



**PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
FACULTAD DE MATEMÁTICAS
DEPARTAMENTO DE ESTADÍSTICA**

Proyecto de demanda eléctrica ELECDA

Autor: Daniel Franzani Cerda.

1. Introducción

El presente trabajo busca modelar los datos enmarcados en la demanda eléctrica, específicamente los pertenecientes a **ELECDA – Crucero** [MWh]. La importancia de modelar este tipo de datos, radica en la infraestructura de alimentación eléctrica en la zona. Lo que permite tanto un consumo, abastecimiento necesario y ganancia retribuida. Para ello, se proponen 2 metodologías básicas las cuales son SARIMA y ARMAX. Adicionalmente se propone utilizar la metodología SARIMAX con el fin de aumentar las posibilidades de elegir un proceso el modelamiento de los datos.

Durante la confección de los modelos, se reportan procedimientos como la diferenciación ya sea para la tendencia o estacionalidad, la cual es analizadas mediante periodogramas. Del mismo modo, se dan a conocer los coeficientes de cada modelo y sus respectivos intervalos de estimación. Además, cada proceso es sometido a un análisis de residual para la validación de supuesto de estacionariedad e independiencia, para luego compararlos mediante criterios de selección como AIC, BIC y error de predicción.

2. Resumen de los datos

Los datos de **ELECDA – Crucero** [MWh] están comprendidos entre las fechas de enero 2009 a noviembre 2015. En la figura 1, la dispersión de los datos muestra una tendencia creciente a largo plazo. En el histograma apreciamos la distribución de los datos junto a la curva de densidad, se aprecia un comportamiento alejado de la normalidad (test de normalidad: $p - \text{value} = 0,01$). Respecto al boxplot, se aprecia una media de 55264,3 MWh y una caja ($Q_1 - Q_3$) asimétrica, lo cual es coincidente a lo observado en el histograma.

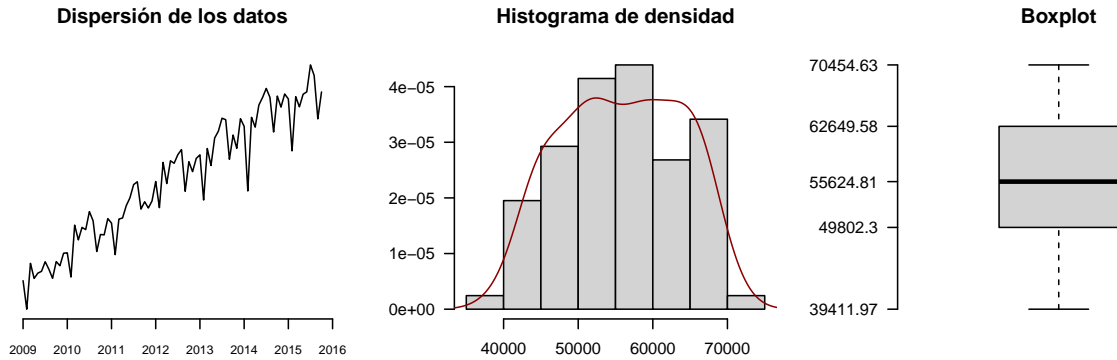


Figura 1: Distribución de los datos.

Para modelar los datos, se analiza en primer lugar la tendencia, para ello se realiza un ajuste de polinomios cúbicos locales (*splines*) (Figura 2), en el cual, se refleja una tendencia al alza. De un modo similar, se modela la estacionalidad de los datos mediante el *factor mes*, mostrando claramente una periodicidad anual (Figura 2).

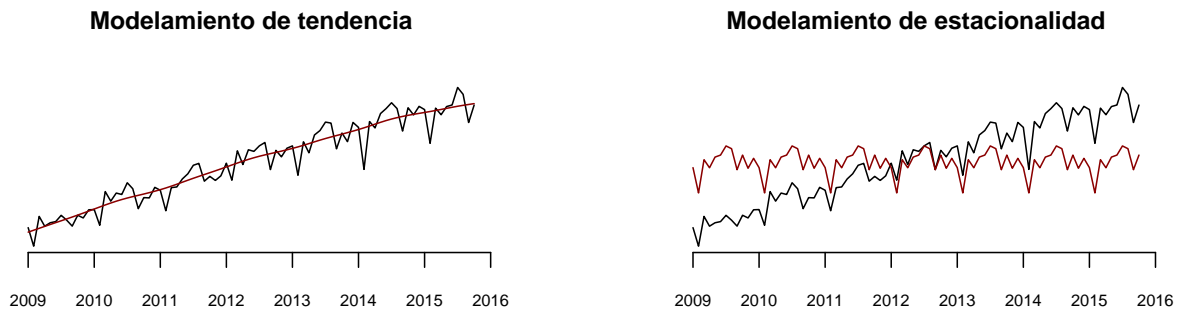


Figura 2: Modelo de tendencia y estacionalidad.

Por otro lado, se estudian los regresores PIB e IMACEC, los cuales que se podrían usar para explicar diversos procesos en el modelamiento de los datos. En la figura 3, se muestran los histogramas de cada una de variables. Apreciamos que ambos casos las variables no tienen un comportamiento normal.

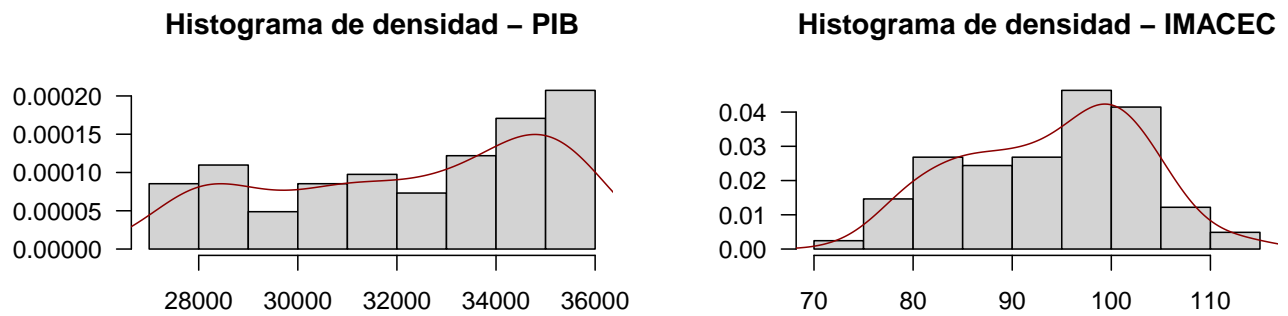


Figura 3: PIB - IMACEC.

3. Análisis Estadístico

3.1. Propuesta de modelos

Para modelar los datos se abarca una propuesta SARIMA y otra ARMAX. Estos modelos son adecuados dada la fuerte estacionalidad presente en los datos. Así, una serie X_t es un proceso SARIMA $(p, d, q) \times (P, D, Q)_s$ si, $Y_t = (1 - B)^d(1 - B^D)X_t$ es un proceso ARMA definido como

$$\phi(B)\xi(B^s)Y_t = \theta(B)\Theta(B^s)\varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2)$$

En el cual, el miembro izquierdo de la igualdad corresponde al proceso ARMA regular (p, d, q) de los datos y, el miembro derecho corresponde al proceso ARMA estacional $(P, D, Q)_s$.

En primer lugar, se evalúa la presencia tendencia con el fin de aplicar diferenciación. En la figura 4, se aprecia el periodograma de los datos, el cual tiende a infinito positivo cuando la frecuencia tiende a cero, siendo esto evidencia de la presencia de tendencia. Al aplicar el operador $(1 - B)$, en el periodograma del centro se aprecia el comportamiento estacional. Finalmente, al aplicar el operador $(1 - B^{12})$ se observa que se elimina la estacionalidad presente en los datos.

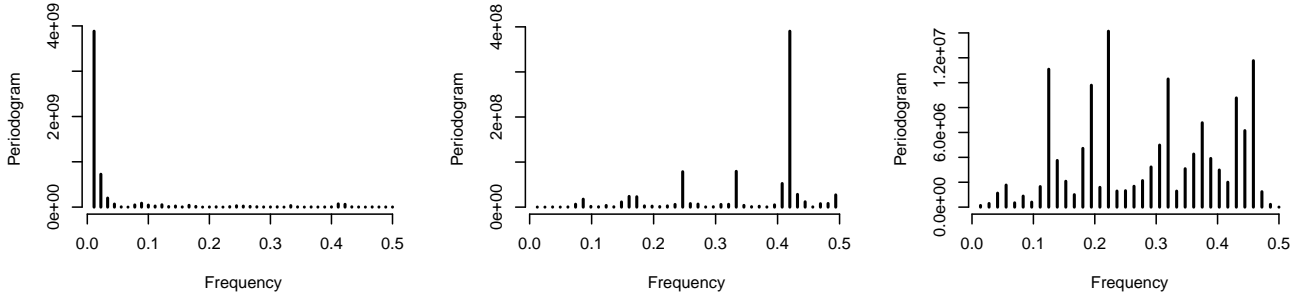


Figura 4: Periodogramas

Al eliminar tendencia y estacionalidad, se aplican a los datos las funciones ACF y PACF. Esto sugiere un proceso ARMA(1,1) regular y, un proceso MA(1) estacional, tal como se muestra en la figura 5.

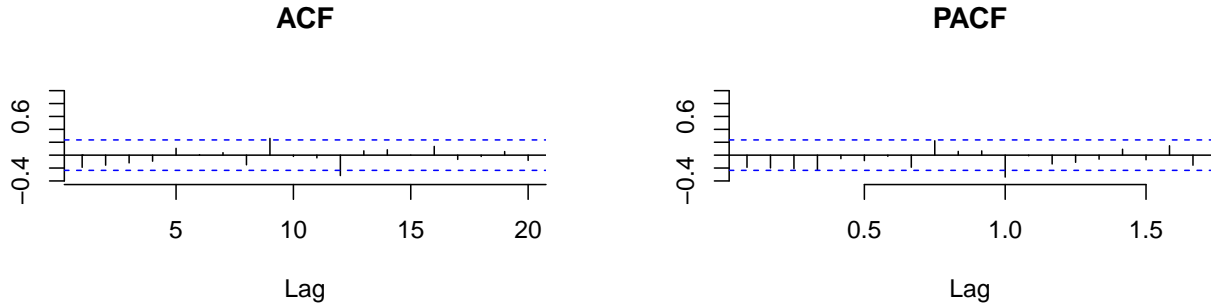


Figura 5: ACF y PACF

El modelo ajustado tiene todos los coeficientes significativos (mayores a 2), tal como se muestra en la tabla 1.

Parámetros	ar1	ma1	sma1
Estimados	0.4519	-0.9053	-0.4337
Desviaciones	0.1366	0.0752	0.1276
Cuantil T-student	3.31	12.05	3.40

Tabla 1: Coeficientes proceso SARIMA

De este modo, en la figura 6 se aprecia que los residuos cumplen con el supuesto de blancura (primera fila), sin embargo muestra una estructura en decaimiento en las observaciones 9 y 11. Es por esto, que corrigiendo bajo interacciones se obtiene el modelo SARIMA $(8, 1, 4) \times (0, 1, 1)_s$, el cual muestra un mejor estructura de independencia en ambos gráficos (segunda fila).

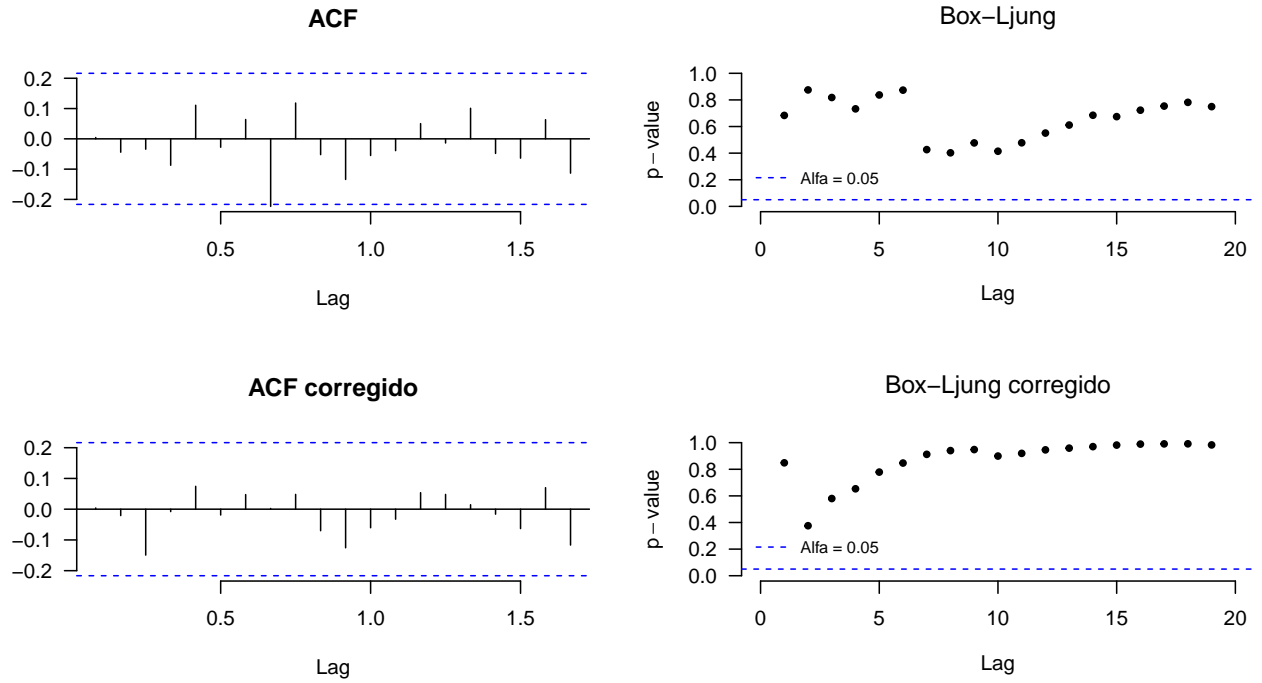


Figura 6: ACF y Box – Ljung, residual

Los coeficientes del modelo corregido son todos significativos, tal como se muestra en la tabla 2.

Parámetros	ar1	ar8	ma2	ma4	sma1
Estimados	-0.3177	-0.2542	-0.5029	-0.2753	-0.49171
Desviaciones	0.1159	0.1146	0.1312	0.1267	0.1289
Cuantil T-student	2.7411	2.2176	3.8327	2.173	3.8148
IC ($\alpha = 0,5$)	$(-0,55; -0,09)$	$(-0,48; -0,03)$	$(-0,76; -0,25)$	$(-0,52; -0,03)$	$(-0,74; -0,24)$

Tabla 2: Coeficientes proceso SARIMA corregido

En la figura 7, se puede observar que los residuos del modelo SARIMA (ya corregido) presentan observaciones fueran de las bandas ± 2 y ± 3 lo cuales corresponde a $\{02 - 2012, 07 - 2013\}$ y $\{08 - 2014\}$ respectivamente. Por otro lado, el gráfico de cuantiles muestra normalidad en los residuos, así lo confirma el test Shapiro – Wilk con un valor – p igual a 0.1909, por lo que, no se rechaza la hipótesis nula de normalidad con un 95 % de confianza. Adicionalmente, en la tabla 3, se muestran los estadísticos para los supuestos de estacionariedad e independencia serial.

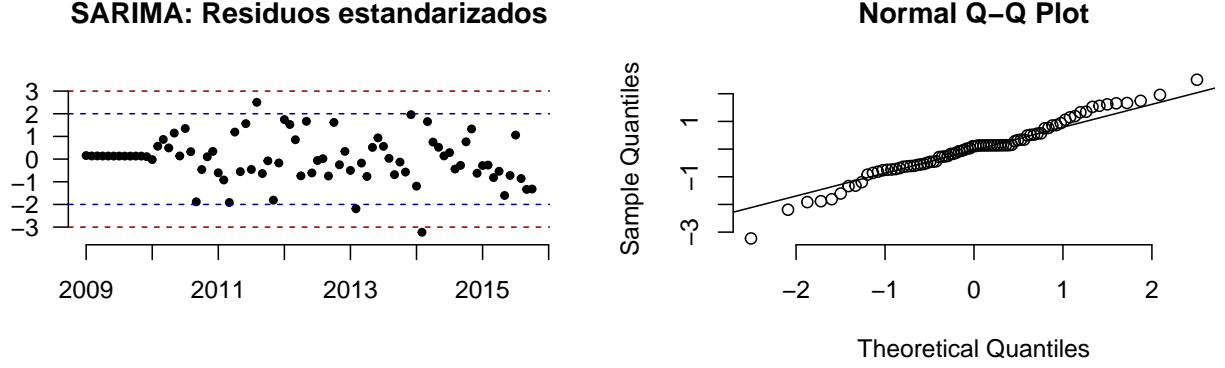


Figura 7: Residuos, SARIMA

Residuos SARIMA	Box - Ljung	Box-Pierce	Breusch-Pagan	lm ($H_0 : \beta_1 = 0$)
Valor-p	0,9714	0,9719	0,2029	0,101

Tabla 3: Box test: Box-Ljung y Box-Pierce, Breusch-Pagan y lm ($H_0 : \beta_1 = 0$), al 95 % de confianza

Es claro, que tanto en el gráfico ACF, en el de valores-p del estadístico Box-Ljung y en los test para hipótesis de supuestos, se confirma la independencia serial con un 95 % de confianza, al igual que para el test de varianza y media constante.

Una propuesta distinta es el proceso ARMAX, el cuál se define como un proceso ARMA regular mediante el uso de un regresor. Para ello, se utilizan los dos regresores mencionados en el resumen de los datos. Sin embargo, ambos se plantean en dos regresiones.

- PIB: $MWh \sim \beta_1 PIB + \beta_2 I_{Mes(j), j=1, \dots, 12} + \beta_3 \text{Año}$
- IMACEC: $MWh \sim \beta_1 IMACEC + \beta_2 I_{Mes(j), j=1, \dots, 12} + \beta_3 \text{Año}$

Analizando los residuos de ambas regresiones, se obtiene que en ambos casos el modelo que ajusta adecuadamente es un proceso ARMA(1,1), sin embargo, bajo interacciones se logra mejorar el modelo obteniendo en ambos casos un proceso ARMA(11,0,3). En la figura 8 se muestran los valores-p asociados al test de Box - Ljung, además en la tabla 4 se muestran los parámetros estimados.

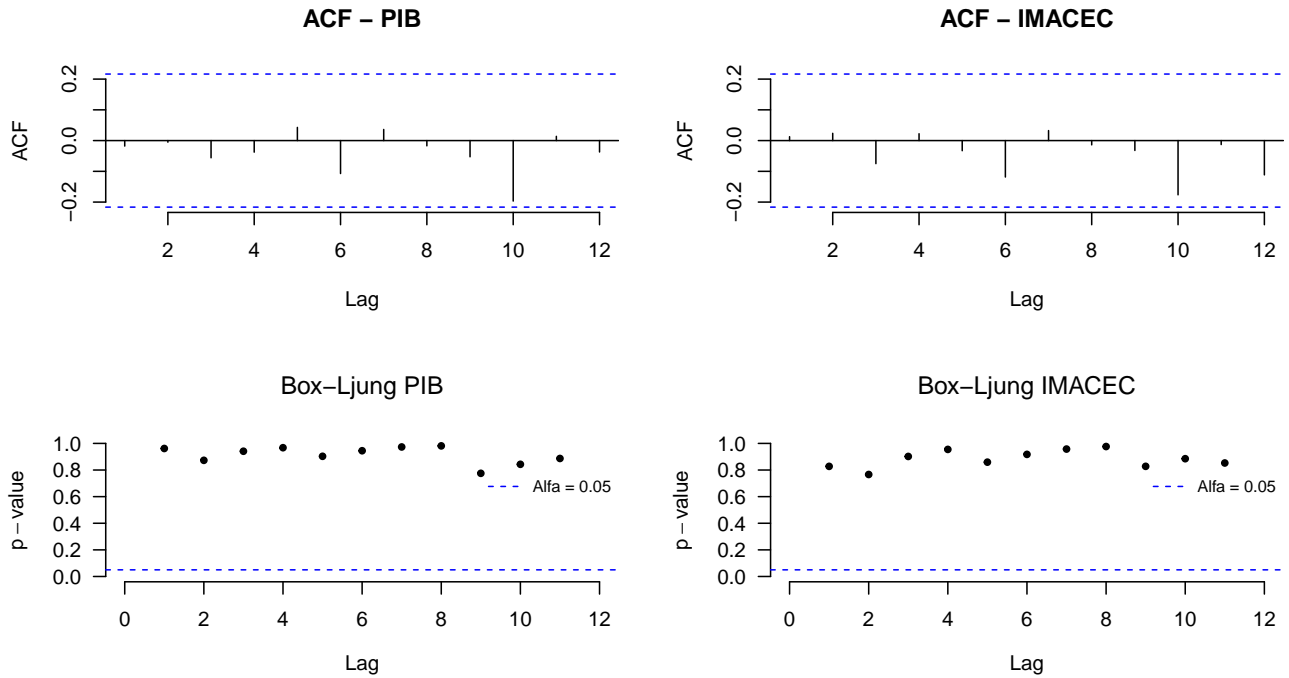


Figura 8: Regresores PIB - IMACEC, análisis residual

Parámetros	ar3	ar8	ar11	ma1	ma3
Estimados PIB	-0.3330	-0.3585	-0.3084	0.5845	0.4145
Desviaciones	0.1249	0.1058	0.1153	0.1013	0.0957
Cuantil T-student	2.6673	3.389	2.6747	5.7695	4.3326
IC ($\alpha = 0,5$)	(-0,58; -0,09)	(-0,57; -0,15)	(-0,54; -0,08)	(0,38; 0,78)	(0,22; 0,60)
Estimados IMACEC	-0.4234	-0.3243	-0.2450	0.5134	0.4879
Desviaciones	0.1211	0.1094	0.1199	0.1055	0.1016
Cuantil T-student	3.4978	2.9646	2.0443	4.8647	4.8009
IC ($\alpha = 0,5$)	(-0,66; -0,18)	(-0,53; -0,11)	(-0,48; -0,02)	(0,30; 0,72)	(0,29; 0,69)

Tabla 4: Coeficientes proceso ARMAX, PIB, IMACEC

En la figura 9, se muestran el análisis supuestos residuales, en el cual observamos que los residuos superior a ± 2 están presentes solo en IMACEC en {03-2011}. En el caso de los residuos ± 3 , estos están en ambos y corresponden a la misma observación {08-2014}. Además, ambos residuos no rechazan las pruebas de normalidad Shapiro-Wilk con un valor - p de 0,3473 y 0,2102 respectivamente. Por otra parte, en la tabla 5 se muestran los estadísticos para los supuestos de estacionariedad e independencia.

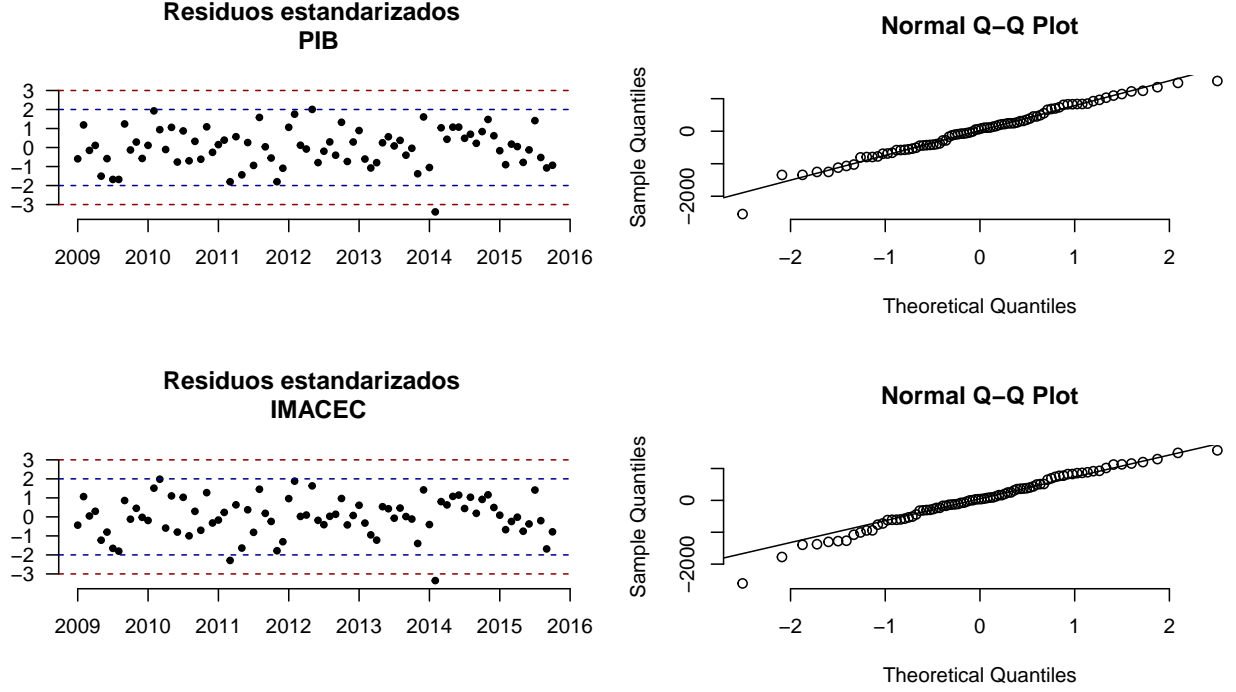


Figura 9: Residuos - Regresores PIB, IMACEC

Residuos	Box - Ljung	Box-Pierce	Breusch-Pagan	$\text{lm} (H_0 : \beta_1 = 0)$
PIB: Valor-p	0,8691	0,8714	0,9943	0,714
IMACEC: Valor-p	0,9073	0,909	0,8321	0,746

Tabla 5: Box test: Box-Ljung y Box-Pierce, Breusch-Pagan y $\text{lm} (H_0 : \beta_1 = 0)$, al 95 % de confianza

Es claro, que tanto en el gráfico ACF, en el de valores-p del estadístico Box-Ljung y en los test para hipótesis de supuestos, se confirma la independencia serial con un 95 % de confianza, al igual que para el test de varianza y media constante, tanto para el modelo con regresor PIB como para el modelo con regresor IMACEC.

Finalmente, para ampliar la gama de modelos se propone alternativamente un modelo SARIMAX, con el finde generar una metodología innovadora en el modelamiento de los datos. Este proceso esta construido con el regresor PIB. El modelo propuesto, tiene la estructura $(2, 0, 8) \times (0, 1, 1)_s$, en la figura 10 se muestran los periodogramas (solo con regresor, y con diferenciacion $(1 - B^{12})$), los cuales muestras finalmente una clara estructura libre de tendencia y estacionalidad.

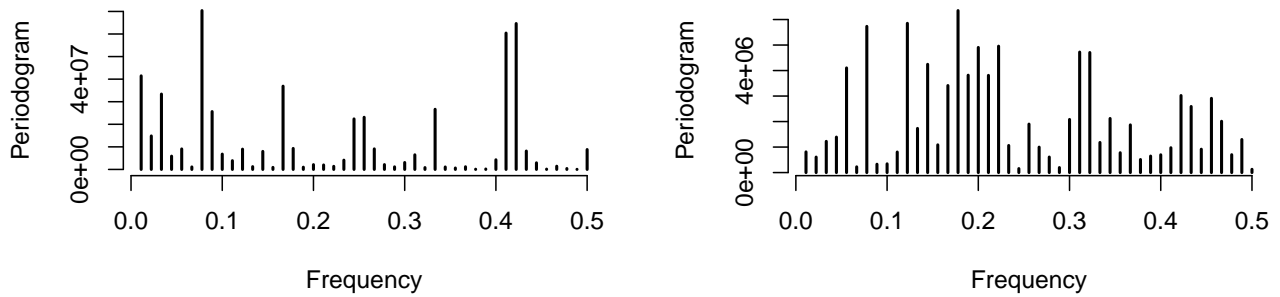


Figura 10: SARIMAX, periodogramas

Por otro lado, los gráficos ACF y valores - p del estadístico Box - Ljung (Figura 11) muestran independencia residual, a pesar de estar cerca del límite en el estadístico Box - Ljung.

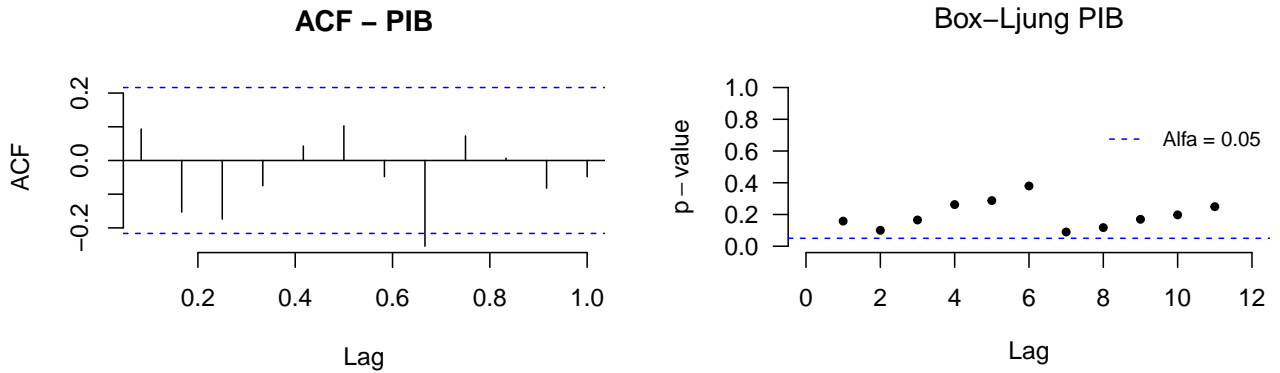


Figura 11: SARIMAX, análisis residual

En la figura 12, se muestran el análisis supuestos residuales, en el cual observamos que los residuos superior a ± 2 están presentes solo en IMACEC en {03-2011}. En el caso de los residuos ± 3 , estos están en ambos y corresponden a la misma observación {08-2014}. Además, los residuos no rechazan la prueba de normalidad Shapiro-Wilk con un valor - p de 0,1156 Por otra parte, en la tabla 6 se muestran los estadísticos para los supuestos de estacionariedad e independencia.

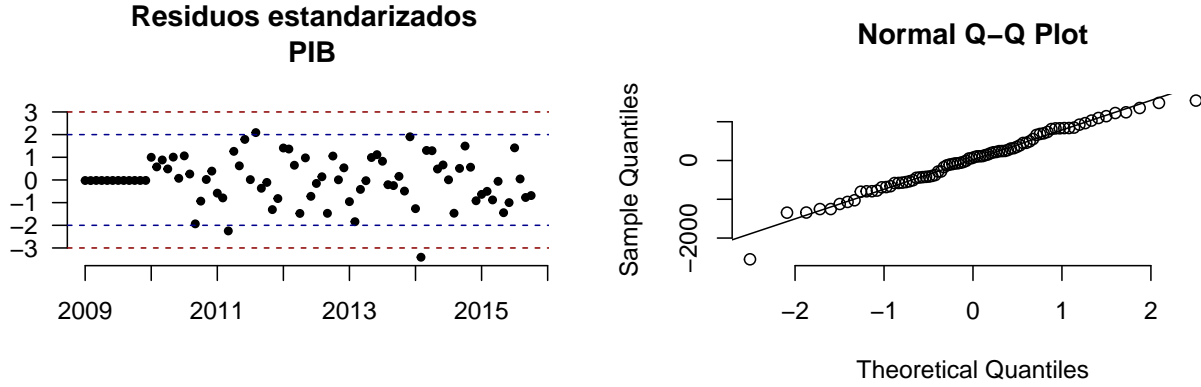


Figura 12: Residuos, SARIMAX

Residuos	Box - Ljung	Box-Pierce	Breusch-Pagan	lm ($H_0 : \beta_1 = 0$)
PIB: Valor-p	0,389	0,3976	0,1477	0,333

Tabla 6: Box test: Box-Ljung y Box-Pierce, Breusch-Pagan y lm ($H_0 : \beta_1 = 0$), al 95 % de confianza

Es claro, que tanto en el gráfico ACF, en el de valores-p del estadístico Box-Ljung y en los test para hipótesis de supuestos, se confirma la independencia serial con un 95 % de confianza, al igual que para el test de varianza y media constante. Los coeficientes estimados del modelos son lo que se muestran en la tabla 7.

Parámetros	ar2	ma1	ma2	ma8	sma1
Estimados	0.9818	0.6788	-0.5127	0.1940	-0.4445
Desviaciones	0.0238	0.1706	0.1852	0.0504	0.1303
Cuantil T-student	41.2647	3.9800	2.7681	8.8151	7.5337
IC ($\alpha = 0,5$)	(0,93; 1,03)	(0,35; 1,01)	(-0,87; -0,15)	(0,09; 0,29)	(-0,70; -0,18)

Tabla 7: Coeficientes proceso SARIMAX

Cabe destacar, que el coeficiente del regresor no es significativo al momento de analizar los coeficientes, por lo que el modelo final no contiene dicho parámetro, sin embargo se debe utilizar para posteriormente predecir datos en caso de desearse.

3.2. Selección de modelos

Para seleccionar el modelo más adecuado, se compara mediante AIC, BIC y error de predicción de los datos en las fechas {11-2015 a 10-2016}. En la tabla 8 se muestra un resumen de lo mencionado.

	df	AIC	df	BIC	Max.error %	Min.error %	Avg.error %
SARIMA	6.00	1170.54	6.00	1184.98	6.84	2.06	5.26
ARMAX PIB	6.00	1336.18	6.00	1350.62	8.77	0.02	4.01
ARMAX IMACEC	6.00	1340.62	6.00	1355.06	9.12	0.64	3.91
SARIMAX	6.00	1199.87	6.00	1214.31	3.03	1.02	2.16

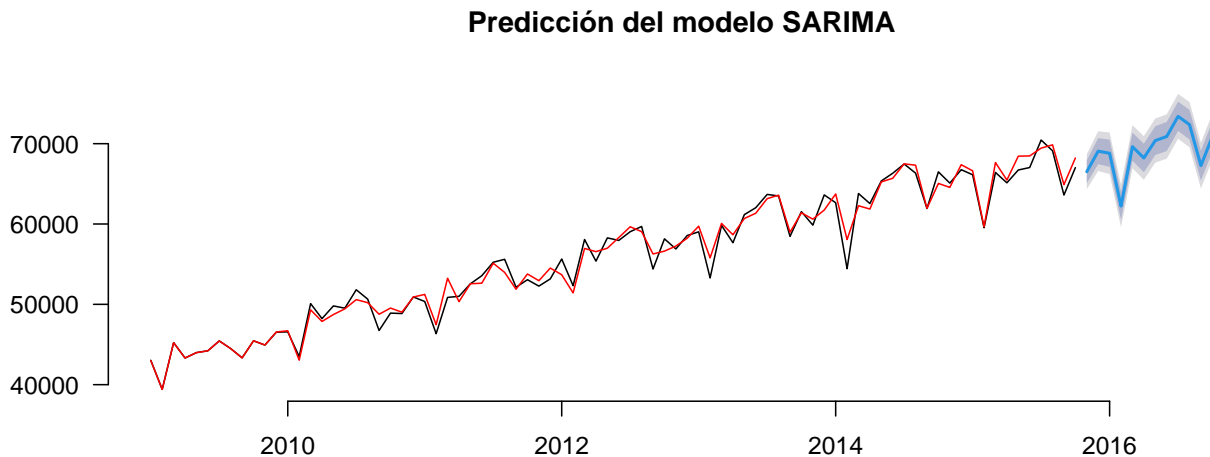
Tabla 8: Selección de modelos

Es claro que dentro de la metodología propuesta, el proceso SARIMA es el que tiene menor AIC y BIC dentro de lo señalado en un principio, además de tener el menor de los máximos errores porcentuales de predicción, aunque tiene el mayor error de predicción promedio. Es por esto, que es el modelo más adecuado para modelar los datos, principalmente a corto plazo.

Por otro lado, el proceso SARIMAX añadido como innovación a la metodología principal, presenta un desempeño mucho mejor al SARIMA teniendo un mejor AIC y menor margen de error en la predicción. Sin embargo, cabe recordar que dicho modelo descartó la presencia del PIB, por lo cual es un poco confuso en tratamiento que da el modelo a los datos finalmente.

4. Predicción

En la figura 13, se muestra la predicción del modelo SARIMA en los meses {11-2015 a 11-2017}.



5. Conclusiones

El modelamiento de los datos destaca el uso de diversas técnicas, en este caso en particular los modelos usados se caracterizan principalmente por el uso o no uso de regresores, es por esto que llama la atención el desempeño de cada uno, teniendo las siguiente fortalezas y debilidades.

1. Fortalezas:

- a)* Variada cantidad de procesos para modelar los datos.
- b)* Distintos desempeños y técnicas de ajuste.

2. Debilidades:

- a)* Bajo impacto de regresores, lo que sugiere un tratamiento distinto de las variables, como por ejemplo transformar regresores.
- b)* Los procesos no pueden clasificarse en su desempeño previamente a conocer los datos.
- c)* Iteración de parámetros para ajustar procesos a los datos.