afirma que la serie de tiempo tiene un patrón de tendencia. A primera observación se puede inferir una función bien definida del tiempo y que puede ser cuadrática, una tendencia determinística.

Respecto al **patrón estacional** intuimos que las observaciones a partir del último trimestre de 2007 se repiten regularmente en ciertas épocas del año, oscilan alrededor de la tendencia, configurando un patrón cíclico; factor significativo que tiene coherencia con la naturaleza del dato, que expresa como puede variar la demanda laboral respecto a los cambios cíclicos en la economía.

Claramente estas afirmaciones sobre la tendencia y la estacionalidad son derivadas del proceso empírico de observación e intuición, puesto que estos parámetros pueden ser identificados claramente a través de criterios estadísticos como los de análisis de auto correlación maestra.

**Figura1: Serie de tiempo total población empleada hombres en Suiza**



1) 25/25

## 1. Pruebas de Incorrección

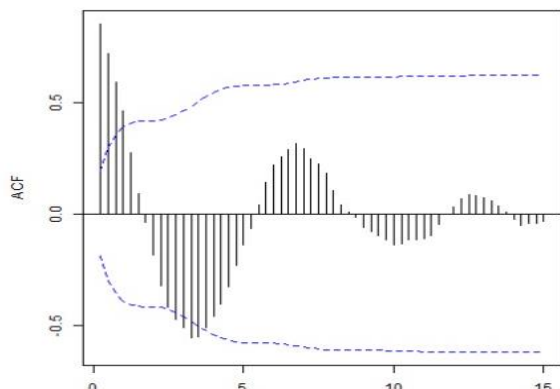
### Diagnóstico gráfico: Barras de Barlett

Este método ayuda a encontrar evidencia sobre la existencia de autocorrelaciones y para ello establece unos límites inferior y superior, definidos como  $(-2/\sqrt{T}, 2/\sqrt{T})$  con una probabilidad del 95%, que se expresan dentro de la gráfica de la fac (gráfica de correlaciones)

En el siguiente gráfico se aprecia que existen correlaciones que sobrepasan esas líneas límites de Barlett; inclusive en esos primeros rezagos sobrepasa en demasía las bandas por lo que podría afirmarse que las correlaciones son fuertes. Todo ello indica que hay evidencia de que los residuos estructurales no son ruido blanco y que poseen una dinámica autocorrelacionada. Este primer diagnóstico indica que la serie podrá ser modelada con los modelos especiales de ARMA o SARMA, aprovechando las autocorrelaciones de

los residuos distintos de 0 para mejorar los pronósticos estructurales.

Figura 2: Gráfica de correlaciones FAC de los residuos



### Prueba de Ljung-Box

Tabla 1: Reportes de valores -Prueba de Ljung Box.

X-Squared	Grados de libertad	Valor P
503.87	25	<2.2e-16

El valor de  $P < 0.05$  por tanto se rechaza la hipótesis nula de la prueba y se confirma que los residuos no son ruido blanco y evidencia algún tipo de autocorrelación de los residuos de la serie.

### Prueba de Durbin-Watson

Tabla 2: Reportes de valores -Prueba de Durbin-Watson.

DW	Valor P	Hipótesis Alternativa
0.25711	<2.2e-26	True correlation is greater than 0

El valor de  $P < 0.05$  por tanto se rechaza la hipótesis nula de la prueba y se confirma que los residuos del modelo de componentes están correlacionados.

### Conclusión Pruebas de In correlación

El resultado del uso de diagnóstico gráfico y de las pruebas de correlación certifican que hay evidencia de que los residuos estructurales del modelo de componentes no son ruido blanco y que poseen dinámica auto correlacionada: pueden ser utilizados para modelar la serie.

2) 25/25

## 2. Identificación de modelos ARMA-SARMA

Identificación con la función *auto.arima*, de la librería *Forecast*, en el software R:

Tabla 3: Reportes de coeficientes función *auto.arima*

ar1	ar2	sma1
0.9966	-0.1807	0.4191
0.0992	0.0993	0.0970

Sigma^2	1.44e+08
Log likelihood	-1145.67
AIC	2299.34
AICc	2299.74
BIC	2309.99

Se obtiene el siguiente modelo que corresponde a un modelo **SARMA(2,0,0)(0,0,1)[4] with zero mean**.

Al analizar la significancia de sus coeficientes se observa que el segundo rezago autorregresivo no es significativo, mientras que el primero sí;

Tabla 4: Reporte test de significancia de coeficientes

	Estimate	Std.Error	Z value	Pr(> z )
ar1	0.996570	0.099215	10.0446	<2.2e-16***
ar2	-0.180671	0.099321	-1.8191	0.0689
sma1	0.419075	0.097004	4.3202	1.559e-05***

disminuye el orden autorregresivo pero no el valor p. Por ello se propone disminuir en una unidad al Valor P asociado al rezago autorregresivo de este modelo obteniendo así el siguiente resultado

Tabla 5: Reportes de coeficientes función *auto.arima* modelo ii

ar1	sma1	Intercept
0.8442	0.4080	4596.008
0.0532	0.0941	10037.109

Sigma^2	1.44e+08
Log likelihood	-1147.19
AIC	2300.39

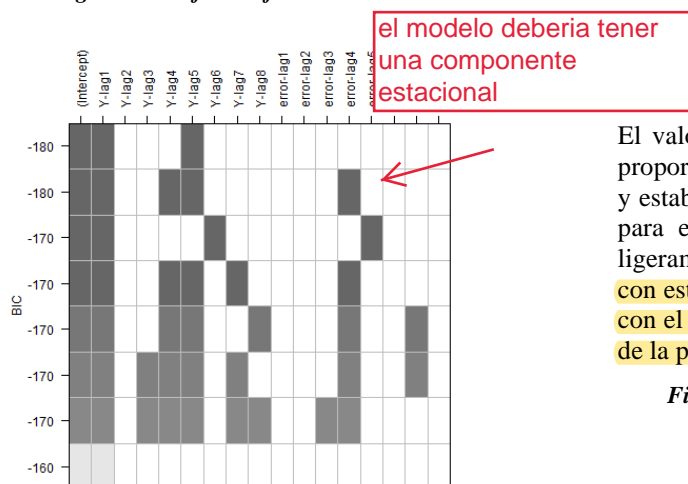
Tabla 6: Reporte test de significancia de coeficientes-modelo ii

	Estimate	Std.Error	Z value	Pr(> z )
ar1	8.4417e-01	5.3250e-02	15.8529	<2.2e-16***
sma1	4.0803e-01	9.4103e-02	4.3360	1.451e-05***
intercept	4.5960e+03	1.0037e+04	0.4579	0.647

Tras realizar este procedimiento de disminución en un grado al rezago autorregresivo se obtiene evidencia que todos los rezagos asociados al modelo son significativos, criterio que permite elegir este modelo, obtenido de depurar ligeramente al modelo conseguido mediante la función *auto.arima*

**Identificación con la función *armasubsets* ,de la librería *TSA*, en el software R:**

**Figura 3: Gráfica de función *armasubsets***



Desde la observación del gráfico resultante de esta función y bajo el principio de Parsimonia el único rezago significativo sería el primer rezago de la parte autorregresiva, ya que el grado próximo más alto significativo asociado al rezago autorregresivo sería el rezago quinto, si se tiene en cuenta que esta serie tiene estacionalidad  $S=4$ , por lo que el modelo elegido corresponde a:

**SARMA (1,0,0)(0,0,0)[4]"**

3) 25/25

**3.Estimación de los modelos con la función *auto.arima***

**Tabla 7: Reportes de coeficientes función *auto.arima***

ar1	intercept
0.8766	4271.338
0.0470	9661.168

Sigma^2	165867854
Log likelihood	-1154.25
AIC	2312.51

**Tabla 8: Reporte test de significancia de coeficientes**

	Estimate	Std.Error	Z value	Pr(> z )
ar1	8.7659e-01	4.6956e-02	18.6683	<2e-16***
intercept	4.2713e+03	9.6612e+03	0.4421	0.6584

Al observar los coeficientes asociados se consigue que el **único rezago autorregresivo que se incluye en el modelo es significativo**, por lo que sin vacilaciones se elige este modelo.

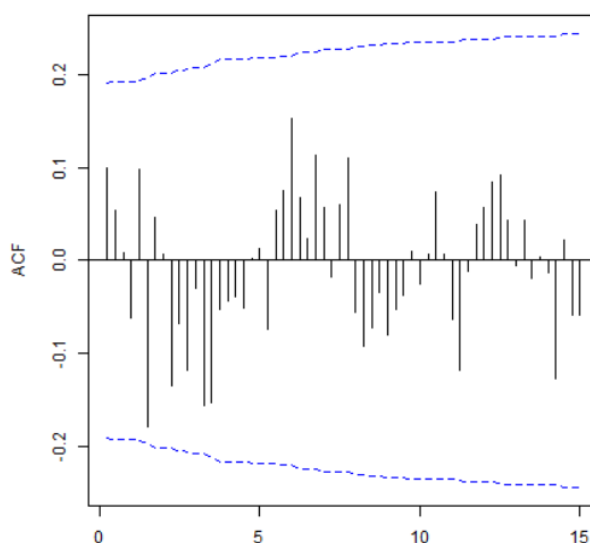
Cálculos de los valores AIC de los modelos

**Tabla 9 reporte de los valores AIC de los modelos**

SARMA(2,0,0)(0,0,1) [4]	2302.387
SARMA (1,0,0)(0,0,0)[4]	2314.508

El valor resultante del Criterio de Información de Akaike, proporciona un enfoque para comparar el ajuste de modelos y establece que debe elegirse aquel con el mínimo ACI, que para este caso corresponde a aquel obtenido de depurar ligeramente el modelo resultado de la función *auto.arima*; con este modelo se procederá a la validación de sus residuos con el diagnóstico gráfico de autocorrelación y la aplicación de la prueba de Ljung-Box. ✓

**Figura 4: Gráfica de correlaciones FAC de los residuos del modelo SARMA**



En el gráfico se aprecia que todas las autocorrelaciones se ubican dentro de las bandas de Barlett, lo que confirma que hay evidencia de que los residuos estructurales son ruido blanco, por lo tanto, los pronósticos sólo podrían realizarse con la parte estructural de la serie, ya que el pronóstico del ruido es 0.

**Prueba de Ljung-Box**

**Tabla 10: Reportes de valores -Prueba de Ljung Box-modelo SARMA.**

X-Squared	Grados de libertad	Valor P
24.172	25	0.5094

**El valor de  $P>0.05$**  por tanto no se tiene evidencia para rechazar la hipótesis nula de la prueba: los residuos obtenidos mediante el modelo SARMA son ruido blanco; resultado que coincide con el del diagnóstico gráfico. ✓

#### 4. Pronósticos

Cálculo de los pronósticos

**Tabla 11 Datos originales de la serie**

	Qtr1	Qtr2	Qtr3	Qtr4
<b>2018</b>	2691016	2722026	2731000	2726401
<b>2019</b>	2723137	2751394	2760200	2754092

Pronósticos modelo C

**Tabla 12 Pronósticos del modelo C**

Pronóstico 1	Pronóstico 2	Pronóstico 3	Pronóstico 4
2645224	2668375	2682328	2653585
Pronóstico 5	Pronóstico 6	Pronóstico 7	Pronóstico 8
2644423	2666107	2678557	2648278

Pronósticos modelo C+ARMA

**Tabla 13 Pronósticos del modelo C+ARMA**

Pronóstico 1	Pronóstico 2	Pronóstico 3	Pronóstico 4
2671813	2690677	2705497	2682808
Pronóstico 5	Pronóstico 6	Pronóstico 7	Pronóstico 8
2669808	2688252	2697968	2665380

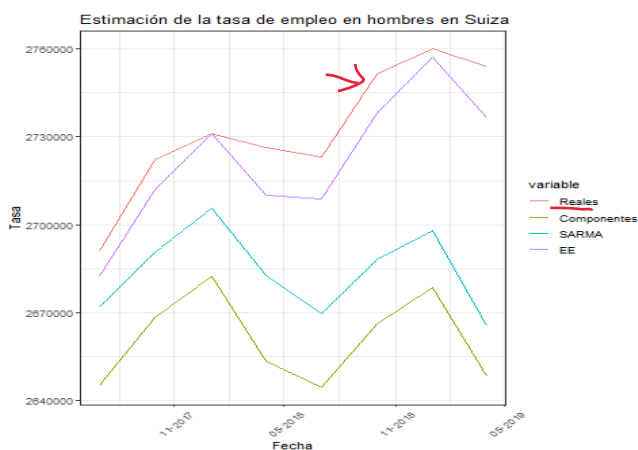
Pronósticos modelo EE

**Modelo ETS-AAA**

**Tabla 14 Pronósticos del modelo EE**

Pronóstico 1	Pronóstico 2	Pronóstico 3	Pronóstico 4
2682489	2711904	2731005	2710191
Pronóstico 5	Pronóstico 6	Pronóstico 7	Pronóstico 8
2708795	2738210	2757311	2736497

**Figura 5: Reporte gráfico de los valores de pronóstico de los 3 modelos**



El gráfico exhibe que entre todos los modelos el de mayores calidades de pronósticos, que se ajusta mejor a los datos reales de la serie, corresponde al modelo de Espacio de Estados. Con la inclusión del modelo SARMA se consigue un modelo intermedio que aparentemente arroja valores más consistentes que los obtenidos mediante el modelo de componentes, pero no lo suficiente como para competir con el modelo de Espacio de Estados.

Finalmente, para otorgar mayores soportes argumentativos y teórico a la conclusión inicial construida a partir del análisis gráfico, se computan los estadísticos de calidad RMSE, MAPE y Utheil:

**Tabla 5: Estadísticos de Calidad**

Modelo	RMSE	MAPE	Utheil
C	74143.86	2.6899376	0.027134178
C+SARMA	53004.75	1.8040858	0.019397969
EE	11913.17	<b>0.3810203</b>	0.004359824

Analisis de estadísticos de calidad confirman que el modelo de Espacio de Estados es el mejor modelo para pronosticar: de todos ellos presenta los menores valores en las medidas RMSE y MAPE, inclusive el estadístico U de Theil es el más cercano a 0, lo que determina que es deseable cuando se evalúa la buena precisión de los pronósticos.

El modelo de componentes+SARMA ocupa un valor intermedio en todas las medidas de calidad.

#### Bibliografía

Adhikari, R., & R. K., A. (s.f.). *An Introductory Study on Time Series Modeling and Forecasting*.

Dataserries.org. (s.f.). Obtenido de [http://www.dataserries.org/?main=ch\\_comb\\_jobs&split=596&select=tot,1,0&var=&freq=quarterly&start=&end=&trans=none](http://www.dataserries.org/?main=ch_comb_jobs&split=596&select=tot,1,0&var=&freq=quarterly&start=&end=&trans=none)

Giraldo, N. (s.f.). *Repositorio Institucional UNAL*. Obtenido de <https://repositorio.unal.edu.co/handle/unal/76272>

Hyndman, R., Koehler, A., Snyder, R., & Grose, S. (s.f.). *A state space framework*.

Kleinbaum, D., & Klein, M. (s.f.). *Survival Analysis*.

Springer.

donde se cita ésta?

Pérez Ramírez, F. (s.f.). *Métodos no paramétricos para el análisis de series de tiempo*. Medellín: Universidad de Medellín.

Universidad Autónoma de Madrid. (s.f.). *Instrumentos estadísticos avanzados. Facultad de Ciencias Económicas y Empresariales*. Obtenido de <https://www.estadistica.net/ECONOMETRIA/SERIES-TEMPORALES/modelo-arima.pdf>

mejorar la redaccion de referencias. Ejemplo referencia tipo APA. varias referencias tienen el año, no son s.f.

Schrager, D. F. (2006). Affine stochastic mortality. *Insurance: mathematics and economics*, 38(1), 81-97.

