Trabajo No 3 de Estadística III, 3009137
Modelos de Componentes versus Modelos ARIMA-SARIMA

1. Características del Trabajo

Este Trabajo es una continuación de los Trabajos No's 1 y 2. Consiste en ajustar modelos SARIMA a la serie asignada con la misma estrategia de validación cruzada considerando los $n = 216$ primeros datos con los cuales ajustó en los Trabajos 1 y 2 y comparar su ajuste y sus pronósticos con los que fueron obtenidos con los mejores modelos de cada trabajo, incluyendo los ajustes locales implementados en el trabajo 1.

2. Puntos a desarrollar

La presentación de la solución de los puntos a desarrollar y que se enuncian a continuación, se hará en modalidad de exposición oral entre el 6 al 8 de septiembre de 2021. Sin embargo, independiente de la fecha y hora específica en la que cada grupo hará su exposición, todos los grupos deberán subir en la plataforma moodle la presentación en archivo pdf, entre las 00:00 horas del 3 de septiembre hasta las 8:00 horas del 3 de septiembre. Antes de la semana de exposiciones se les informará una agenda de horarios en los que se puede convenir las exposiciones durante la semana establecida, incluyendo horarios de asesorías. El orden de temas a presentar en la exposición es igual al de los puntos que se piden desarrollar. ***Recuerde además, que el documento que suben debe estar completo tanto en resultados numéricos, como gráficos, en conclusiones en cada sección, formulación de pruebas, etc.*** Cada grupo contará con un máximo de 45 minutos para su presentación. ***Para el desarrollo de los siguientes puntos deberá trabajar sobre la serie recortada con la misma longitud usada en trabajos previos, es decir, con los primeros*** $n = 216$ ***datos. Esta vez no se desarrollará una introducción.***

1. Análisis descriptivo y test HEGY de raíces unitarias estacionales, usando los primeros $n = 216$ datos de la serie.

   a) Describa brevemente los patrones de la serie: tendencia, estacionalidad, presencia de ciclos y la varianza; explique por qué se puede o no considerar la tendencia y la estacionalidad como de tipo global.

   b) Si la serie es de varianza constante, presente y analice las siguientes gráficas

   - de la serie recortada y su ACF muestral,
   - de su primera diferencia regular ∇Y_t y su ACF muestral
   - de su primera diferencia estacional $\nabla_{12} Y_t$ y su ACF muestral
   - de la serie diferenciada por tendencia y estacionalidad (o sea $\nabla\nabla_{12} Y_t$) y su ACF muestral

   Para las ACFs use $m = 36$. El objetivo de estos análisis es mostrar que es necesario y suficiente los órdenes de diferencias regular y estacional dados en la Tabla 1, para llegar a un proceso estacionario. Para ello, en cada una de las series (la serie y sus diferencias arriba citadas) debe analizar y concluir en términos de estacionariedad o no estacionariedad, tanto de la estructura regular como de la estructura estacional, evaluando si:

   1) ¿Media es constante? Recuerde que la media de un proceso puede cambiar en el tiempo no sólo por tendencia sino también por patrones periódicos exactos o casi exactos, por lo tanto, el hecho de que el nivel de una serie es estable (es decir la tendencia es una recta de pendiente cero), no implica que la media es estable ya que puede aún tener cambio en el tiempo por patrón estacional periódico exacto o casi exacto.

2) ¿Varianza es constante?

3) ¿El proceso es ergódico? Aquí debe evaluar la ACF, separando las conclusiones sobre la parte regular (la cual se inspecciona en $k = 1, 2, 3, 4, 5, 6$) de la parte estacional (la cual se inspecciona en $k = 12, 24, 36$). Por tanto, puede pasar que la parte regular del proceso es ergódico pero no la estacional, o viceversa.

Nota 1: Con relación a la PACF, sólo presente y analice la correspondiente a la serie filtrada con el filtro $\nabla^d \nabla_{12}^D$, con d, D según indica la Tabla 1, únicamente en el punto 2a) de identificación y también use $m = 36$. Además, en su análisis se debe evaluar por separado los patrones en la parte regular (la cual se inspecciona en $k = 1, 2, 3, 4, 5, 6$) de la estacional (la cual se inspecciona en $k = 12, 24, 36$), indicando claramente si hay patrón corte o cola y qué tipo de cola en este último caso.

Nota 2: Si la serie es de varianza no constante, en lugar de trabajar en la escala original, considere a $\log(Y_t)$ para resolver este punto y los siguientes.

Nota 3: Ver Apéndice B para diferencias y ACFs pedidas en esta sección.

c) Test HEGY: Evalúe este test sobre la serie recortada, o sobre su logaritmo si es el caso (***ver en Apéndice A de esta guía cómo correr este test*** y la teoría sobre este test en el material publicado en moodle).

1) Plantee claramente el modelo AR (de orden $p = \infty$, ¿para cuál variable?) y el modelo de regresión (¿para cuál variable?) asociado a este test.

2) Reporte y analice los resultados conforme a las hipótesis que se prueban, criterios de rechazo y las conclusiones en cada prueba que se realiza dentro del test HEGY y responda a la cuestión ¿conviene diferenciar a la serie (o su logaritmo, si es el caso) tanto con filtro diferencia regular como con el filtro estacional de periodo 12?

Nota 4: Si la serie (o su logaritmo, si es el caso) tiene una posible raíz unitaria en su parte regular, la prueba para detectarla está incluida en el test HEGY (la primera prueba). En caso de detectar la raíz unitaria regular y cualquiera de las raíces unitarias estacionales, este test dice que sería necesario diferenciar a la serie en forma regular y estacional, ***pero la decisión final de aplicar esta diferencia mixta debe tener en cuenta también el análisis realizado en el literal anterior.***

2. Identificación de modelos SARIMA(p,d,q) × (P,D,Q)[12]: Sobre la serie (o para su logaritmo natural, si es el caso) recortada a $n = 216$ datos y filtrada con el filtro $\nabla^d \nabla_{12}^D$, con los órdenes d, D indicados en Tabla 1, usarán las funciones `acf - pacf` y `armasubsets` para identificar modelos arma estacionales estacionarios de media cero y luego, darán la ecuación del modelo ARIMA estacional para la serie (o para su logaritmo natural, si es el caso). Si $d = D = 1$, los modelos van sin deriva ya que en R sólo es posible incluirla cuando se usa únicamente uno de los dos tipos de diferencias, es decir, si $d = 0$, $D \neq 0$, o bien, si $d \neq 0$, $D = 0$. Por otro lado, usarán la función `auto.arima` directamente sin aplicar diferencias sobre la serie (o sobre su logaritmo natural, si es el caso) recortada a $n = 216$ datos; estos modelos pueden resultar con uno de los dos órdenes d ó D con valor cero y podrían incluir una deriva δ.

a) Examen de la ACF y la PACF de la serie recortada (o de su logaritmo, según sea el caso y use $m = 36$) pero diferenciada con el filtro $\nabla^d \nabla_{12}^D$, con d, D según indique Tabla 1. Chequee si el modelo que identifican es o no igual al modelo que se les asigna en la Tabla 1 bajo el nombre de modelo 1 (lea bien las observaciones o notas al final de la Tabla 1). Recuerde que para sustentar, se debe mostrar análisis de ACF y PACF parte regular (sólo inspeccione patrones en $k = 1, 2, 3, 4, 5, 6$) y de la ACF y PACF parte estacional (sólo inspeccione en $k = 12, 24, 36$), indicando claramente dónde hay patrón cola y el tipo de cola, y donde haya patrón de corte debe decir hasta cuál k es este patrón no nulo estadísticamente. Recuerde que si alguna de las dos estructuras, regular o estacional, es tipo AR o tipo MA, debe explicar claramente cuál es el orden y por qué.

b) La función `auto.arima()` de la librería `forecast`, aplicada a la serie recortada sin diferenciarla (o sobre su logaritmo, según sea el caso), usando todas las combinaciones posibles de los argumentos `ic` y `seasonal.test`:

```
auto.arima(yt,ic="aic",seasonal.test="ocsb")
auto.arima(yt,ic="aic",seasonal.test="ch")
auto.arima(yt,ic="aic",seasonal.test="seas")
auto.arima(yt,ic="bic",seasonal.test="ocsb")
auto.arima(yt,ic="bic",seasonal.test="ch")
auto.arima(yt,ic="bic",seasonal.test="seas")
```

donde `yt` corresponde a la serie recortada para el ajuste con validación cruzada. *En el documento de la presentación debe mostrar un print-screen de la consola R donde se vea la ejecución de estas seis líneas y el respectivo resultado.* Recuerde que si la serie es de varianza no constante debe realizar lo anterior sobre el logaritmo natural de la serie recortada, $\log(Y_t)$. Tenga en cuenta que uno de los modelos resultantes debe coincidir con el que aparece en la Tabla 1 bajo el nombre de modelo 2. Aunque sólo trabajarán los modelos que se les propone en la Tabla 1, en esta sección debe presentar las ecuaciones de todos los modelos resultantes con `auto.arima` y además deberá discutir si estos son modelos apropiados, pues algunos dan $D = 0$ ó $d = 0$, con o sin deriva.

Nota 5: Asegúrese que cuenta con la librería `uroot`, la cual es necesaria para que pueda usarse el argumento `ch` en la función `auto.arima`.

c) La función `armasubsets()` de la librería `TSA` usando criterio BIC y argumento `ar.method='ols'`, de la serie recortada (o de su logaritmo, según sea el caso) diferenciada con el filtro $\nabla^d \nabla_{12}^D$, con d, D según indica Tabla 1. Obtenga las figuras variando los argumentos `nar` y `nma`, según lo que se indica en la Tabla 1 para los modelos 3 y 4, respectivamente, y use el renglón que también se indica en esta tabla, para cada uno de los dos tableros pedidos. Además, si en la Tabla 1 se indica agregar algún parámetro, dé primero la ecuación antes de modificar y luego la ecuación ingresando los términos correspondientes. Recuerde también que en los modelos identificados con esta función algunos coeficientes deben ser fijados en cero. Verifique que las dos figuras que obtienen son las mismas que aparecen para cada serie en el Apéndice F de este documento (los modelos tienen que ser tomados de las figuras que se les está mostrando en esta guía y que deberían ser las mismas que ud. obtiene cuando use su propio programa R).

Repase en diapositivas de clase sobre procesos estocasticos estacionales, cómo se usa esta función para identificar Modelos ARMA(p,q)(P,Q)[12] y en el Apéndice G de esta guía, vea su uso para identificar un SARIMA y cómo se realiza su ajuste, a través de los ejemplos que se presentan allí. En el Apéndice H vea una isntrucción sobre cómo generar y editar las gráficas de estos tableros con un formato apropiado para incluirlas en documento de Word o Powerpoint. Vea además Apéndice D.

Nota 6: Para cada modelo escriba en esta sección sólo la ecuación teórica para Y_t (o para $\log(Y_t)$ cuando sea el caso) en la forma $\phi_p(B)\Phi_P(B^s)\nabla_s^D\nabla^d Y_t = \theta_q(B)\Theta_Q(B^s)E_t$ (modelo sin deriva) o bien $\phi_p(B)\Phi_P(B^s)\nabla_s^D\nabla^d Y_t = \delta + \theta_q(B)\Theta_Q(B^s)E_t$ (modelo con deriva) con sus respectivos supuestos, pero escribiendo por extensión cada polinomio y diferencia. Por ejemplo, si se identifica para una serie mensual Y_t un modelo ARIMA(2,1,3)(2,1,2)[12] con todos sus parámetros (OJO: no es posible incluir una deriva al usar las dos diferencias), escribir

$$\begin{aligned}(1-\phi_1 B-\phi_2 B^2)(1-\Phi_1 B^{12}-\Phi_2 B^{24})(1-B)(1-B^{12})Y_t &= \\ (1+\theta_1 B+\theta_2 B^2+\theta_3 B^3)(1+\Theta_1 B^{12}+\Theta_2 B^{24})E_t,\ \text{con } E_t \text{ RB } N(0,\sigma^2). \quad (1)\end{aligned}$$

3. Ajuste de modelos con validación cruzada: Sólo considere los cuatro modelos que se les indica en la Tabla 1. Ajuste los modelos asignados usando la función `Arima()` mediante método 'ML' (argumento method='ML'), excepto en los casos donde se le indica usar otro método de estimación (Ver Tabla 1) y construya la tabla de parámetros ajustados con la función R `coeftest()` de la librería `lmtest` (***Ver Apéndices C y D***). Como las ecuaciones finales a las que se llegaría despejando a Y_t en los casos aditivos, o a $\log(Y_t)$ en los casos multiplicativos, exigen la multiplicación de polinomios AR con las diferencias, y entre los polinomios MA, implicando muchas operaciones algebraicas cuyos resultados son

expresiones muy extensas, deje la ecuación teórica como se les indicó en el numeral anterior y no dé ecuación ajustada (o sea, no escriba ecuaciones para $\widehat{Y}_t$). Evalúe significancia para los parámetros y compare calidad de ajuste tanto por las gráficas de la serie ajustada en su escala original, como por las medidas AIC y BIC calculadas usando $\exp(C_n^*(p))$ (ver Apéndice E). Tenga en cuenta que para modelos ajustados sobre logaritmo de la serie también es necesario aplicar el factor de corrección por transformación lognormal al traer valores ajustados a la escala original, es decir, $\widehat{Y}_t \approx \exp\left(\widehat{\log(Y_t)}\right) \times \exp(\hat{\sigma}^2/2)$.

Recuerde: Como ecuación ajustada es una expresión algebraica muy extensa, no reportar estas ecuaciones.

4. Análisis de residuales y validación de supuestos: Para los cuatro modelos propuestos, presente los gráficos de residuales, ACF, PACF y test de Ljung-Box (use $m = 36$) construidos con sus residuos de ajuste. Examine los residuos en cada modelo y evalúe el supuesto de ruido blanco y test de normalidad sobre E_t. Recuerde que la normalidad se prueba sólo si no se detectan autocorrelaciones ni autocorrelaciones parciales que se consideren estadísticamente distintas de cero, y se hace mediante la gráfica de probabilidad normal y test Shapiro-Wilk. En todos los tests, formule claramente las hipótesis, estadísticos de prueba y criterios de decisión, como también las respectivas conclusiones.

5. Pronósticos para la validación cruzada: Para los cuatro modelos propuestos en la Tabla 1 presente los resultados y análisis de pronósticos: Tablas de pronósticos puntuales y por intervalos, sus medidas de precisión, interpretación de pronósticos y de las medidas de precisión, gráfico comparativo de los pronósticos. Tenga en cuenta que para modelos ajustados sobre logaritmo de la serie también es necesario aplicar el factor de corrección por transformación lognormal al traer valores pronosticados a la escala original, es decir, $\widehat{Y}_n(L) \approx \exp\left(\widehat{\log(Y_n(L))}\right) \times \exp(\hat{\sigma}^2/2)$.

   Sobre ecuación de pronóstico: Como son algebraicamente muy extensas, no reportar estas ecuaciones.

6. Conclusiones finales: Compare los modelos obtenidos a lo largo del semestre: Mejor modelo de regresión global del trabajo 1, mejor modelo de regresión global con errores ARMA en el trabajo 2, mejor modelo SARIMA del trabajo 3 y el mejor de los modelos locales (descomposición & loess y Holt-Winters). Para ello

   - dé las ecuaciones teóricas de cada uno de estos modelos (puede organizarlas en una tabla como se ha ilustrado en documentos de ejemplos de clase),
   - las gráficas de los ajustes y los valores de los criterios AIC y BIC presentados en una tabla, calculados por $\exp(C_n^*(p))$,
   - una gráfica comparativa de los pronósticos puntuales,
   - Tabla reportando las medidas de cobertura y amplitud de los pronósticos por intervalos y las medidas de precisión de pronósticos puntuales MAE, MAPE y RMSE, todo esto en la escala original de los datos,
   - comparación resumida de los resultados de la validación de supuestos: ¿los errores de ajuste provienen de un proceso de ruido blanco con distribución normal? No muestre gráficos de residuos ni ACFs, ni PACFs, sino un resumen en una tabla indicando las conclusiones sobre validez de supuesto "error de ajuste ruido blanco" y sobre supuesto "error de ajuste distribuye normal", en cada modelo.
   - Aspectos a favor y en contra de cada uno de los modelos probados
   - Conclusión final: Ésta es muy importante, es la conclusión de los tres trabajos y en la cual debe dar una recomendación respecto a cómo se debe modelar esta serie para construir pronósticos, teniendo en cuenta toda la información con relación a validez de supuestos, calidad de ajuste y de pronósticos y sobre todo cómo es el comportamiento de la serie en el tiempo vs. cómo cada modelo representa o no sus patrones ¿Cuál tipo de modelación se aproxima mejor a la dinámica de la serie en el tiempo?

Nota 7:

- Recuerde que para el caso con series transformadas por logaritmo natural, debe calcular ajustes y pronósticos en la escala original, es decir, $\hat{Y}_t$ y $\hat{Y}_n(L)$, respectivamente, y para ello se exponencian los valores hallados en escala logarítmica multiplicados por factor de corrección $\exp(\hat{\sigma}^2/2)$, esto es, $\widehat{Y}_t = \exp\left(\widehat{\log Y_t}\right) \times \exp\left(\hat{\sigma}^2/2\right)$ y $\widehat{Y}_n(L) = \exp\left(\widehat{\log Y_n(L)}\right) \times \exp\left(\hat{\sigma}^2/2\right)$, donde $\hat{\sigma}^2$ es la estimación de la varianza del ruido blanco del modelo. Sin embargo, la validación de supuestos se realiza usando los residuales del modelo en escala logarítmica.

- Todas las ACF, PACF deben realizarse hasta $k = 36$. Así mismo el test Ljung-Box debe fijarse con máximo $m = 36$. Sólo en la ACF y PACF usada en la identificación sobre $\nabla^d \nabla_{12}^D Y_t$, con d, D según indique la Tabla 1, en los casos aditivos ó sobre $\nabla^d \nabla_{12}^D \log(Y_t)$ en los casos multiplicativos, coloque líneas de referencia en múltiplos de $s = 12$, por ejemplo, como se ilustra a continuación, donde difdD12 representa a un objeto R que guardó una serie mensual diferencia apropiadamente por tendencia y/o estacionalidad.

```
win.graph(width=3.7,height=2.8)
acf(as.numeric(difdD12),ci.type="ma",lag.max=36,lwd=2,main="",cex.lab=0.5,cex.axis=0.5)
title(main="ACF difdD12",cex.main=0.5)
abline(v=c(12,24,36),lty=2,col=2)

win.graph(width=3.7,height=2.8)
pacf(as.numeric(difdD12),lag.max=36,lwd=2,main="",cex.lab=0.5,cex.axis=0.5)
title(main="PACF difdD12",cex.main=0.5)
abline(v=c(12,24,36),lty=2,col=2)
```

3. Los grupos para el Trabajo y Asignaciones

Se mantienen los grupos y asignación de los datos.

Tabla 1: Modelos SARIMA a ajustar

			modelo 1 (a)	**modelo 2 (b)**	**modelo 3 (a)**	**modelo 4 (a)**
serie	$\log(Y_t)$	**Difer.**	**ACF-PACF o sustituto**	**auto.arima**	**armasubsets 1 (c)**	**armasubsets 2 (c)**
Datos1	no	d=0, D=1	ARIMA(5,0,0)(2,1,0)[12] con deriva	ARIMA(2,0,2)(0,1,2)[12] con deriva	12x12, renglón 5, método 'ols', e incluir a Φ_1, Φ_2 y deriva	24x24, renglón 4, método 'ols' e incluir a ϕ_2 y deriva
Datos3	sí	d=D=1	ARIMA(2,1,1)(2,1,1)[12]*	ARIMA(1,1,3)(2,1,1)[12]	18x18, renglón 2, método 'ols' e incluir a Φ_2	24x24, renglón 1, método 'ols'
Datos4	no	d=D=1	ARIMA(0,1,2)(2,1,1)[12]*	ARIMA(2,0,3)(2,1,2)[12] con deriva	12x12, renglón 2, método 'ols', e incluir a Φ_2	18x18, renglón 2, método 'ols', e incluir Φ_2, Θ_2
Datos6	sí	d=D=1	ARIMA(6,1,0)(2,1,1)[12]*	ARIMA(1,0,2)(1,1,0)[12] con deriva	12x12, renglón 6, método 'ols', e incluir a Φ_2. Estimar por 'CSS-ML' en Arima()	18x18, renglón 2, método 'ols' e incluir a Φ_2
Datos9	sí	d=D=1	ARIMA(10,1,0)(0,1,1)[12]*	ARIMA(1,0,2)(2,1,2)[12] con deriva	12x12, renglón 2, método 'ols'	18x18, renglón 4, método 'ols'
Datos10	no	d=D=1	ARIMA(0,1,1)(1,1,1)[12]*	ARIMA(0,1,1)(0,1,2)[12]	12x12, renglón 3, método 'ols'	24x24, renglón 1, método 'ols' e incluir ϕ_1, θ_1
Datos11	sí	d=D=1	ARIMA(0,1,1)(0,1,1)[12]	ARIMA(1,0,1)(1,1,1)[12] con deriva	12x12, renglón 1, método 'ols'	24x24, renglón 1, método 'ols' e incluir a θ_1
Datos12	sí	d=D=1	ARIMA(0,1,1)(0,1,1)[12]	ARIMA(0,1,1)(1,1,1)[12]	12x12, renglón 3, método 'ols' e incluir a ϕ_{10}	24x24, renglón 2, método 'ols' e incluir a θ_1
Datos13+	no	d=D=1	ARIMA(0,1,6)(0,1,1)[12]	ARIMA(1,1,1)(0,0,2)[12] con deriva	12x12, renglón 1, método 'ols'. Estimar por 'CSS-ML' en función Arima()	18x18, renglón 2, método 'ols'. Estimar por 'CSS' en función Arima()
Datos14	no	d=D=1	ARIMA(2,1,7)(0,1,1)[12]*	ARIMA(2,1,0)(2,0,0)[12]	12x12, renglón 1, método 'ols'	12x12, renglón 3, método 'ols'
Datos15	no	d=D=1	ARIMA(5,1,0)(0,1,2)[12]	ARIMA(1,0,1)(0,1,2)[12] con deriva	12x12, renglón 2, método 'ols'	24x24, renglón 4, método 'ols'

Tabla 1 (Continuación)

serie	$\log(Y_t)$	Difer.	**modelo 1 (a)** **ACF-PACF o sustituto**	**modelo 2 (b)** **auto.arima**	**modelo 3 (a)** **armasubsets 1 (c)**	**modelo 4 (a)** **armasubsets 2 (c)**
Datos16	no	d=D=1	ARIMA(4,1,4)(1,1,1)[12] *	ARIMA(3,0,1)(2,1,2)[12] con deriva	18x18, renglón 3, método 'ols' e incluir ϕ_2	24x24, renglón 2, método 'ols' e incluir Θ_1
Datos17	no	d=D=1	ARIMA(7,1,4)(1,1,1)[12] *	ARIMA(2,0,1)(0,1,1)[12] con deriva	18x18, renglón 1, método 'ols' e incluir ϕ_6	24x24, renglón 1, método 'ols' e incluir Θ_3
Datos18	sí	d=D=1	ARIMA(2,1,0)(0,1,3)[12]	ARIMA(1,1,2)(2,1,2)[12]	18x18, renglón 1, método 'ols' e incluir Φ_3	18x18, renglón 4, método 'ols'
Datos19	no	d=D=1	ARIMA(2,1,2)(1,1,3)[12]*	ARIMA(2,0,1)(0,1,2)[12] con deriva	12x12, renglón 1, método 'ols' e incluir Φ_3	18x18, renglón 1, método 'ols' e incluir a Φ_3
Datos20	no	d=D=1	ARIMA(8,1,2)(0,1,1)[12] *	ARIMA(1,0,2)(0,1,1)[12] con deriva	12x12, renglón 1, método 'ols'	18x18, renglón 2, método 'ols'
Datos21	no	d=D=1	ARIMA(5,1,1)(2,1,3)[12]*	ARIMA(1,0,2)(1,1,0)[12] con deriva	12x12, renglón 2, método 'ols'	18x18, renglón 1, método 'ols'
Datos22+	no	d=D=1	ARIMA(6,1,2)(0,1,2)[12]*	ARIMA(1,0,2)(2,1,0)[12] con deriva	12x12, renglón 1, método 'ols' e incluir, $\theta_3, \Phi_1, \Theta_2$	12x12, renglón 2, método 'ols' e incluir, $\theta_3, \Phi_1, \Theta_2$. Ajustar por 'CSS' en función Arima
Datos23	no	d=D=1	ARIMA(8,1,2)(0,1,2)[12]*	ARIMA(1,0,2)(2,1,0)[12] con deriva	12x12, renglón 2, método 'ols' e incluir a ϕ_6, Φ_2	24x24, renglón 1, método 'ols', incluir a $\phi_1, \phi_2, \phi_5, \theta_{11}, \Theta_1$. Estimar por 'CSS-ML' en Arima()
Datos25	no	d=D=1	ARIMA(7,1,0)(0,1,1)[12]*	ARIMA(4,1,1)(2,0,0)[12]. Ajustar por 'CSS-ML' en Arima()	12x12, renglón 4, método 'ols'	18x18, renglón 1, método 'ols'
Datos28	sí	d=D=1	ARIMA(4,1,0)(1,1,1)[12] *	ARIMA(2,0,2)(1,1,0)[12] con deriva	24x24, renglón 1, método 'ols' e incluir a Θ_2	24x24, renglón 3, método 'ols' e incluir a Θ_2
Datos29	no	d=D=1	ARIMA(4,1,0)(1,1,1)[12]*	ARIMA(2,0,2)(1,1,0)[12] con deriva	12x12, renglón 5, método 'ols' e incluir a Φ_1	18x18, renglón 3, método 'ols'
Datos31	no	d=D=1	ARIMA(7,1,2)(1,1,2)[12]*	ARIMA(2,0,2)(1,1,0)[12] con deriva	12x12, renglón 4, método 'ols'	18x18, renglón 1, método 'ols' e incluir a ϕ_3
Datos32	sí	d=D=1	ARIMA(6,1,0)(3,1,0)[12]*	ARIMA(4,1,0)(2,1,2)[12]	18x18, renglón 2, método 'ols'	24x24, renglón 1, método 'ols'

(a) Para la identificación con estos métodos se trabaja sobre la serie (su logaritmo natural en el caso multiplicativo), recortada a $n = 216$ observaciones y filtrada con el filtro $\nabla^d \nabla_{12}^D$, con d y D según se indica en esta tabla

(b) Verifique que este modelo resulta con auto.arima por alguno de los criterios 'aic', ó 'bic' combinado con alguno de los tests de raíces unitarias estacionales: 'ch', 'ocsb' ó 'seas', sobre la serie (su logaritmo en el caso multiplicativo) con sólo los primeros $n = 216$ datos.

(c) Método 'ols' significa que debe usar el argumento ar.method='ols' en la función armasubsets. Verifique que los tableros resultantes con esta función coinciden con los del Apéndice F

+ En el modelo 4 sobre esta serie, al evaluar normalidad tanto por gráfico de probabilidad como por Shapiro Wilk, debe hacerlo retirando los primeros 35 residuos, es decir, sobre residuals(modelo4)[-c(1:35)]

* No resultan directamente de ACF-PACF, son propuestos basados en varios ensayos

Nota: Excepto en los casos donde se indica otro método, los modelos se estiman por método `'ML'` en la función `Arima`

Referencias

[1] Bowerman, B. L, O'Connell, R. T y Koehler, A. B. (2009) *Pronósticos, Series de Tiempo y Regresión. Un Enfoque Aplicado. 4 ed.* CENGAGE Learning

[2] Chatfield, C. (2019) *The Analysis of Time Series. An Introduction with R, Seventh edition.* CRC Press-USA.

[3] Diebold, F. (2001) *Elementos de Pronósticos.* International Thomson Editores, México.

[4] Cryer, J. D. and Chan, K-S. (2008) *Time Series Analysis With Applications in R.* Springer.

[5] González, N. G. (2013) *Notas de Clase Estadística III 3009137.* Escuela de Estadística, Universidad Nacional de Colombia Sede Medellín.

[6] Shumway, R. H. and Stoffer, D. S. (2017) *Time Series Analysis and Its Applications. With R Examples.* Fourth ed. Springer

APÉNDICE

A. Ejecución del test HEGY

Actualmente en la la librería `pdR` está disponible la función R `HEGY.test`, úsela con argumento Pmax=12 así:

1. Para aplicar el test a la serie recortada Y_t, si ésta es de componentes aditivas (la lectura de datos es como corresponda a la serie asignada, adapte según su caso):

```
#Lectura de los datos
#Este es un ejemplo, adapte lo que sea necesario según sus datos
datos=read.table(file.choose(),header=T,skip=7,sep=';',dec=",",colClasses=c(rep("NULL",10),"numeric",rep("NULL",2)))

datos=ts(datos,freq=12,start=c(2001,1))
#Defina longitud serie recortada
n=length(datos)-12 #En este ejemplo se recortan 12 datos
t=1:n
#Serie recortada
yt=ts(datos[t],freq=12,start=c(2001,1))
#Test Hegy sobre serie recortada
library(pdR)
HEGY.test(wts=yt,itsd=c(0,0,c(0)),selectlags=list(mode="aic", Pmax=12))$stats
```

2. Para aplicar test sobre $\log(Y_t)$, si la serie es de componentes multiplicativas, entonces seguir el siguiente ejemplo de programación (la lectura de datos es como corresponda a la serie asignada, adapte según su caso):

```
#Lectura de los datos
#Este es un ejemplo, adapte lo que sea necesario según sus datos
datos=read.table(file.choose(),header=T,skip=7,sep=';',dec=",",colClasses=c(rep("NULL",10),"numeric",rep("NULL",2)))

datos=ts(datos,freq=12,start=c(2001,1))
#defina longitud serie recortada
n=length(datos)-12 #en este ejemplo se recortan 12 datos
t=1:n
#Serie recortada
yt=ts(datos[t],freq=12,start=c(2001,1))
#Test Hegy sobre log de serie recortada
library(pdR)
HEGY.test(wts=log(yt),itsd=c(0,0,c(0)),selectlags=list(mode="aic", Pmax=12))$stats
```

B. Cómo aplicar en R diferencias regulares y estacionales a una serie de tiempo

Tenga en cuenta que en R la función `diff()` es la que permite generar series diferenciadas tanto regular como estacionalmente, así (se supone que `Yt` es un objeto serie de tiempo, para su caso, la serie recortada para el ajuste):

- $\nabla^d Y_t$: `diff(Yt,difference=d)`
- ∇Y_t: `diff(Yt)`
- $\nabla_{12}^D Yt$: `diff(Yt,lag=12,difference=D)`
- $\nabla_{12} Y_t$: `diff(Yt,lag=12)`
- $\nabla^d \nabla_{12}^D Y_t = \nabla_{12}^D \nabla^d Y_t$: `diff(diff(Yt,lag=12,difference=D),difference=d)`

Por ejemplo si $d = D = 1$ y $s = 12$, es decir, queremos obtener a $\nabla\nabla_{12} Y_t$, basta lo siguiente
`difdD12=diff(diff(Yt,lag=12))`
o bien,
`difdD12=diff(diff(Yt),lag=12).`
En el punto 1 de análisis descriptivo construya las ACFs de la serie y sus diferencias, y en el punto 2a) de identificación con

ACF-PACF de la serie debidamente diferenciada con filtro $\nabla^d\nabla_{12}^D$, con d, D según Tabla 1, construya estas dos gráficas, y en todos estos casos tenga la precaución de usar la función `as.numeric()` sobre estos objetos, por ejemplo (con $m = 36$),

```
win.graph(width=3.7,height=2.8)
acf(as.numeric(Yt),ci.type="ma",lag.max=36,lwd=2,main="",cex.lab=0.5,cex.axis=0.5)
title(main="ACF Yt",cex.main=0.5)

win.graph(width=3.7,height=2.8)
acf(as.numeric(difd1),ci.type="ma",lag.max=36,lwd=2,main="",cex.lab=0.5,cex.axis=0.5)
title(main="ACF diferencia regular",cex.main=0.5)
abline(v=c(12,24,36),lty=2,col=2)

win.graph(width=3.7,height=2.8)
pacf(as.numeric(difd1),lag.max=36,lwd=2,main="",cex.lab=0.5,cex.axis=0.5)
title(main="PACF diferencia regular",cex.main=0.5)
abline(v=c(12,24,36),lty=2,col=2)

win.graph(width=3.7,height=2.8)
acf(as.numeric(difdD12),ci.type="ma",lag.max=36,lwd=2,main="",cex.lab=0.5,cex.axis=0.5)}\\
title(main="ACF diferencia regular y estacional (d=D=1)",cex.main=0.5)
abline(v=c(12,24,36),lty=2,col=2)

win.graph(width=3.7,height=2.8)
pacf(as.numeric(difdD12),lag.max=36,lwd=2,main="",cex.lab=0.5,cex.axis=0.5)
title(main="ACF diferencia regular y estacional (d=D=1)",cex.main=0.5)
abline(v=c(12,24,36),lty=2,col=2)
```

Nota 8: De nuevo, recuerde que para las series con componentes multiplicativas debe trabajar sobre $\log(Y_t)$, en lugar de Y_t.

C. Cómo ajustar y pronosticar un SARIMA(p,d,q) $\times$ (P,D,Q)$_s$ sin deriva y con todos los coeficientes

Para el ajuste se usa la función `Arima()` especificando con el argumento `order=c(p,d,q)` el orden p, d, q para la parte regular y con el argumento `seasonal=list(order=c(P,D,Q))` el orden P, D, Q para la parte estacional. Por ejemplo, para ajustar sin deriva y pronosticar un SARIMA(2,1,1) $\times$ (1,1,2)[12], y asumiendo que Y_t ya tiene formato de serie de tiempo con frecuencia $s = 12$ **(ojo, los valores P son bajo la distribución $N(0, 1)$, es decir $P(|Z| > |Z_0|)$, con $Z \sim N(0, 1)$),**

```
modelo=Arima(yt,order=c(2,1,1),seasonal=list(order=c(1,1,2)),method="ML")
coeftest(modelo) #Tabla de parámetros estimados; valores P bajo la N(0,1)

#fecha de inicio de los pronósticos es enero 2019 y horizonte de 12 periodos
pronóstico=ts(as.data.frame(forecast(modelo,h=12,level=95)),freq=12,start=c(2019,1))
pronóstico
```

D. Cómo ajustar y pronosticar un SARIMA(p,d,q) $\times$ (P,D,Q)[s] sin deriva y con algunos de los coeficientes fijos en cero

Además de los descrito en C, se debe usar el argumento `fixed=`, donde con un vector de longitud $p + q + P + Q$, se indica con NA ó 0 cuáles coeficientes deben estimarse y cuáles fijarse en cero, respectivamente, en el siguiente orden: los primeros p valores para los ϕ_j, $j = 1, 2, \ldots, p$; los siguientes q valores para los θ_i, $i = 1, 2, \ldots, q$; los siguientes P valores para los Φ_k, $k = 1, 2, \ldots, P$ y los últimos Q valores para los Θ_l, $l = 1, 2, \ldots, Q$.

Por ejemplo, para ajustar sin deriva y pronosticar un SARIMA(9,1,3) $\times$ (1,1,2)[12], con $\phi_j \neq 0$, para $j = 2, 3, 9$, $\theta_i \neq 0$, para $i = 1, 3$, $\Phi_1 \neq 0$ (obviamente) y $\Theta_l \neq 0$ para $l = 2$, y asumiendo que Y_t ya tiene formato de serie de tiempo con frecuencia $s = 12$,

```
modelo=Arima(yt,order=c(9,1,3),seasonal=list(order=c(1,1,2)),
             fixed=c(0,NA,NA,rep(0,5),NA,NA,0,NA,NA,0,NA),method="ML")
coeftest(modelo) #Tabla de parámetros estimados; valores P bajo N(0,1)

#Fecha de inicio de los pronósticos es enero de 2019 y horizonte de 12 periodos
pronóstico=ts(as.data.frame(forecast(modelo,h=12,level=95)),freq=12,start=c(2019,1))
pronóstico
```

Nota 9:

1. Puede usar dentro del argumento `fixed=` la función `rep(a,b)` con la que se indica que el valor a se debe repetir b veces. Por ejemplo `rep(0,5)` genera el vector de longitud 5 y todos sus valores iguales a 0. También si fuese necesario puede usarse con NA. Por ejemplo, `rep(NA,5)` genera un vector de longitud 5 con todos sus valores iguales a NA.

2. Para incluir una deriva, R lo permite únicamente en presencia de sólo uno de los dos tipos de diferencias, y basta usar en la función `Arima()` el argumento `include.drift=TRUE`. Si la deriva se va a incluir en Modelos 3 y 4, es necesario agregar en el argumento `fixed` otro `NA` al final del vector que se construye para indicar parámetros que deben ir en el modelo.

3. Recuerde que si es necesaria la transformación logaritmo natural sobre la serie, entonces los modelos Arima se identifican y se ajustan primero en esa escala, así como los pronósticos, pero luego debe traer a la escala original estos resultados exponenciando y multiplicando por el factor de corrección. Vea cómo se hace esto en los ejemplos presentados en el Apéndice G (donde la serie que se muestra para ilustrar es estacional multiplicativa).

E. Cálculo de $\exp(C_n^*(p))$ en modelos ARIMA(p,d,q)(P,D,Q)[s]:

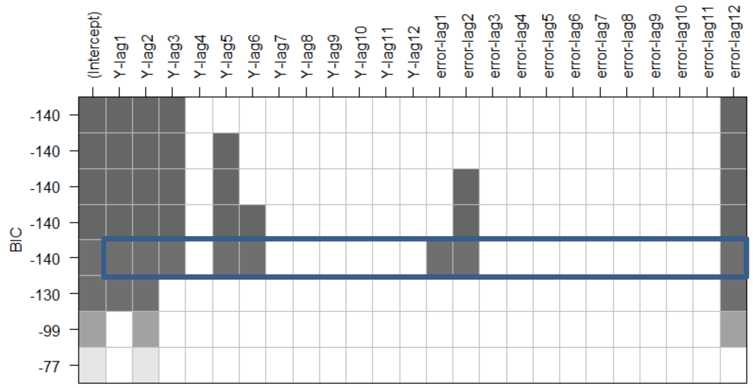
```
#Creando función usuario crit.inf.resid() para calcular C*n(p)
crit.inf.resid=function(residuales,n.par,AIC="TRUE"){
if(AIC=="TRUE"){
#Calcula AIC
CI=log(mean(residuales^2))+2*n.par/length(residuales)
}
if(AIC=="FALSE"){
#Calcula BIC
CI=log(mean(residuales^2))+n.par*log(length(residuales))/length(residuales)
}
CI
}
#Tome k igual al total de parámetros del modelo, contando incluso la deriva cuando el modelo tenga este parámetro
#Si no transformó a Yt:
AICmodelo=exp(crit.inf.resid(residuales=residuals(modelo),n.par=k))
BICmodelo=exp(crit.inf.resid(residuales=residuals(modelo),n.par=k,AIC="FALSE"))

#Si transformó a Yt con logaritmo natural
Yhat=exp(modelo$fitted)*exp(modelo$sigma2/2) #valores ajustados en escala original
res.orig=Yt-Yhat #Pseudo residuos
AICmodelo=exp(crit.inf.resid(residuales=res.orig,n.par=k))
BICmodelo=exp(crit.inf.resid(residuales=res.orig,n.par=k,AIC="FALSE"))
```

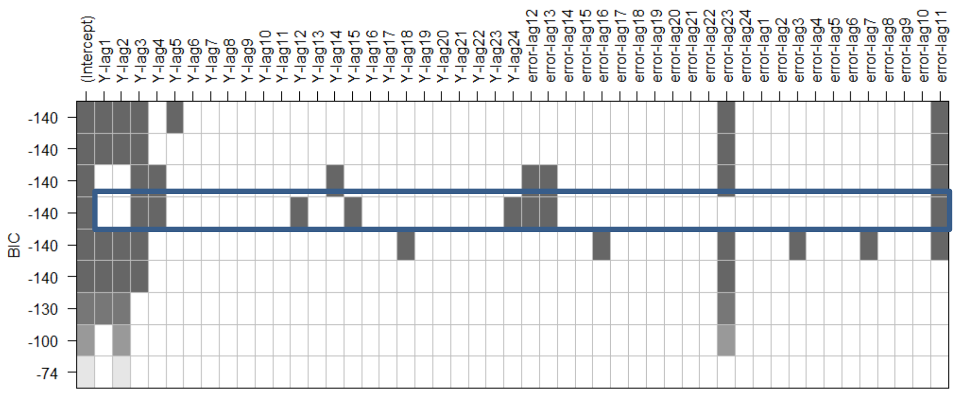
F. Gráficas de armasubsets aplicadas a $\nabla^{d}\nabla_{12}^{D}Y_t$ (caso aditivo) o sobre $\nabla^{d}\nabla_{12}^{D}\log(Y_t)$ (caso multiplicativo), usando sólo los primeros $n = 216$ datos (Recuerde que los valores de d, D son de acuerdo a lo que indica la Tabla 1 de asignación de modelos)

En Datos 1:

Modelo 3: Tablero 12x12, método 'ols', renglón 5 e incluir a Φ_1, Φ_2 y deriva δ

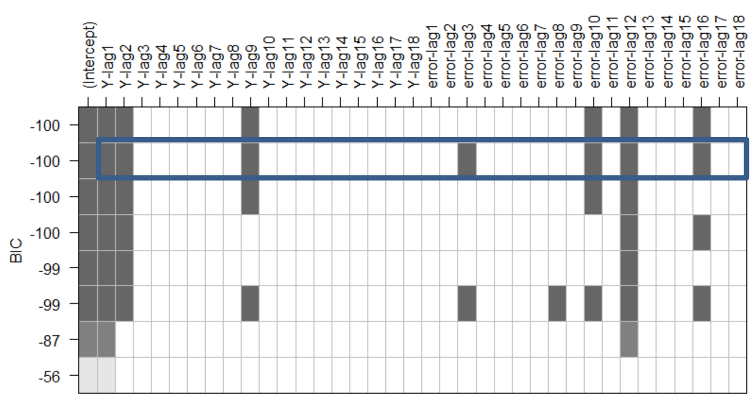


Modelo 4: Tablero 24x24, renglón 4, método 'ols', renglón 2, e incluir a ϕ_2 y deriva δ

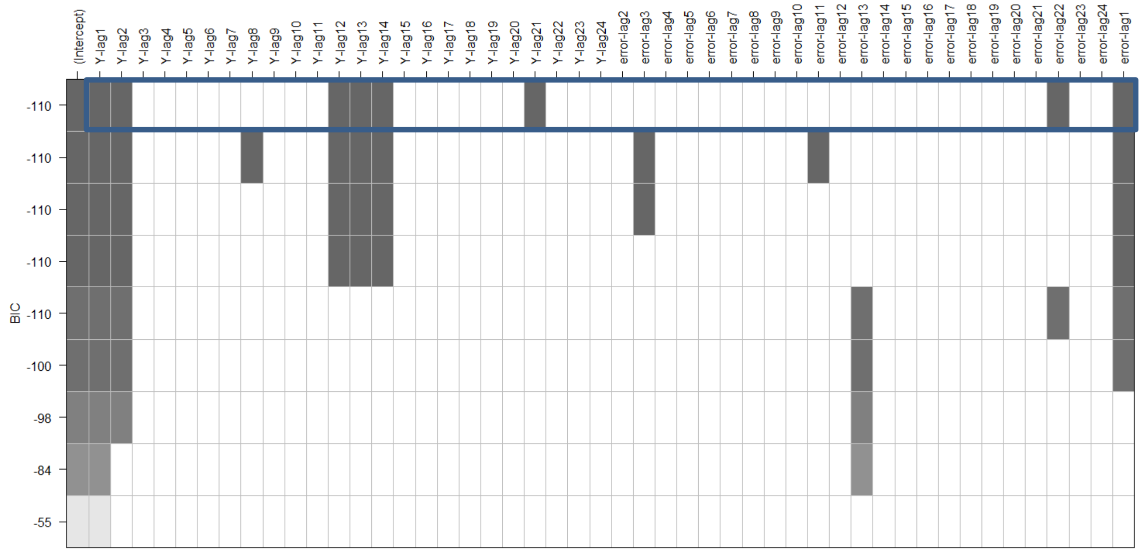


En Datos 3:

Modelo 3: Tablero 18x18, método 'ols', renglón 2, e incluir a Φ_2

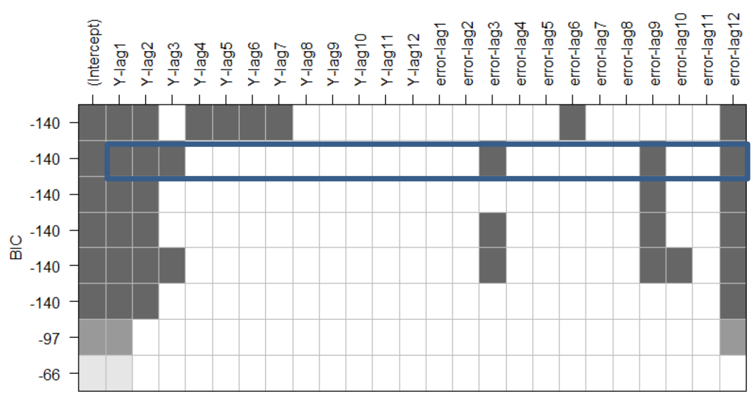


Modelo 4: Tablero 24x24, método 'ols', renglón 1

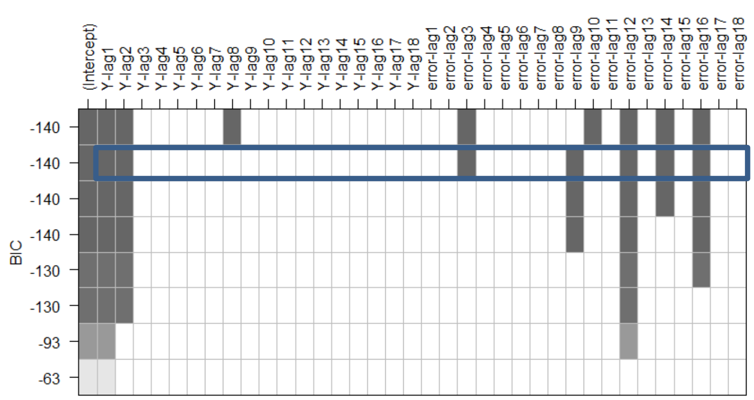


En Datos 4:

Modelo 3: Tablero 12x12, método 'ols', renglón 2, e incluir a Φ_2

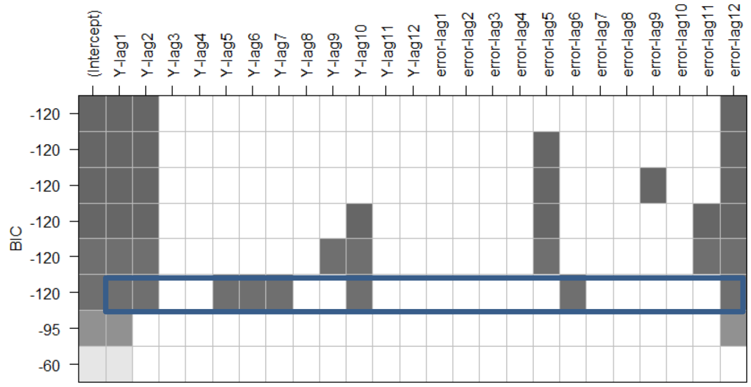


Modelo 4: Tablero 18x18, método 'ols', renglón 2, e incluir a Φ_2, Θ_2

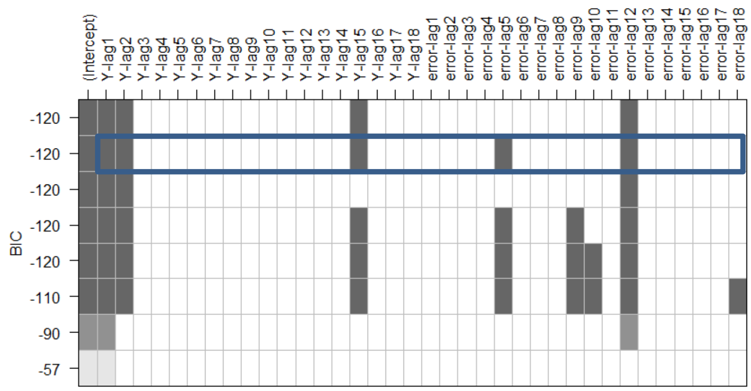


En Datos 6:

Modelo 3: Tablero 12x12, método ‘ols’, renglón 6 e incluir a Φ_2

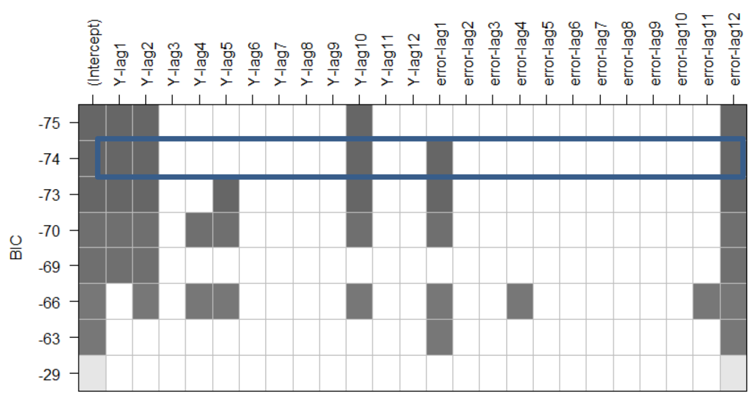


Modelo 4: Tablero 18x18, método ‘ols’, renglón 2 e incluir a Φ_2

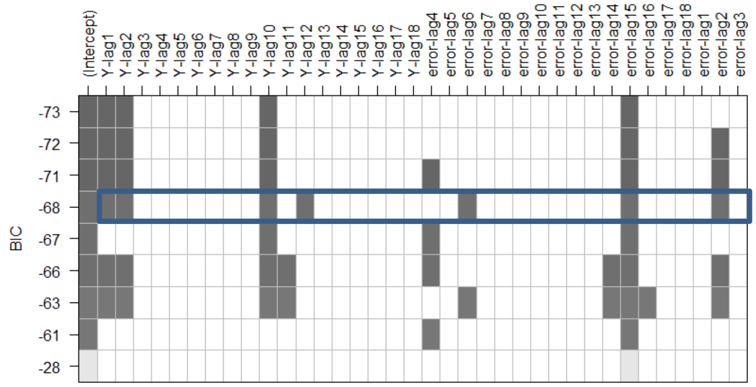


En Datos 9:

Modelo 3: Tablero 12x12, método ‘ols’, renglón 2

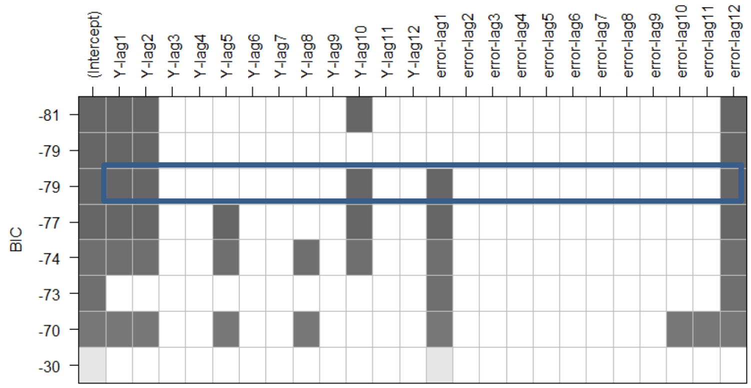


Modelo 4: Tablero 18x18, método 'ols', renglón 4,

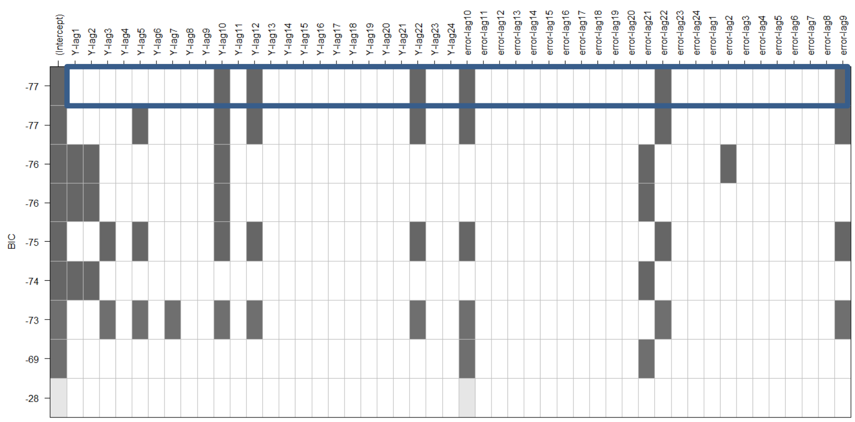


En Datos 10:

Modelo 3: Tablero 12x12, método 'ols', renglón 3

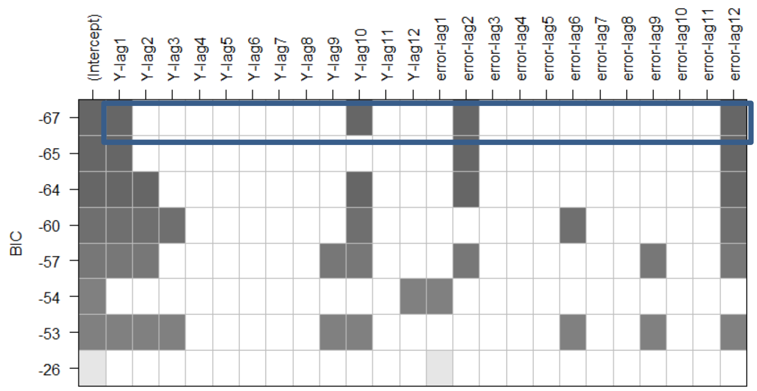


Modelo 4: Tablero 24x24, método 'ols', renglón 1 e incluir ϕ_1, θ_1

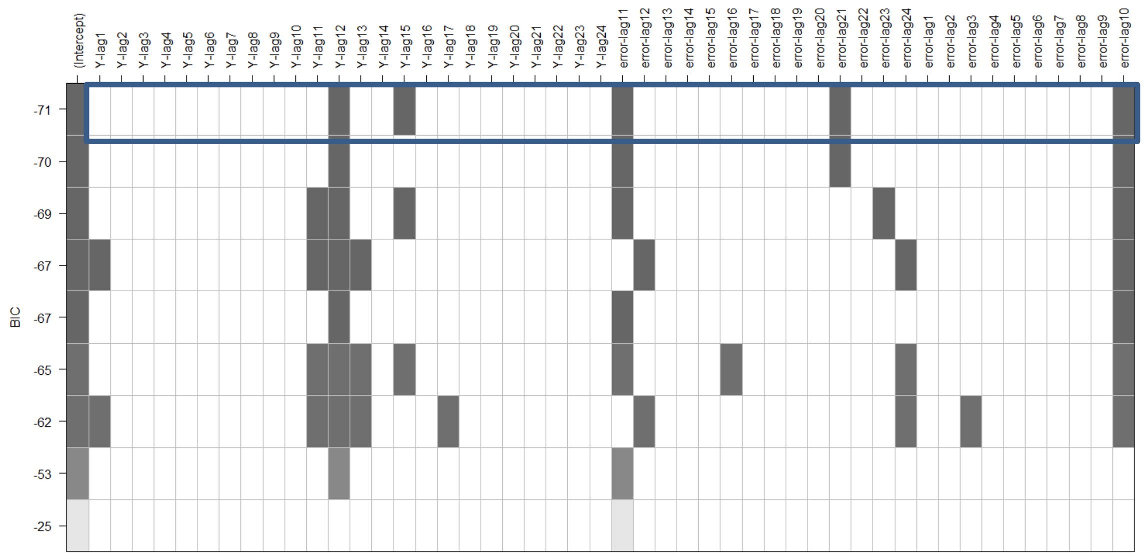


En Datos 11:

Modelo 3: Tablero 12x12, método ‘ols’, renglón 1

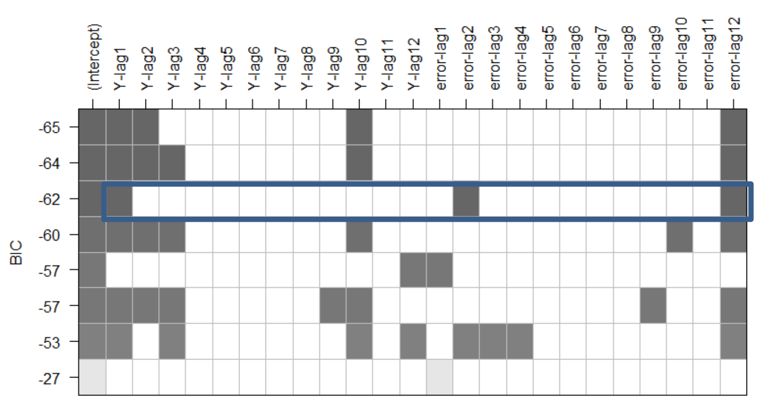


Modelo 4: Tablero 24x24, método ‘ols’, renglón 1 e incluir a θ_1

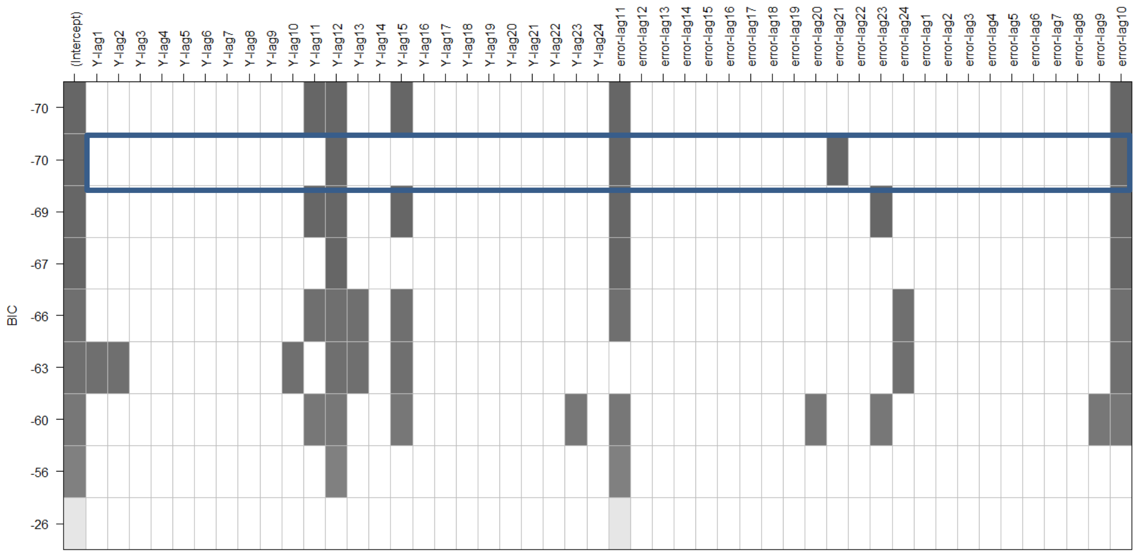


En Datos 12:

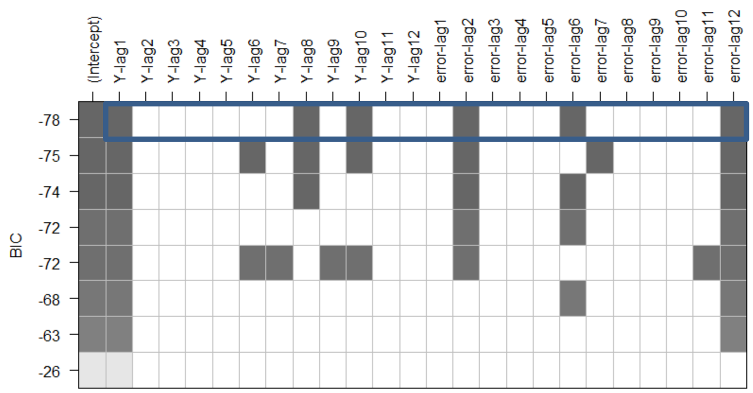
Modelo 3: Tablero 12x12, método ‘ols’, renglón 3 incluir a ϕ_{10}



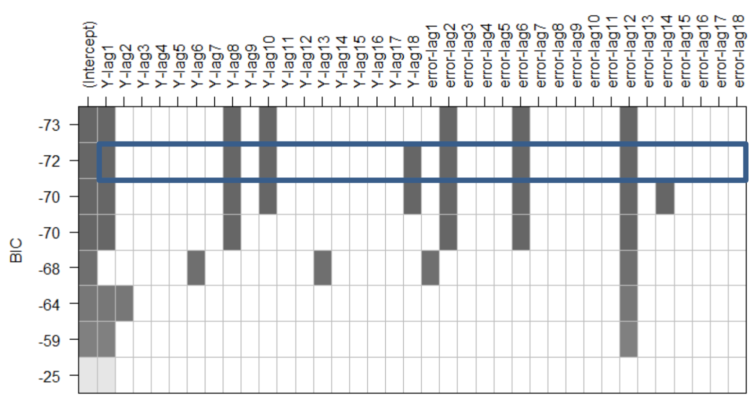
Modelo 4: Tablero 24x24, método ‘ols’, renglón 2 e incluir a θ_1



En Datos 13:
Modelo 3: Tablero 12x12, renglón 1, método ‘ols’

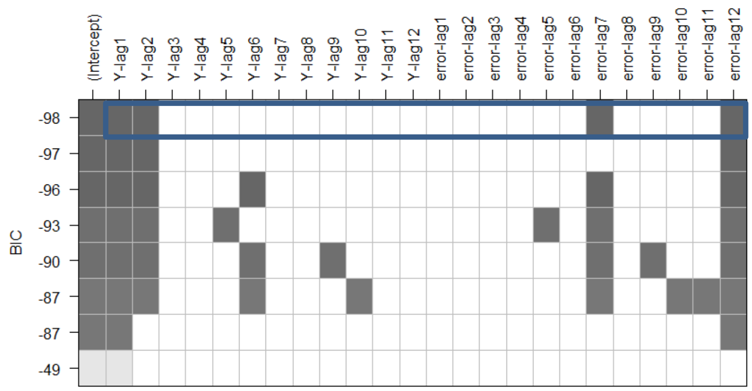


Modelo 4: Tablero18x18, renglón 2, método ‘ols’

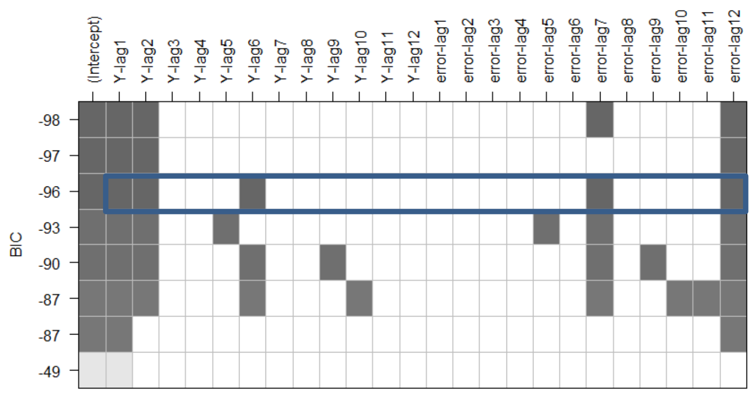


En Datos 14:

Modelo 3: Tablero 12x12, renglón 1, método ‘ols’

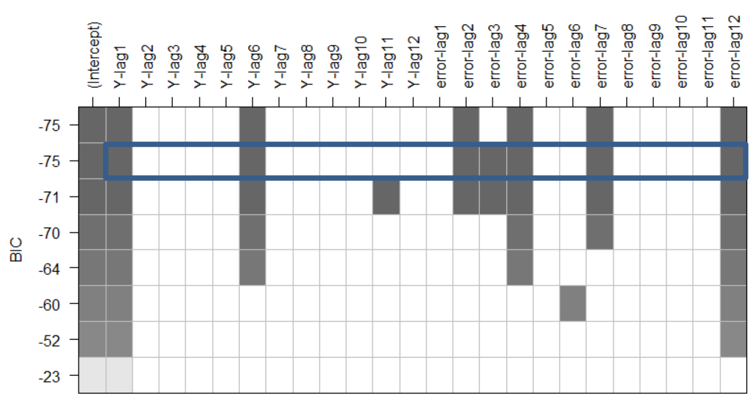


En modelo 4: Tablero 12x12, renglón 3, método ‘ols’

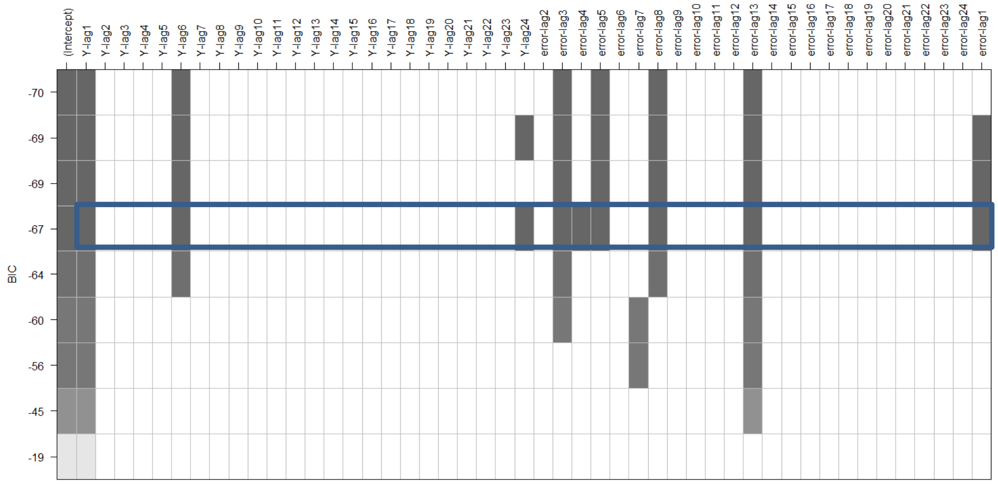


En Datos 15:

Modelo 3: Tablero 12x12, renglón 2, método ‘ols’

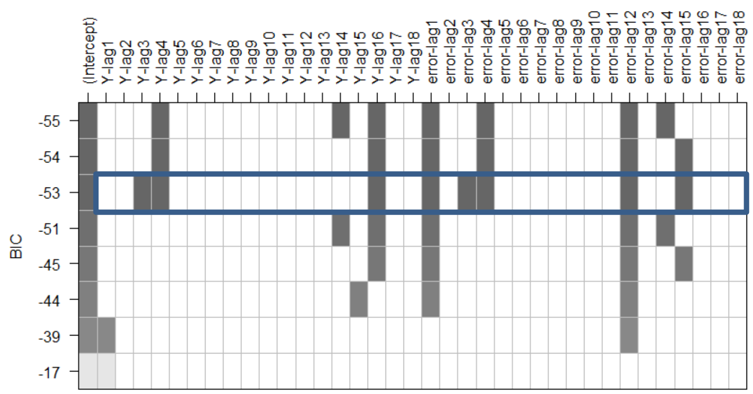


Modelo 4: tablero 24x24, renglón 4, método ‘ols’

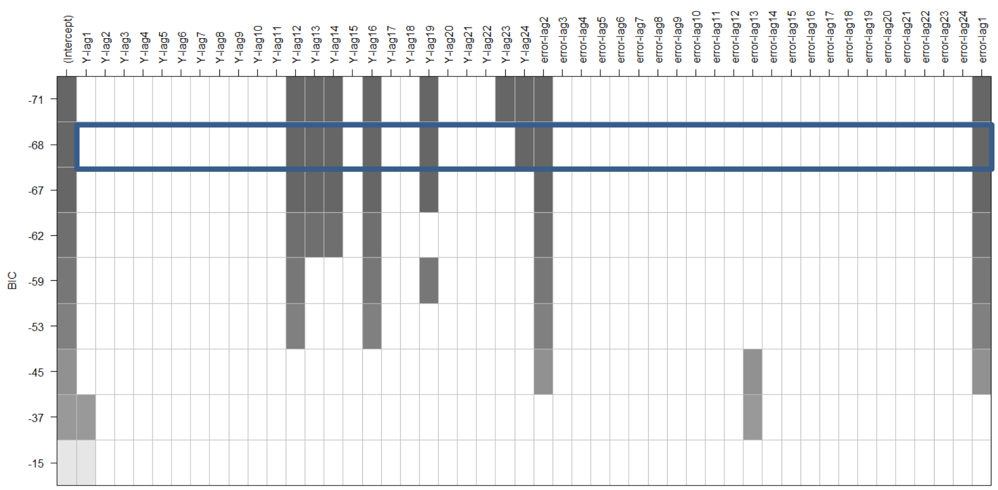


En Datos 16:

Modelo 3: tablero 18x18, renglón 3, método ‘ols’ e incluir a ϕ_2

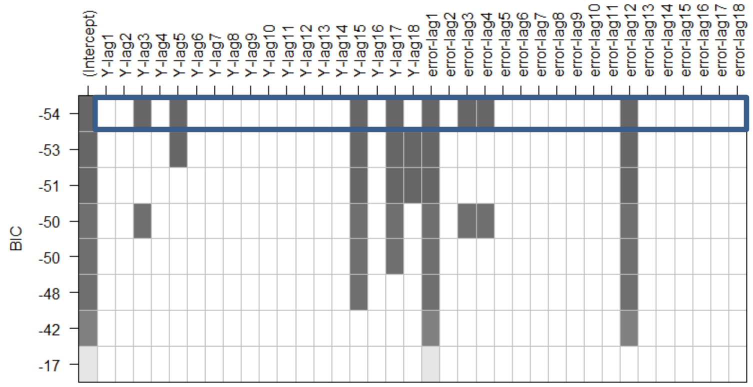


Modelo 4: Tablero 24x24, método ‘ols’, renglón 2, e incluir Θ_1

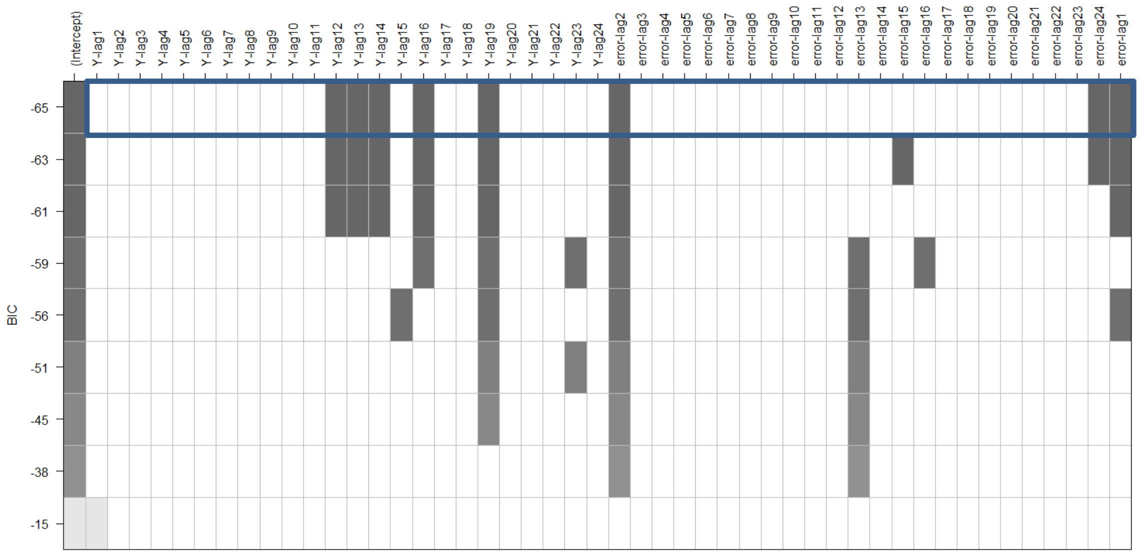


En Datos 17:

Modelo 3. Tablero 18x18, método ‘ols’, renglón 1 e incluir a ϕ_6

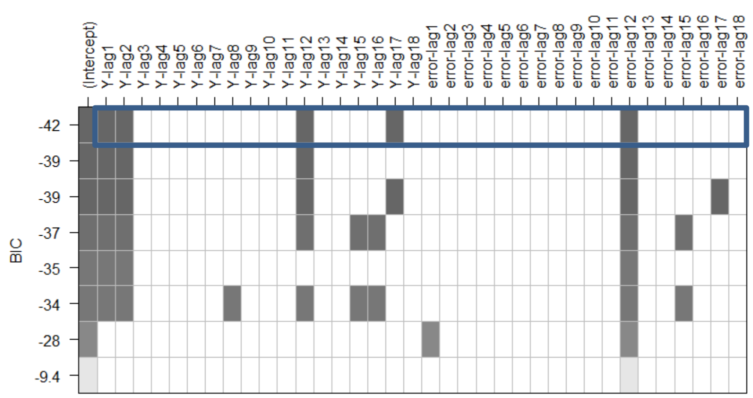


Modelo 4: Tablero 24x24, método ‘ols’, renglón 1 e incluir a Θ_3

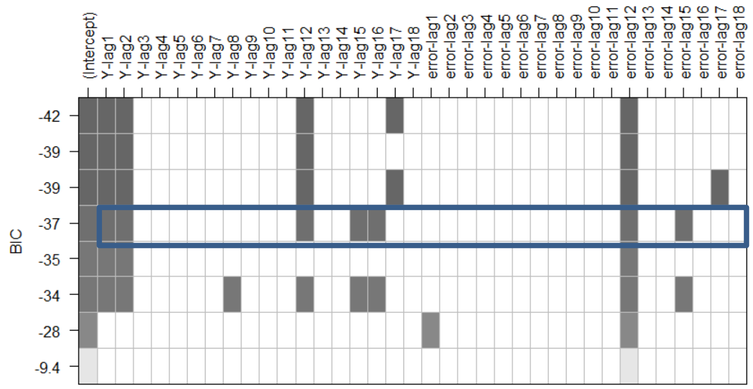


En Datos 18:

Modelo 3: Tablero 18x18, método ‘ols’, renglón 1 e incluir a Φ_3

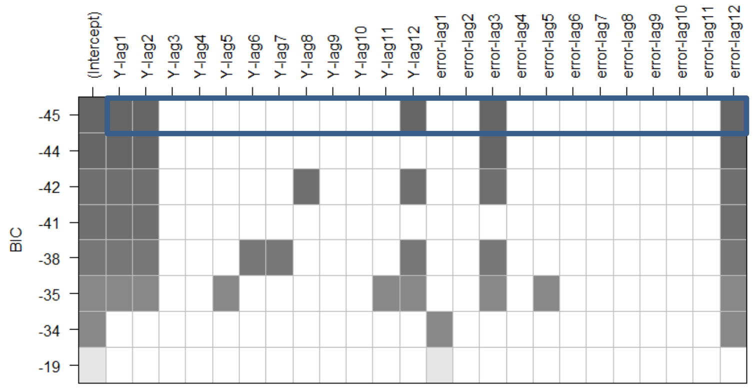


Modelo 4: Tablero 18x18, método ‘ols’, renglón 4

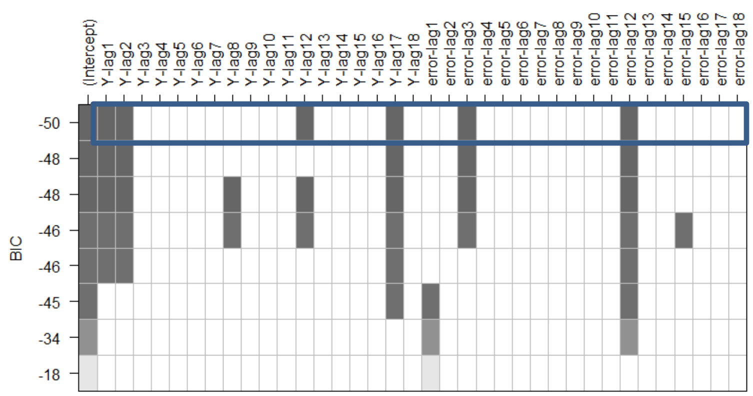


En Datos 19:

Modelo 3. Tablero 12x12, método ‘ols’, renglón 1 e incluir a Φ_3

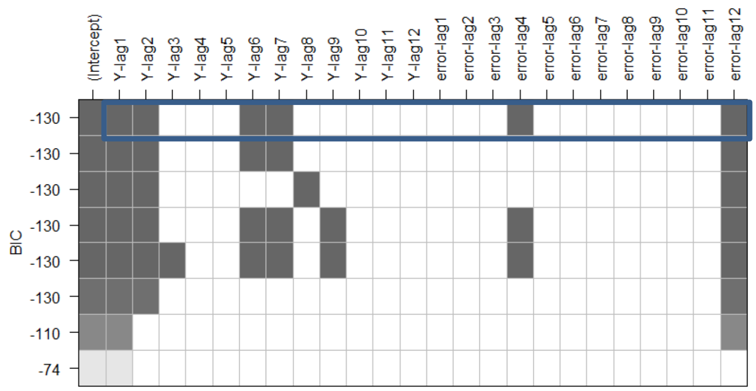


Modelo 4: Tablero 18x18, método ‘ols’, renglón 1 e incluir a Φ_3

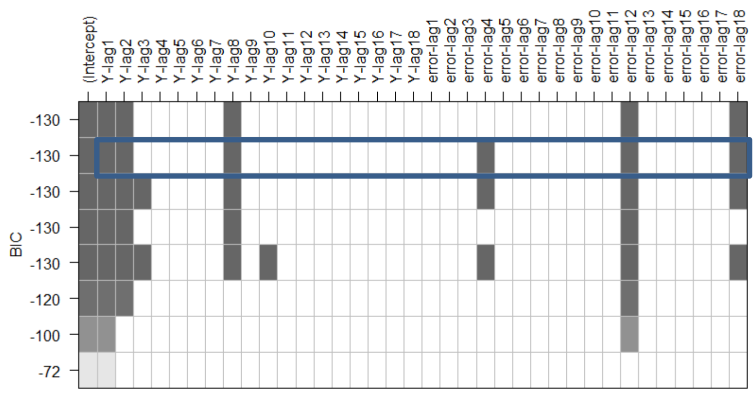


En Datos 20:

Modelo 3: Tablero 12x12, método 'ols', renglón 1

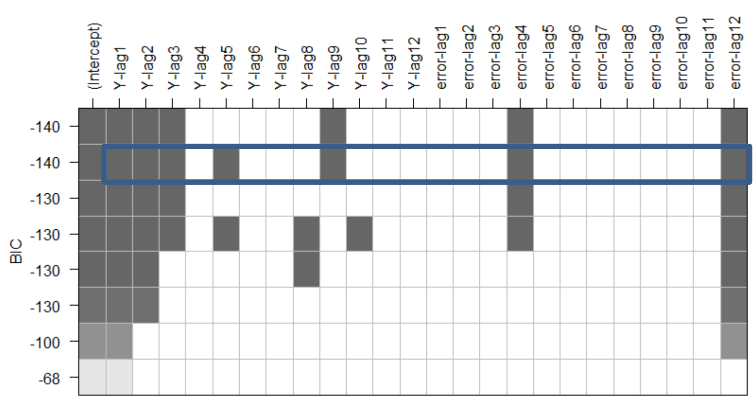


Modelo 4: Tablero 18x18, método 'ols', renglón 2

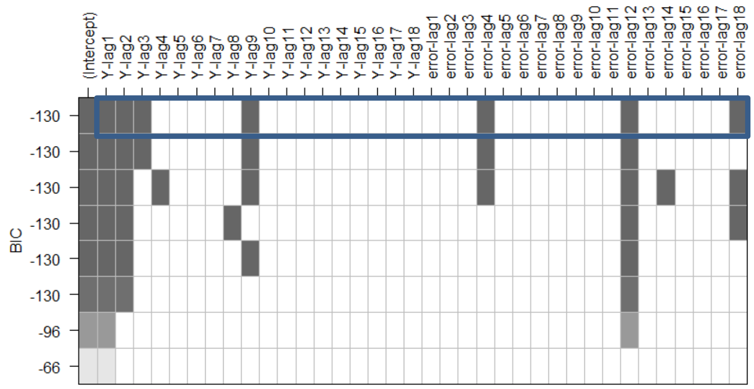


En Datos 21:

Modelo 3: Tablero 12x12, método 'ols', renglón 2

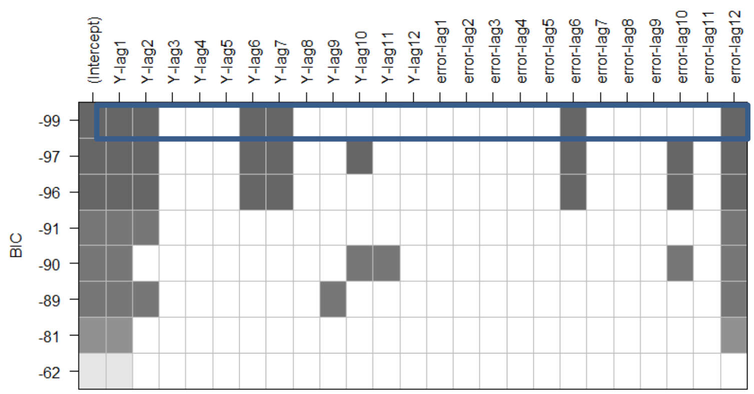


Modelo 4: Tablero 18x18, método ‘ols’, renglón 1

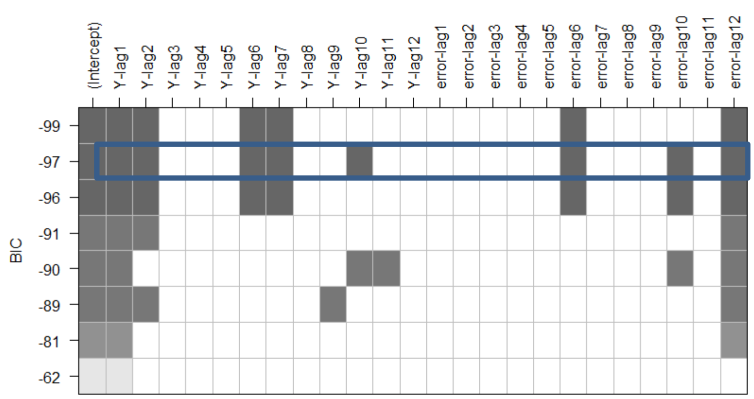


En Datos 22:

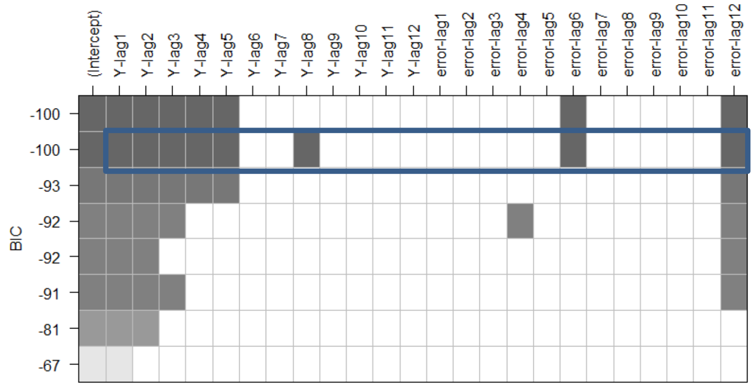
Modelo 3: Tablero 12x12, método ‘ols’, renglón 1 e incluir a $\theta_3, \Phi_1, \Theta_2$



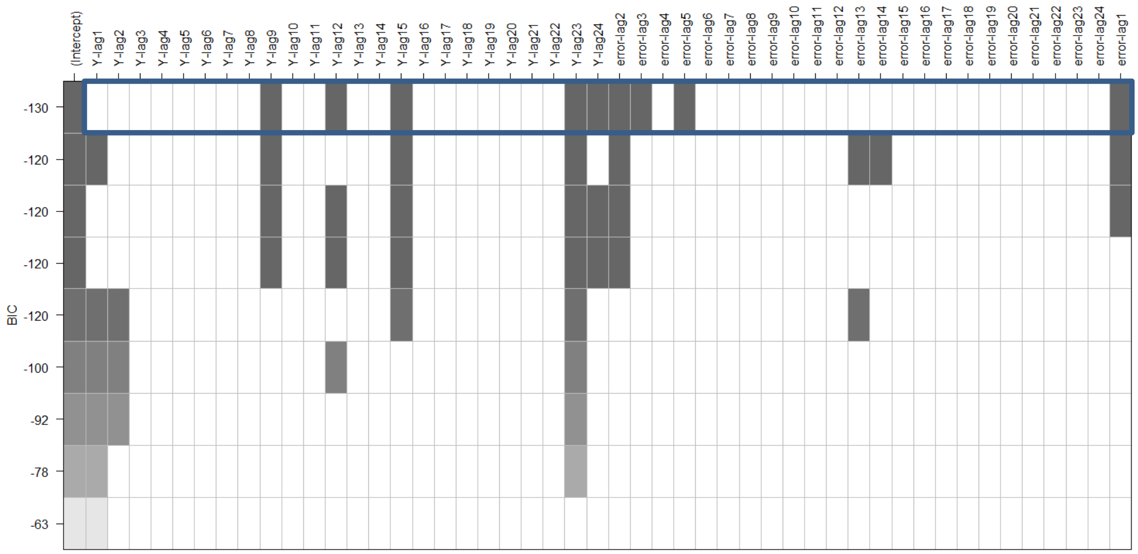
Modelo 4: Tablero 12x12, método ‘ols’, renglón 2 e incluir a a $\theta_3, \Phi_1, \Theta_2$



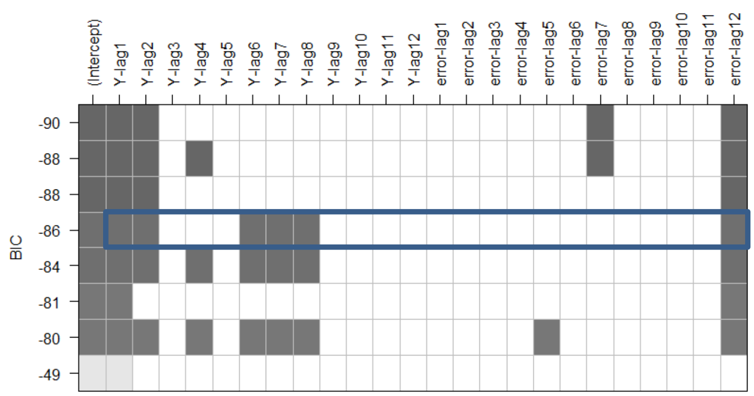
En Datos 23:
Modelo 3: Tablero 12x12, método 'ols', renglón 2 e incluir ϕ_6, Φ_2



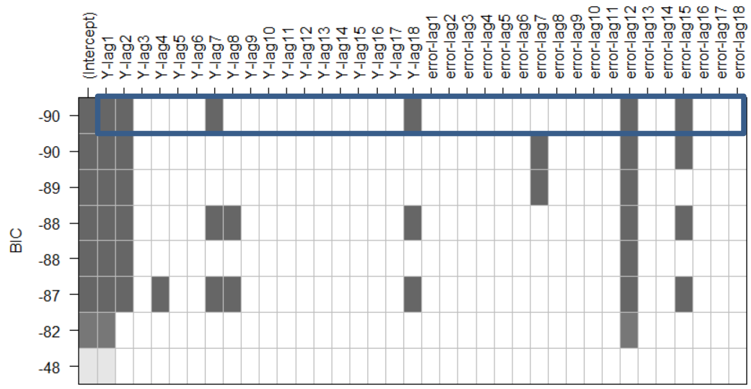
Modelo 4: Tablero 24x24, método 'ols', renglón 1 e incluir $\phi_1, \phi_2, \phi_5, \theta_{11}, \Theta_1$



En Datos 25:
Modelo 3: Tablero 12x12, método 'ols', renglón 4

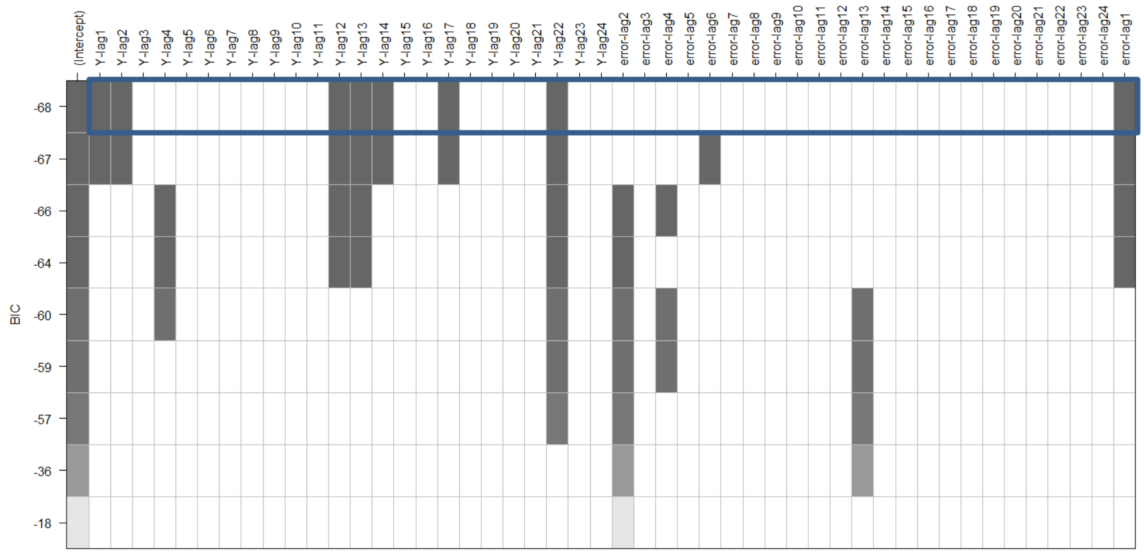


Modelo 4: Tablero 18x18, método ‘ols’, renglón 1

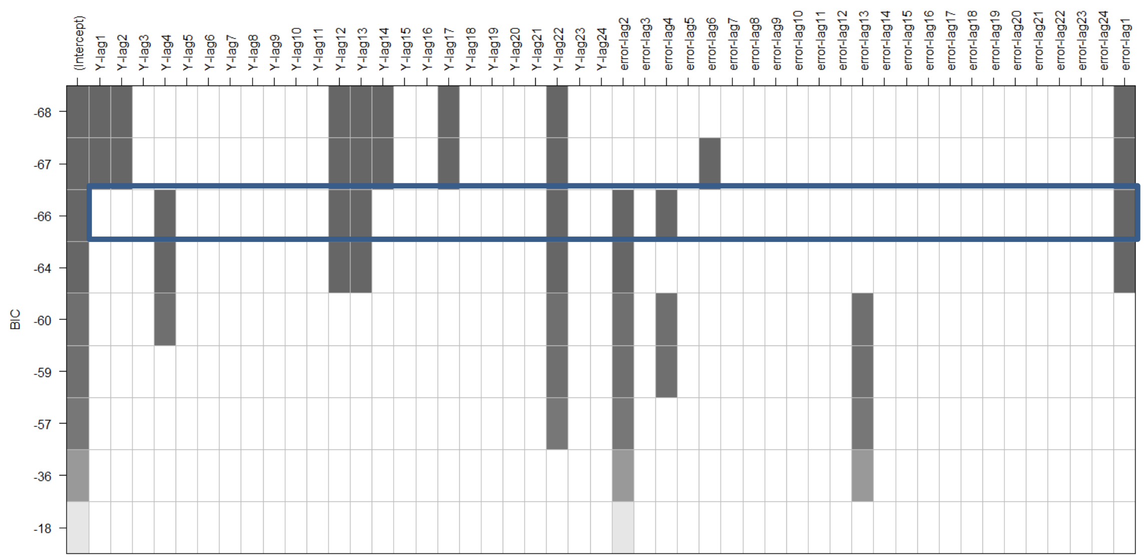


En Datos 28:

Modelo 3: Tablero 24x24, método ‘ols’, renglón 1 e incluir a Θ_2

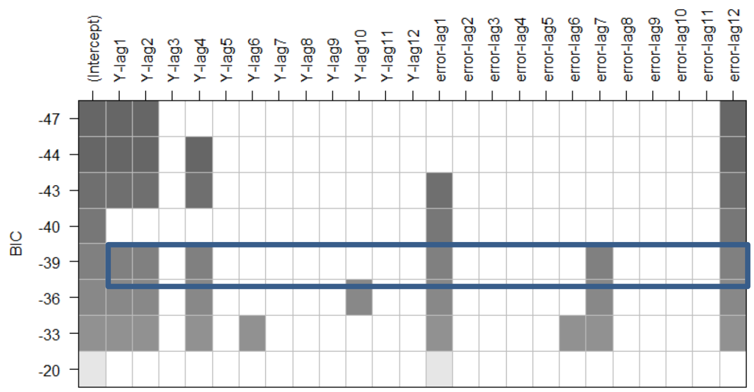


Modelo 4: Tablero 24x24, método ‘ols’, renglón 3 e incluir a Θ_2

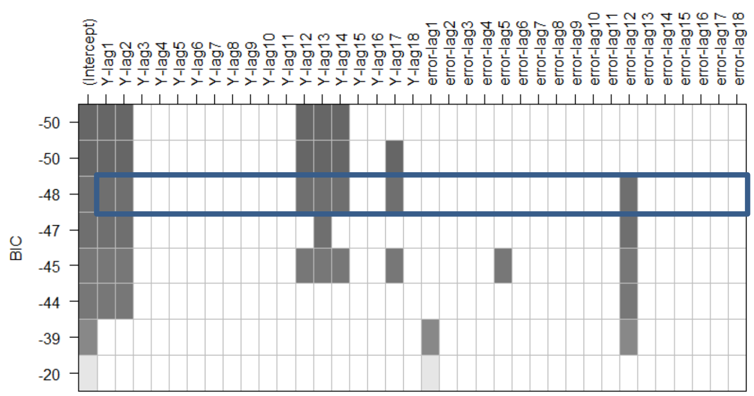


En Datos 29:

Modelo 3: Tablero 12x12, método 'ols', renglón 5 e incluir a Φ_1

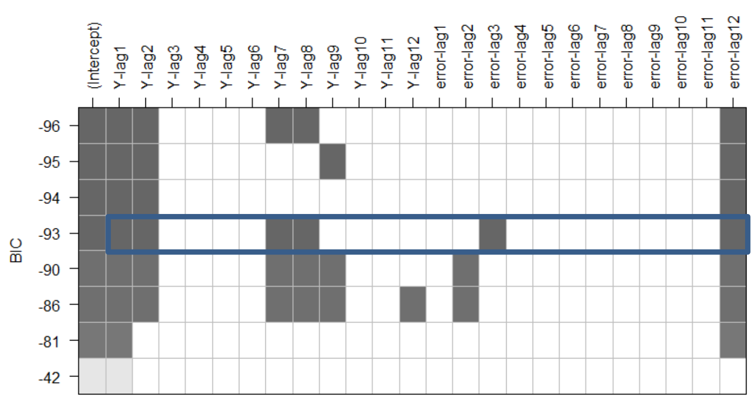


Modelo 4: Tablero 18x18, método 'ols', renglón 3

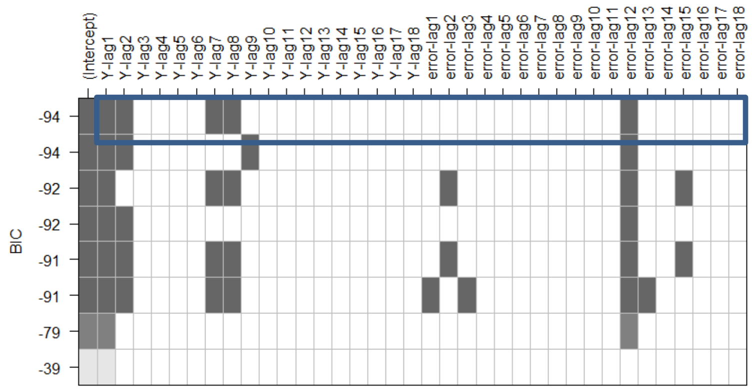


En Datos 31:

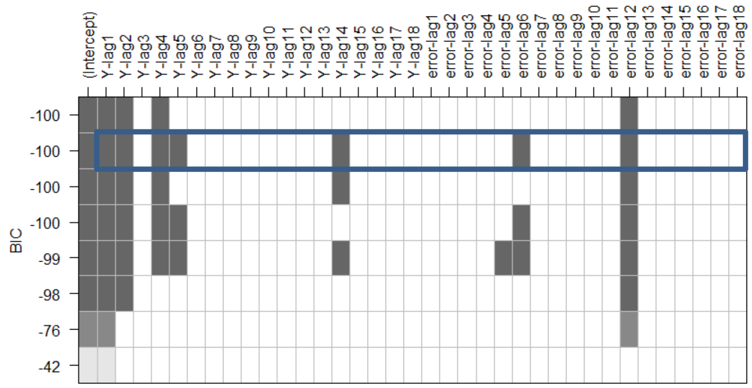
Modelo 3: Tablero 12x12, método 'ols', renglón 4



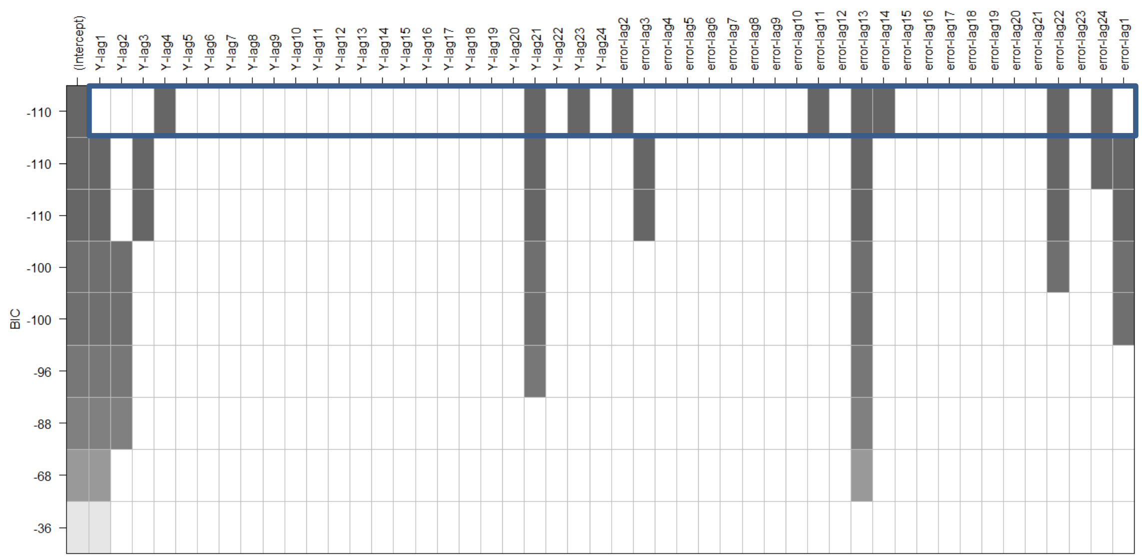
Modelo 4: Tablero 18x18, método ‘ols’, renglón 1 e incluir a ϕ_3



En Datos 32:
Modelo 3: Tablero 18x18, método ‘ols’, renglón 2



Modelo 4. Tablero 24x24, método ‘ols’, renglón 1



G. Ejemplos de identificación con armasubsets(). Ajuste y pronóstico del SARIMA de período s=12, resultante con algunos coeficientes AR ó MA nulos (fijados en cero).

Considere la serie mensual (s=12) que se ilustra a continuación sobre No. de pasajeros transportados en aerolíneas internacionales (en miles), enero de 1959 a Diciembre de 1960.

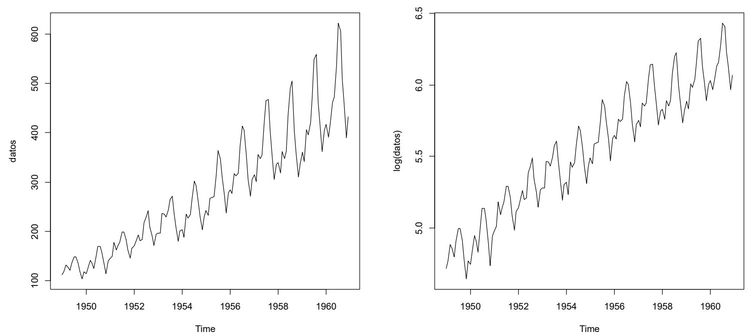


Figura 1. Serie "Air Passengers"

Se ha identificado los órdenes p, q, P y Q de un modelo SARIMA para el logaritmo de la serie arriba con sólo la primera diferencia regular (es decir, $\nabla \log(Y_t)$, de donde d=1 y D=0) al aplicar la función armasubsets sobre $\nabla \log(Y_t)$, usando sólo los primeros $n = 132$ datos (de enero de 1959 a diciembre de 1959),

```
library(TSA);library(lmtest);library(forecast)
SERIESG=ts(scan(file.choose(),skip=2),freq=12,start=c(1949,1)) #Archivo SERIESG.1.DAT
plot(SERIESG);plot(log(SERIESG))
n=length(SERIESG)-12 #validación cruzada será con los últimos 12 datos
t=1:n
yt=ts(SERIESG[t],freq=12,start=c(1949,1)) #Datos para el ajuste
tnuevo=(n+1):length(SERIESG)
ytf=ts(SERIESG[tnuevo],freq=12,start=c(1960,1)) #Datos para la validación cruzada
diflog1=diff(log(yt)) #Primera diferencia regular
plot(armasubsets(diflog1,nar=18,nma=18))
```

se obtiene la siguiente gráfica:

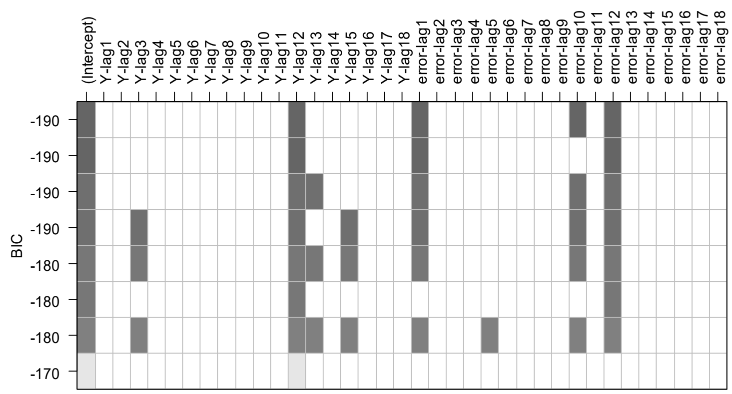


Figura 2. armasubsets 18x18 sobre $\nabla \log(Y_t)$

Apliquemos las reglas vistas en clase al modelo en el primer renglón superior: en la parte AR sólo es señalada la casilla j=12 (se asigna a Φ_1) y en la parte MA son señaladas las casillas i=1 (se asigna a θ_1), i=10 (se asigna a θ_{10}) e i=12 (se asigna a Θ_1). Por tanto, se obtiene que en la parte autorregresiva p=0, P=1 y en la parte MA se tiene q=10 y Q=1 con los siguientes polinomios:

- Polinomio AR regular no hay
- Polinomio MA regular $\theta_{10}(B) = 1 + \theta_1 B + \theta_{10} B^{10}$
- Polinomio AR estacional $\Phi_1(B^{12}) = 1 - \Phi_1 B^{12}$
- Polinomio MA estacional $\Theta_1(B^{12}) = 1 + \Theta_1 B^{12}$

Para ajustar el modelo SARIMA(0,1,10)(1,0,1)$_{[12]}$ con los parámetros identificados usamos lo siguiente:

```
ml1=Arima(log(yt),order=c(0,1,10),seasonal=list(order=c(1,0,1)),fixed=c(NA,rep(0,8),NA,NA,NA),method='ML')
coeftest(ml1) #Tabla de parámetros estimados, valores P bajo distribución N(0,1)
```

Note en Arima() el vector que se especifica en el argumento fixed=: Potencialmente se tienen en total 10 parámetros MA regular, uno en la parte AR estacional y uno en la parte MA estacional para un total de 12 parámetros, pero teniendo en cuenta que para la parte

MA regular sólo θ_1, y θ_{10} son distintos de cero, indicamos con NA's los parámetros que deben ser estimados y con 0's los parámetros que se deben fijar en cero, así:

$(\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6,\theta_7,\theta_8,\theta_9,\theta_{10},\Phi_1,\Theta_1)$: `fixed=c(NA,0,0,0,0,0,0,0,0,NA,NA,NA)` o abreviadamente, `fixed=c(NA,rep(0,8),NA,NA,NA).`

El resultado es el siguiente

```
z test of coefficients:
       Estimate Std. Error  z value  Pr(>|z|)
ma1  -0.2992084  0.0963389  -3.1058  0.001898 **
ma10 -0.0309524  0.1023838  -0.3023  0.762410
sar1  0.9903978  0.0053262 185.9470 < 2.2e-16 ***
sma1 -0.5608384  0.0812253  -6.9047  5.03e-12 ***
```

Suponga ahora que se desea correr el modelo en el renglón 5 de arriba hacia abajo. Siguiendo las reglas vistas en clase, se identifica en la parte autorregresiva que p=3, P=1, y en la parte MA leemos q=10, Q=1, así: Las casillas indicadas en la parte AR son j=3 (se asigna a ϕ_3), j=12 (se asigna a Φ_1), y las casillas j=13 y 15 en las cuales se debe aplicar la regla 4 vista en clase:

- para j=13 se tiene $12\boldsymbol{m} < 13 < 12(\boldsymbol{m}+1)$, entonces **m**=1 y $l = j - 12\boldsymbol{m} = 13 - 12 = 1$, de donde se identifican a $\Phi_{\mathrm{m}} = \Phi_1$ y $\phi_l = \phi_1$
- para j=15 se tiene $12\boldsymbol{m} < 15 < 12(\boldsymbol{m}+1)$, entonces **m**=1 y $l = j - 12\boldsymbol{m} = 15 - 12 = 3$, de donde se identifica de nuevo a $\Phi_{\mathrm{m}} = \Phi_1$ y a $\phi_l = \phi_3$.

En la parte MA las casillas indicadas son i=1, 10 y 12 y por tanto los resultados en esta parte son idénticos a los obtenidos en el modelo del primer renglón. Luego, se tienen finalmente lo siguiente:

- polinomio AR regular $\phi_3(B) = 1 - \phi_1 B - \phi_3 B^3$
- Polinomio MA regular $\theta_{10}(B) = 1 + \theta_1 B + \theta_{10} B^{10}$
- polinomio AR estacional $\Phi_1(B^{12}) = 1 - \Phi_1 B^{12}$
- Polinomio MA estacional $\Theta_1(B^{12}) = 1 + \Theta_1 B^{12}$

Para el vector de coeficientes se tiene la siguiente especificación del argumento fixed:

$(\phi_1,\phi_2,\phi_3,\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6,\theta_7,\theta_8,\theta_9,\theta_{10},\Phi_1,\Theta_1)$: `fixed=c(NA,0,NA,NA,0,0,0,0,0,0,0,0,NA,NA,NA),`
o abreviadamente, `fixed=c(NA,0,NA,NA,rep(0,8),NA,NA,NA).`

La programación para el SARIMA(3,1,10)(1,0,1)[12] con los parámetros identificados sería entonces la siguiente:

```
ml3=Arima(log(yt),order=c(3,1,10),seasonal=list(order=c(1,0,1)),fixed=c(NA,0,NA,NA,rep(0,8),NA,NA,NA),method='ML')
coeftest(ml3) #Tabla de parámetros estimados, valores P bajo distribución N(0,1)
```

Con lo que se obtiene lo siguiente

```
z test of coefficients:
       Estimate Std. Error  z value  Pr(>|z|)
ar1    0.1471253  0.3221166   0.4567   0.64785
ar3   -0.1589495  0.0891955  -1.7820   0.07474 .
ma1   -0.4298201  0.2953290  -1.4554   0.14556
ma10  -0.0277739  0.0943700  -0.2943   0.76852
sar1   0.9897186  0.0057252 172.8720 < 2.2e-16 ***
sma1  -0.5598709  0.0862184  -6.4936 8.379e-11 ***
```

Ahora considere el modelo en el quinto renglón de la siguiente gráfica del armasubsets() aplicado al logaritmo natural de la serie con una diferencia regular y una estacional (es decir, $\nabla_{12}\nabla \log(Y_t)$, por tanto, d=D=1):

```
diflog1.12=diff(diff(log(yt)),lag=12)
plot(armasubsets(diflog1.12,nar=18,nma=18))
```

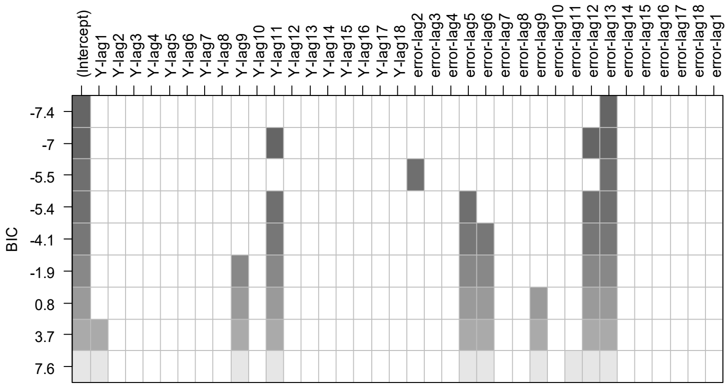


Figura 3. armasubsets 18x18 sobre $\nabla_{12}\nabla\log(Y_t)$

En este renglón en la parte autorregresiva sólo está señalada la casilla j=11 (se asigna a ϕ_{11}) indicando que p=11, P=0, y en la parte MA las casillas señaladas son i=5 (se asigna a θ_5), i=6 (se asigna a θ_6), i=12 (se asigna a Θ_1), e i=13 (debe aplicar la regla 3). Note que para la casilla i=13 aplicando la cuarta de las reglas dadas en clase, se obtiene lo siguiente:
$12\boldsymbol{k} < 13 < 12(\boldsymbol{k}+1)$ entonces $\boldsymbol{k} = 1$ y $r = i - 12\boldsymbol{k} = 13 - 12 = 1$, de donde se identifican a los coeficientes $\Theta_{\mathrm{k}} = \Theta_1$ y a $\theta_r = \theta_1$.
Por tanto, finalmente tenemos que
Polinomio AR regular $\phi_1(B) = 1 - \phi_{11}B^{11}$
Polinomio MA regular $\theta_1(B) = 1 + \theta_1 B + \theta_5 B^5 + \theta_6 B^6$
Polinomio AR estacional no hay
Polinomio MA estacional $\Theta_1(B^{12}) = 1 + \Theta_1 B^{12}$
Para el vector de coeficientes se tiene la siguiente especificación del argumento fixed:

$(\phi_1,\phi_2,\phi_3,\phi_4,\phi_5,\phi_6,\phi_7,\phi_8,\phi_9,\phi_{10},\phi_{11},\theta_1,\theta_2,\theta_3,\theta_4,\theta_5,\theta_6,\Theta_1)$: `fixed=c(rep(0,10),NA,NA,rep(0,3),NA,NA,NA)`

Luego, se ajusta el modelo SARIMA(11,1,6)(0,1,1)$_{[12]}$ con los coeficientes identificados, así:

```
ml2=Arima(log(yt),order=c(11,1,6),seasonal=list(order=c(0,1,1)),fixed= c(rep(0,10),NA,NA,rep(0,3),NA,NA,NA),method='ML')
coeftest(ml2) #Tabla de parámetros estimados, valores P bajo distribución N(0,1)
```

Con lo que se obtiene lo siguiente

```
z test of coefficients:
       Estimate Std. Error z value  Pr(>|z|)
ar11 -0.0019191  0.0946431 -0.0203 0.9838225
ma1  -0.3415655  0.0965592 -3.5374 0.0004041 ***
ma5   0.0433637  0.0962457  0.4506 0.6523122
ma6   0.0100745  0.0946821  0.1064 0.9152619
sma1 -0.5566359  0.0796625 -6.9874   2.8e-12 ***
```

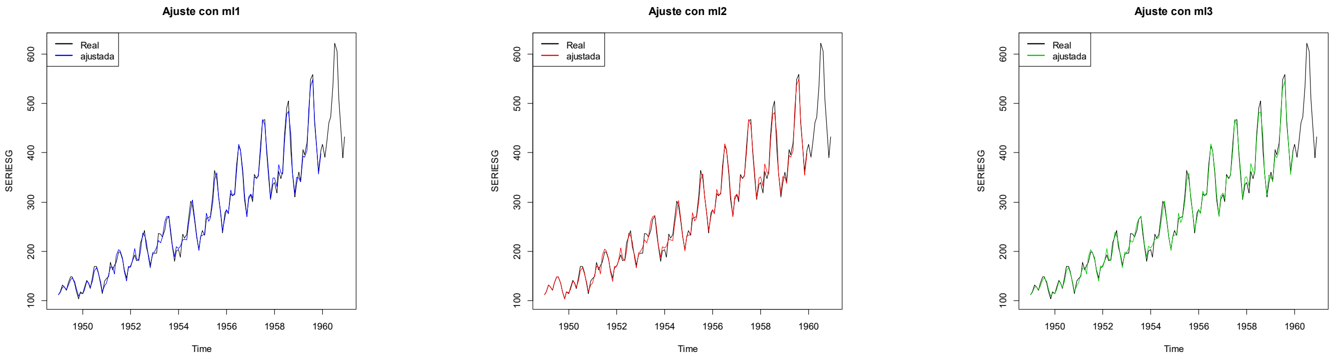
Los ajustes y pronósticos para h=12 de los tres modelos presentados, en la escala original, se obtendrían de la siguiente manera (recuerde que en este ejemplo se recurrió a transformación log):

```
yhatml1=exp(ml1$fitted)*exp(ml1$sigma2/2) #valores ajustados en escala original
pronml1=exp(as.data.frame(forecast(ml1,h=12,level=95)))*exp(ml1$sigma2/2) #pronósticos en escala original
pronml1=ts(pronml1,freq=12,start=c(1960,1)); pronml1
accuracy(pronml1[,1],ytf)

yhatml2=exp(ml2$fitted)*exp(ml2$sigma2/2)
pronml2=exp(as.data.frame(forecast(ml2,h=12,level=95)))*exp(ml2$sigma2/2)
pronml2=ts(pronml2,freq=12,start=c(1960,1)); pronml2
accuracy(pronml2[,1],ytf)

yhatml3=exp(ml3$fitted)*exp(ml3$sigma2/2)
pronml3=exp(as.data.frame(forecast(ml3,h=12,level=95)))*exp(ml3$sigma2/2)
pronml3=ts(pronml3,freq=12,start=c(1960,1)); pronml3
accuracy(pronml3[,1],ytf)
#Gráficas de los ajustes
plot(SERIESG,main="Ajuste con ml1"); lines(yhatml1,col=4)
legend("topleft",legend=c("Real","ajustada"),col=c(1,4),lwd=2)
plot(SERIESG,main="Ajuste con ml2"); lines(yhatml2,col=2)
legend("topleft",legend=c("Real","ajustada"),col=c(1,2),lwd=2)
plot(SERIESG,main="Ajuste con ml3"); lines(yhatml3,col=3)
legend("topleft",legend=c("Real","ajustada"),col=c(1,3),lwd=2)
```

Observe las gráficas de la serie y sus ajustes:



(a) (b) (c)

Figura 4. Ajustes modelos SARIMA sobre serie "Air Passengers". (a) Con modelo ml1; (b) con modelo ml2; (c) con modelo ml3

NOTA: Recuerde que debe escribir la ecuación teórica específica hallada usando sólo los parámetros que fueron identificados y no despeje a Y_t (si identificó directamente sobre la serie sin transformar) o $\log(Y_t)$ (si identificó directamente sobre la serie transformada por logaritmo natural).

H. Cómo manejar en R y en Word un formato apropiado para gráficas del armasubsets

Por ejemplo, con s=12: En R, suponga que **diffdDYt** representa a $W_t = \nabla_{12}^{D}\nabla^{d}Y_t$ (los órdenes de acuerdo a como se le indicaron en la Tabla 1 de asignación de modelos) el cual se considera estacionario, entonces los gráficos de **armasubsets**, obtenerlos así (use los argumentos nar y ma que se le hayan indicado en modelos 3 y 4):

```
win.graph(heigh=5,width=9)
plot(armasubsets(diffdDYt,nar=12,nma=12,y.name='AR',ar.method='ols'))
win.graph(heigh=5,width=9)
plot(armasubsets(diffdDYt,nar=18,nma=18,y.name='AR',ar.method="ols"))
win.graph(heigh=5,width=9)
plot(armasubsets(diffdDYt,nar=24,nma=24,y.name='AR',ar.method='ols'))
```

Para cada tablero, corra líneas de programa y dar click derecho sobre la gráfica resultante para elegir opción de "Copiar como mapa de bits"

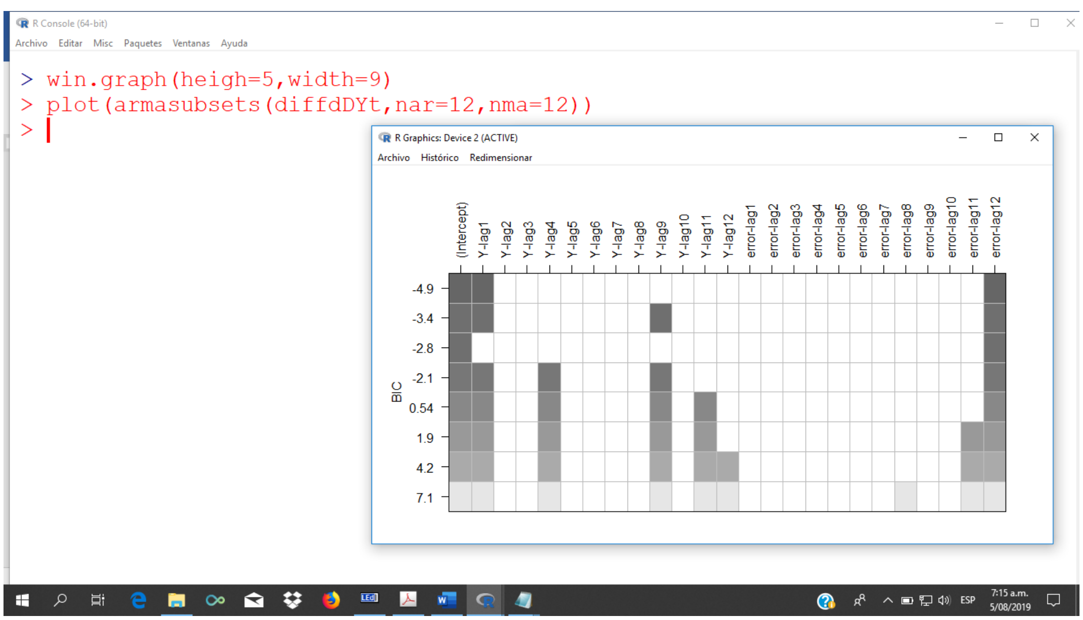


Figura 5. Obtención gráfico armasubsets

Pegar en archivo de Word imagen y con click derecho sobre la imagen, abrir menú de edición y seleccionar en ventana emergente la opción "Tamaño y posición"

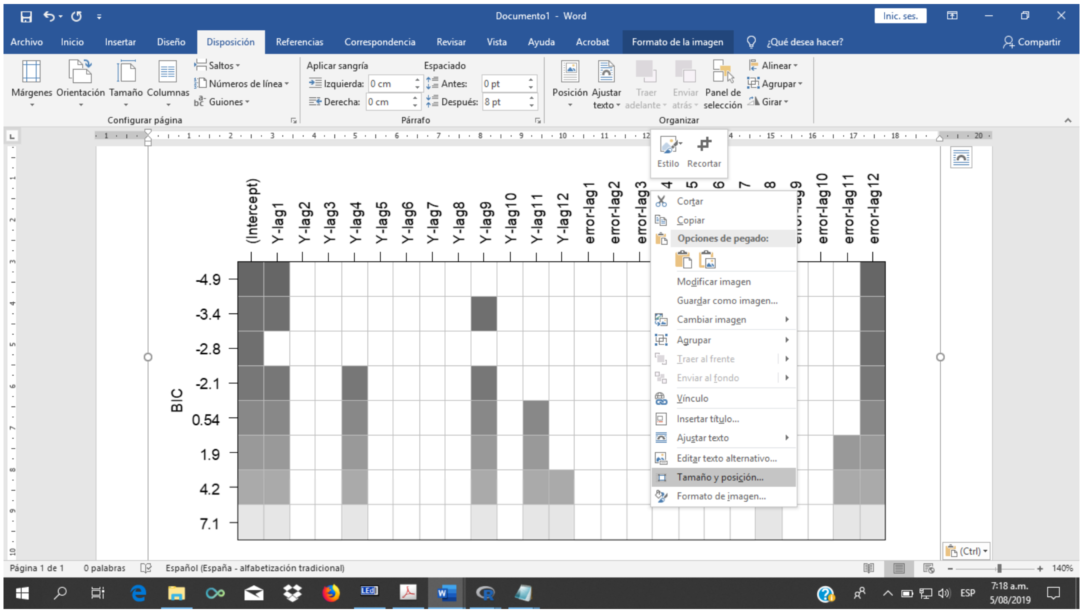


Figura 6. Gráfico de armasubsets pegada en Word y apertura de ventana para editarla

Finalmente, en la ventana emergente llamada "Diseño", deshabilitar las opciones "Bloquear relación de aspecto" y "Proporcional al tamaño original de la imagen" y especificar en "Alto" opción "Absoluto" un valor de 6cm y "Ancho" opción "Absoluto" 10,5cm

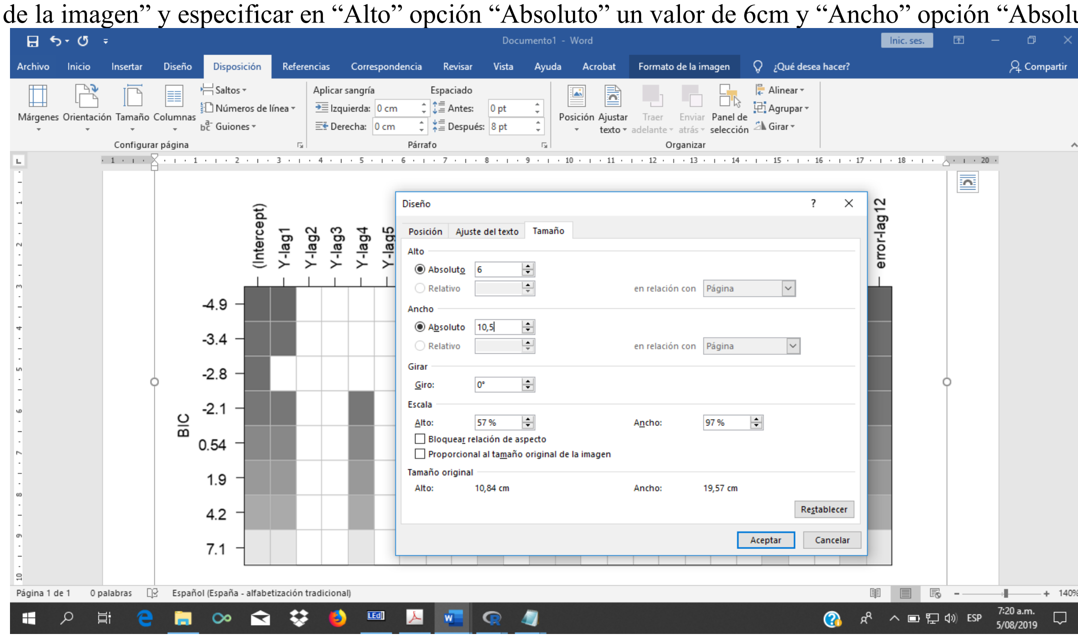


Figura 7. Gráfico de armasubsets pegada en Word, cambio de dimensiones

La imagen resultante es la siguiente:

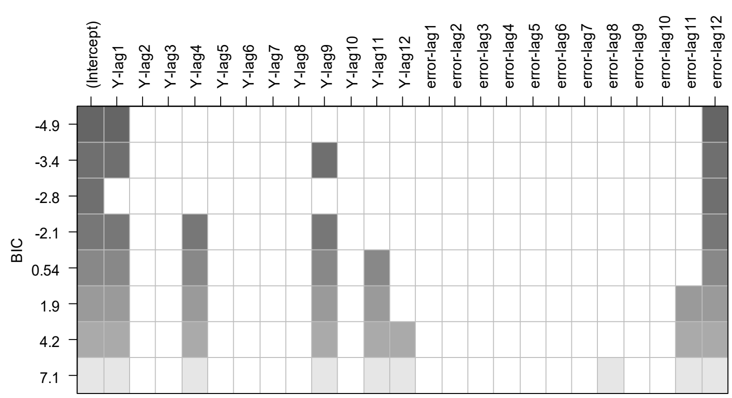


Figura 8. Apariencia final de gráfico de armasubsets 12x12

Siguiendo los mismos pasos para las otras dos gráficas, el resultado es el siguiente:

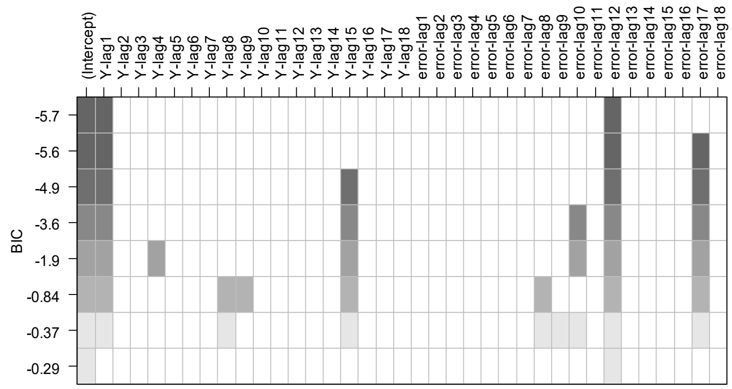


Figura 9. Apariencia final de gráfico de armasubsets de 18x18

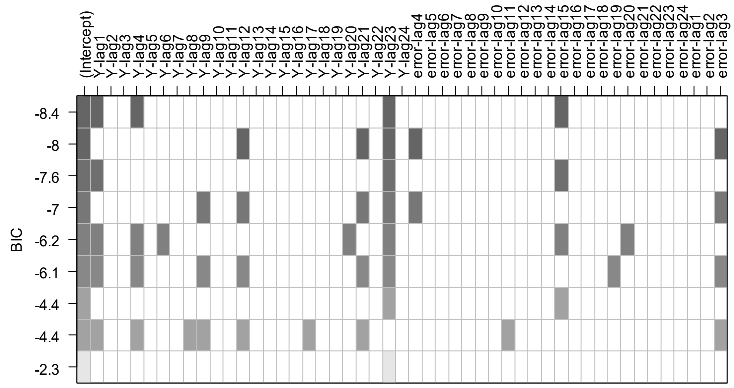


Figura 10. Apariencia final de gráfico de armasubsets de 24x24