



Tarea #3

*Estudiante: Luciano Andres Juárez López***Problema 1**

Supongamos que tienes un vector $Y_t = (Y_{1t}, Y_{2t}, Y_{3t}, Y_{4t})$, donde:

- Y_{3t} tiene un cambio estructural,
- Y_{1t} tiene estacionalidad,
- Y_{2t} es estacionaria,
- Y_{4t} es $I(1)$ (integrada de orden 1).

(Solución)

Para ello generamos las series de la siguiente manera:

```
1  # Tamaño de muestra
2  n <- 100
3
4  # Y1t: serie con estacionalidad
5  Y1t <- ts(sin(2 * pi * (1:n) / 12) + rnorm(n), frequency = 12)
6
7  # Y2t: serie estacionaria
8  Y2t <- arima.sim(list(order = c(1, 0, 1), ar = 0.5, ma = -0.3), n = n)
9
10 # Y3t: serie con cambio estructural
11 Y3t <- c(rnorm(n/2, mean = 5), rnorm(n/2, mean = 10))
12
```

```

13 # Y4t: serie integrada de orden 1
14 Y4t <- cumsum(rnorm(n))

```

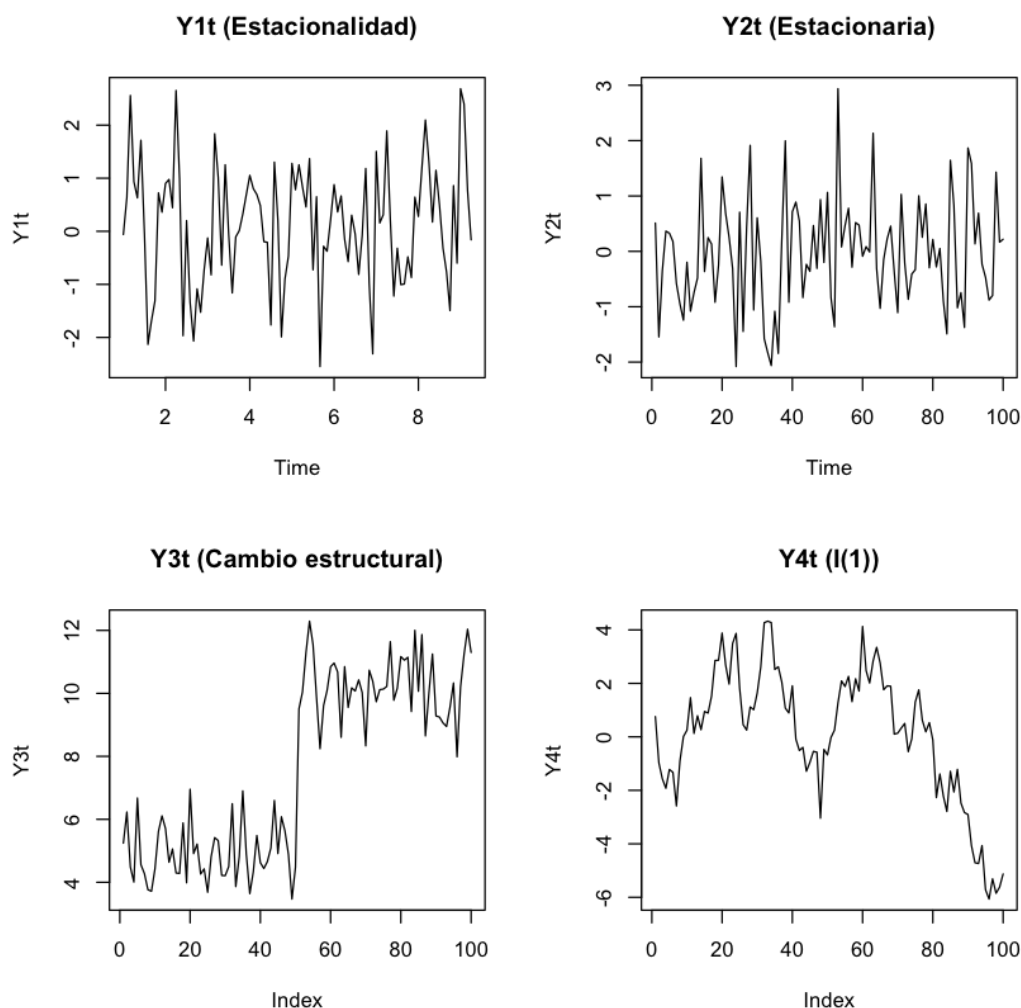


Figura 1: Graficamos las series de tiempo

Por lo que para estimar un modelo VAR, requerimos que las series sean estacionarias, como Y2t ya es estacionaria, requerimos hacer a las demas series estacionarias. Podemos observar en 1 que Y1t parece ser estacionaria, por lo que aplicamos la prueba ADF, obteniendo lo siguiente:

```

1 Augmented Dickey-Fuller Test
2 data: Y1t
3 Dickey-Fuller = -6.3606, Lag order = 4, p-value = 0.01
4 alternative hypothesis: stationary

```

Donde el p-valor es igual a 0.01, por lo que se rechaza la hipotesis nula, concluyendo que es estacionaria.

Para hacer que Y3t y Y4t sean estacionarias, primero aplicamos un retardo a cada una de ellas:

```

1 # Aplicamos una diferenciacion a las series

```

```

2 Y3t _ diff <- diff(Y3t)
3 Y4t _ diff <- diff(Y4t)

```

Y aplicando la prueba ADF a estas series diferenciadas, obtenemos lo siguiente:

Para la serie Y3t, observamos que el p-valor es igual a 0.01, por lo que se rechaza la hipótesis nula, concluyendo que es estacionaria.

```

1 Augmented Dickey-Fuller Test
2 data: Y3t_diff
3 Dickey-Fuller = -5.708, Lag order = 4, p-value = 0.01
4 alternative hypothesis: stationary

```

Para la serie Y4t, observamos que el p-valor es igual a 0.01, por lo que se rechaza la hipótesis nula, concluyendo que es estacionaria.

```

1 Augmented Dickey-Fuller Test
2 data: Y4t_diff
3 Dickey-Fuller = -4.2066, Lag order = 4, p-value = 0.01
4 alternative hypothesis: stationary

```

Una vez que hemos comprobado que las series son estacionarias, procedemos a combinar las series de tiempo y determinar el número óptimo de rezagos:

```

1 series_combinadas <- cbind(Y1t, Y2t, Y3t_diff, Y4t_diff)
2 library(vars)
3
4 # Determinamos el numero optimo de rezagos
5 lag_selection <- VARselect(series_combinadas, lag_max = 8)
6 print(lag_selection$selection)

```

Obteniendo el siguiente resultado:

```

1 AIC(n): 1 HQ(n): 1 SC(n): 1 FPE(n): 1

```

Por lo que el rezago óptimo de acuerdo a las pruebas realizadas es igual a uno, ahora procedemos a estimar el modelo VAR con las series combinadas y el número de rezagos óptimo.

```

1 var_model <- VAR(series_combinadas, p = 1, type = "none")
2 summary(var_model)

```

Del resultado obtenido, solo podemos observar que las raíces del polinomio característico del modelo VAR son menores a 1, lo que nos dice que están dentro del círculo unitario e indica que el modelo es estable.

También observamos que Y1t y Y3t son significativas para sus respectivas ecuaciones.

```

1
2 VAR Estimation Results:
3 =====
4 Roots of the characteristic polynomial:
5 0.3788 0.2256 0.2256 0.1311
6
7 Estimation results for equation Y1t:
8 =====
9 Y1t = Y1t.l1 + Y2t.l1 + Y3t_diff.l1 + Y4t_diff.l1
10
11 Estimate Std. Error t value Pr(>|t|)

```

```

12 Y1t.l1      0.32153      0.09657      3.330      0.00124 **
13 Y2t.l1     -0.23358      0.11259     -2.075      0.04076 *
14 Y3t_diff.l1 0.03646      0.07884      0.463      0.64478
15 Y4t_diff.l1 -0.05683      0.11018     -0.516      0.60722
16 ---
17 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
18
19
20 Residual standard error: 1.086 on 94 degrees of freedom
21 Multiple R-Squared:  0.135, Adjusted R-squared:  0.09817
22 F-statistic: 3.667 on 4 and 94 DF, p-value: 0.008056
23
24 Estimation results for equation Y3t_diff:
25 =====
26 Y3t_diff = Y1t.l1 + Y2t.l1 + Y3t_diff.l1 + Y4t_diff.l1
27
28           Estimate Std. Error t value Pr(>|t|)
29 Y1t.l1      0.07797    0.11712    0.666 0.507197
30 Y2t.l1      0.08299    0.13655    0.608 0.544795
31 Y3t_diff.l1 -0.34770    0.09562   -3.636 0.000451 ***
32 Y4t_diff.l1  0.18155    0.13363    1.359 0.177517
33 ---
34 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
35
36
37 Residual standard error: 1.318 on 94 degrees of freedom
38 Multiple R-Squared:  0.144, Adjusted R-squared:  0.1076
39 F-statistic: 3.954 on 4 and 94 DF, p-value: 0.005201

```

Prueba de normalidad de los residuos

Ahora comprobamos si los residuos siguen una distribución normal, este supuesto es muy importante para la inferencia en el modelo VAR.

```

1 normality_test <- normality.test(var_model)
2 # Imprimir los resultados de las pruebas
3 print(normality_test)

```

De los datos obtenidos, concluimos lo siguiente:

1. Para el **JB-Test**: El p-valor es igual a 0.1519, por lo que no se rechaza la hipótesis de que el modelo sigue una distribución normal multivariada.
2. **Skewness (Asimetría)**: El p-valor es igual a 0.1964, por lo que no hay evidencia suficiente para concluir que los residuos presentan una asimetría significativa.
3. **Kurtosis**: El p-valor es igual a 0.2031, por lo que no se rechaza la hipótesis de que los residuos tienen una kurtosis normal.

Por lo tanto, los resultados sugieren que no hay evidencia significativa para rechazar la normalidad de los residuos del modelo VAR.

Prueba de autocorrelación de los residuos

Verificamos si los residuos del modelo VAR presentan autocorrelación, ya que la presencia de autocorrelación indicaría que el modelo no ha capturado adecuadamente la estructura temporal de los datos. Realizamos la prueba de autocorrelación en series de los residuales:

```
1 arch_test <- arch.test(var_model)
2 # Imprimir los resultados de las pruebas
3 print(arch_test)
```

Obtenemos que el p-valor es igual a 0.1104, entonces no se rechaza la hipótesis nula, pues no hay suficiente evidencia estadística para concluir que existe heterocedasticidad en los residuos del modelo, lo que nos dice que los errores del modelo tienen varianza constante (son homocedásticos).

Causalidad de Granger

Dado que los resultados del modelo VAR sugieren que las relaciones más significativas involucran a $Y1t$ y $Y3t$, nos enfocamos en hacer las pruebas de causalidad de Granger en solo esas dos variables, realizando las siguientes pruebas:

```
1 causality(var_model, cause = "Y1t") # Probamos si Y1t causa a otras
  variables
```

Obtenemos que el p-valor es igual a 0.1443, por lo que no se puede rechazar la hipótesis nula. Lo que nos sugiere que no hay suficiente evidencia para concluir que $Y1t$ causa a las otras variables en el modelo.

```
1 causality(var_model, cause = "Y3t" ) # Probamos si Y3t causa a otras
  variables
```

Obtenemos que el p-valor es igual a 0.5248, por lo que no se puede rechazar la hipótesis nula. Lo que nos sugiere que no hay suficiente evidencia para concluir que $Y3t$ causa a las otras variables en el modelo.

Conclusión

Aunque no se encontró una relación de causa y efecto clara entre las variables, el modelo cumple con los principales supuestos para la inferencia, tales como la estacionariedad de las series temporales una vez aplicado las transformaciones, la normalidad de los residuos, la ausencia de autocorrelación en los mismos y la homocedasticidad (varianza constante de los errores). Esto sugiere que el modelo es adecuado para realizar predicciones y análisis basados en las relaciones observadas entre las variables.

Problema 2

Para un modelo VAR(p) genera un procedimiento para generar matrices A_i estacionarias controlando la estabilidad del modelo como:

- i) Leve
- ii) Moderada
- iii) Fuerte.

Hint: utiliza la representación VAR(1) para un modelo VAR de orden p y estudia los valores propios de la solución de $|I_{Kp} - zA|$.

(Solución)

Para realizar un procedimiento que genere matrices A_i estacionarias controlando la estabilidad del modelo VAR(p), utilizaremos la representación VAR(1) para un modelo VAR de orden p , la representación tiene la siguiente forma:

$$Y_t = AY_{t-1} + \epsilon_t$$

Donde A es la matriz que contiene las matrices A_i para $i = 1, \dots, p$ asociadas al VAR(p). En la literatura, esta matriz A tiene una estructura general ya definida en un modelo VAR(p), la cual esta dada por bloques de matrices autoregresivas, la cual tiene la siguiente forma:

$$A = \begin{pmatrix} A_1 & A_2 & A_3 & \dots & A_p \\ I_k & 0 & 0 & \dots & 0 \\ 0 & I_k & 0 & \dots & 0 \\ 0 & 0 & I_k & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & I_k \end{pmatrix}$$

La matriz A es de tamaño $Kp \times Kp$ (K es el numero de variables y p el orden del modelo VAR), las matrices A_i para $i = 1, \dots, p$ son matrices de coeficientes autoregresivos e I_k es la matriz identidad, todas estas de tamaño $K \times K$.

Generación de las matrices A_i

Se generan matrices A_i de tamaño $K \times K$, controlando la estabilidad de cada matriz a través de la selección de los valores propios λ_i . Los niveles de estabilidad considerados son:

- **Leve:** Los valores propios λ_i se seleccionan de manera aleatoria en el rango $[0.6, 0.9]$.
- **Moderada:** Los valores propios λ_i se seleccionan de manera aleatoria en el rango $[0.3, 0.6]$.

- **Fuerte:** Los valores propios λ_i se seleccionan de manera aleatoria en el rango $[0.0, 0.3]$.

Para ello utilizamos la siguiente descomposición:

$$A_i = BDB^{-1}$$

Aquí D es una matriz diagonal con los valores λ_i generados y la matriz B se genera utilizando la descomposición QR. Una vez obtenidas las matrices D y B , podemos construir las matrices A_i

Estabilidad de A

Dado que se requiere controlar la estabilidad, se torna complicado al tener de esta forma la matriz A , por lo que se requiere realizar operaciones elementales sobre esta matriz A , para obtener una equivalente y mas facil de trabajar, por lo que obtenemos lo siguiente:

$$\begin{aligned}
 A &= \begin{pmatrix} A_1 & A_2 & A_3 & A_4 & \dots & A_p \\ I_k & 0 & 0 & 0 & \dots & 0 \\ 0 & I_k & 0 & 0 & \dots & 0 \\ 0 & 0 & I_k & 0 & \dots & 0 \\ 0 & 0 & 0 & I_k & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & I_k \end{pmatrix} \\
 &\sim \begin{pmatrix} A_1 & A_2 & A_3 & A_4 & \dots & A_p \\ 0 & A_1^{-1}A_2 & A_1^{-1}A_3 & A_1^{-1}A_4 & \dots & A_1^{-1}A_p \\ 0 & I_k & 0 & 0 & \dots & 0 \\ 0 & 0 & I_k & 0 & \dots & 0 \\ 0 & 0 & 0 & I_k & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & I_k \end{pmatrix} \\
 &\sim \begin{pmatrix} A_1 & 0 & A_3 & A_4 & \dots & A_p \\ 0 & A_1^{-1}A_2 & -A_1^{-1}A_3 & -A_1^{-1}A_4 & \dots & -A_1^{-1}A_p \\ 0 & 0 & A_2^{-1}A_3 & A_2^{-1}A_4 & \dots & A_2^{-1}A_p \\ 0 & 0 & I_k & 0 & \dots & 0 \\ 0 & 0 & 0 & I_k & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & I_k \end{pmatrix} \quad \text{considerando } (A_1^{-1}A_2)^{-1} = A_2^{-1}A_1
 \end{aligned}$$

$$\sim \begin{pmatrix} A_1 & 0 & 0 & A_4 & \dots & A_p \\ 0 & A_1^{-1}A_2 & 0 & -A_1^{-1}A_4 & \dots & -A_1^{-1}A_p \\ 0 & 0 & A_2^{-1}A_3 & A_2^{-1}A_4 & \dots & A_2^{-1}A_p \\ 0 & 0 & 0 & A_3^{-1}A_4 & \dots & A_3^{-1}A_p \\ 0 & 0 & 0 & I_k & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & I_k \end{pmatrix}$$

...

$$\sim \begin{pmatrix} A_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & A_1^{-1}A_2 & 0 & 0 & \dots & 0 \\ 0 & 0 & A_2^{-1}A_3 & 0 & \dots & 0 \\ 0 & 0 & 0 & A_3^{-1}A_4 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & A_{p-1}^{-1}A_p \end{pmatrix}$$

Obteniendo así una matriz equivalente a la matriz A , lo cual hace que sea más fácil realizar un análisis de estabilidad, donde el polinomio característico lo podemos ver como el producto de los elementos de la diagonal, como se muestra a continuación:

$$P_A(S) = P_{A_1}(S) \cdot P_{A_1^{-1}A_2}(S) \cdot \dots \cdot P_{A_{p-1}^{-1}A_p}(S)$$

Por lo que la matriz A será estable si cada bloque de la diagonal es estable.

Problema 3

Para un modelo VAR(p) donde K es el número de series de tiempo y T el número de observaciones, propón un criterio empírico para determinar $p_{\max} = f(K, T)$ en la selección de rezagos óptimos y comenta brevemente sus propiedades para algún conjunto de datos simulados.

(Solución)

Se propone el siguiente criterio empírico para determinar $p_{\max} = f(K, T)$, el cual se basa en la autocorrelación de las series simuladas, de la siguiente manera:

1. **Simulación de series de tiempo:** Se generan K series de tiempo independientes, cada una con T observaciones, simuladas a partir de una distribución normal.
2. **Cálculo de la ACF:** Se calcula la función de autocorrelación (ACF) para cada serie simulada, pues sabemos que la ACF mide la correlación entre los valores de la serie en diferentes rezagos.
3. **Intervalo de confianza:** Definimos un nivel de significancia $\alpha = 0.05$ para establecer un intervalo de confianza para la ACF, donde los valores de la ACF que caen fuera de este intervalo indican una correlación significativa en ese rezago.
4. **Determinación del rezago máximo:** Para cada serie, obtenemos aquellos rezagos donde la ACF cae dentro del intervalo de confianza, después para cada serie, se obtiene el último rezago en el que la ACF es significativa.
5. **Salida:** Se devuelve una lista de rezagos máximos para cada una de las K series, que puede usarse para seleccionar el orden óptimo p_{\max} en el modelo VAR(p).

Donde las propiedades para este criterio, son las siguiente:

- **Dependencia de T :** El tamaño de la muestra (T) afecta la precisión con la que se estima la ACF. Pues para series de tiempo cortas, el intervalo de confianza es más amplio, lo que puede llevar a la selección de valores más pequeños de p_{\max} .
- **Dependencia de K :** A medida que aumenta el número de series de tiempo (K), debemos ser mas cuidadosos con el criterio empírico, pues buscamos el rezago que funcione bien para todas las series, por lo que podemos usar el rezago más alto que se repita en todas ellas.
- **Simplicidad:** Este método es fácil de implementar, pero podría no capturar toda la complejidad de los modelos VAR, donde las series pueden estar correlacionadas en varios rezagos.

Problema 4

Se adjuntan dos bases de datos: `Clima.csv` y `Datos.csv`. La primera incluye datos de temperatura y precipitación promedio en México de 1985 a noviembre de 2022 y la segunda contiene información de la inflación, actividad económica mensual, índice de precios y cotizaciones, y tipo de cambio nominal, todas para México de 1980 a mayo 2022. A partir de estos datos, se pide contestar lo siguiente:

1. ¿La temperatura y la precipitación son causales en el sentido de Granger?
2. ¿Hay relación causal entre alguna de las series de tiempo climáticas con alguna de las series económicas o financieras? ¿Tiene sentido lo encontrado?
3. Genera un modelo VAR(p) de predicción para la temperatura y evalúa su funcionamiento fuera de muestra en los últimos 24 meses. ¿Funciona mejor que un SARIMA(p, d, q)?

(Solución)

- a) Concluimos que la temperatura y la precipitación son causales en el sentido de Granger, ya que se rechazó la hipótesis nula al realizar la prueba de Granger, dado que los p-valores fueron menores a 0.05.
- b) De acuerdo con lo obtenido, generando el modelo VAR y aplicando la prueba de causalidad de Granger, podemos concluir que tanto la temperatura como la precipitación están significativamente relacionadas de manera causal con las series económicas o financieras analizadas en el modelo VAR. Lo hallado en esta sección tiene sentido en varios aspectos, ya que las variables climáticas pueden influir significativamente en sectores clave como la agricultura, la producción de energía y el consumo de bienes y servicios. Estas influencias pueden, a su vez, afectar variables económicas como la inflación y, por ende, impactar en índices de precios y cotizaciones.

Sin embargo, esto no implica que los cambios en la temperatura o la precipitación sean la única causa de los cambios en las variables económicas. Otros factores también pueden estar influyendo.
- c) Graficamos las predicciones de los últimos 24 meses con respecto a los valores reales, para observar qué método se ajusta mejor.

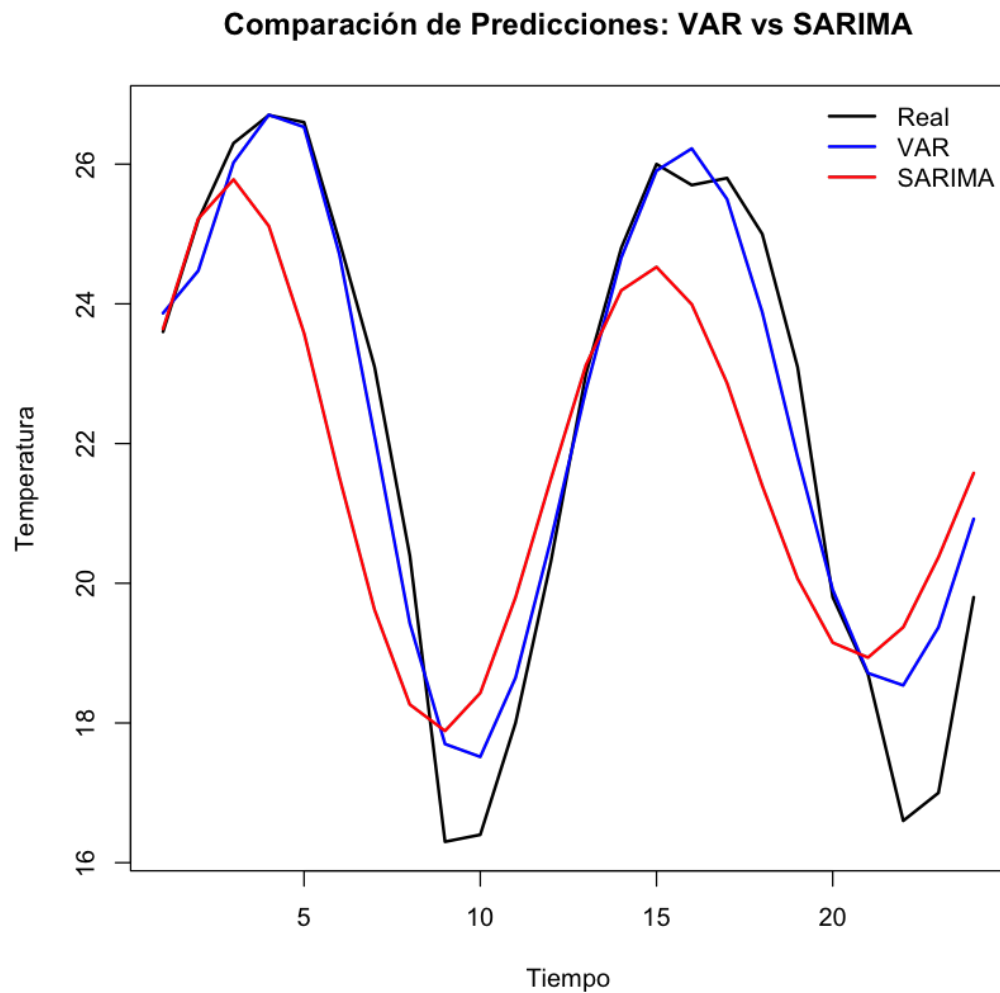


Figura 2: Muestra de los últimos 24 meses y predicciones usando un modelo $\text{VAR}(p)$ y $\text{SARIMA}(p,d,q)$

Por lo que podemos concluir que el modelo $\text{VAR}(p)$ nos da una mejor predicción de los valores que el modelo $\text{SARIMA}(p, d, q)$, pues los datos predichos se acercan más a los valores reales de la muestra.