



Universidad del
Rosario

Analisis Avanzado de Datos

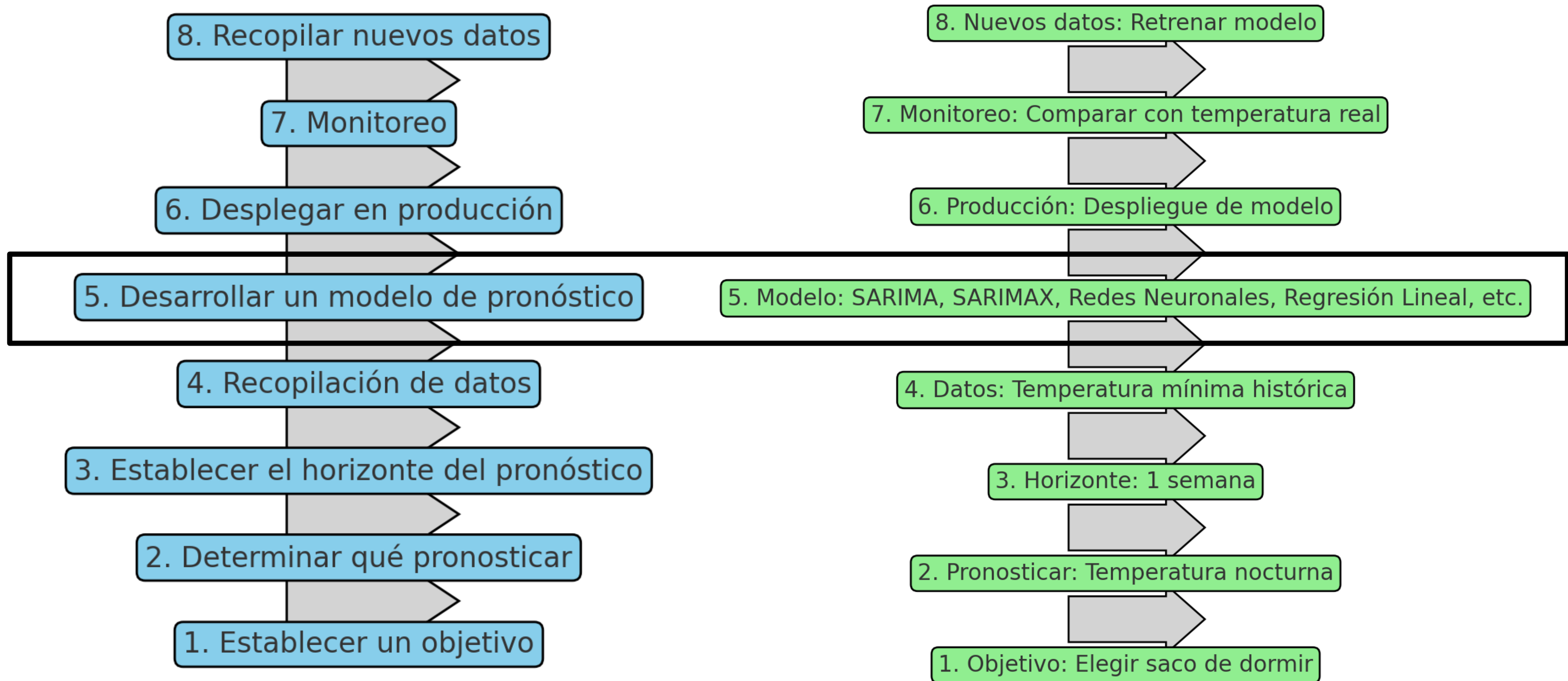
W15. major modelo series de tiempo

FERNEY ALBERTO BELTRAN MOLINA

Escuela de Ingeniería, Ciencia y Tecnología

Matemáticas Aplicadas y Ciencias de la Computación

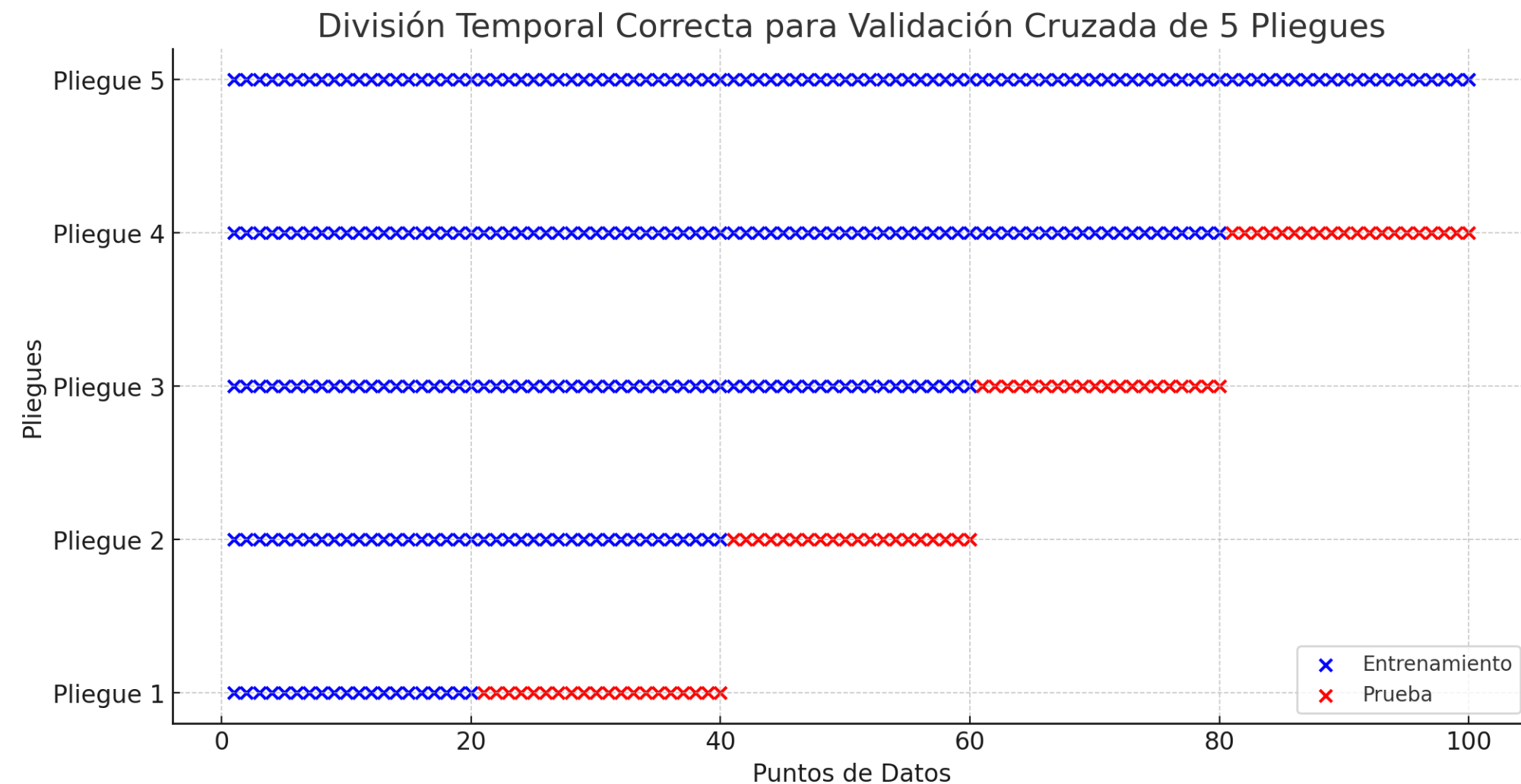
Pasos para pronosticas series de tiempo



**cómo elegiremos el mejor modelo de todos los
modelos que ajustaremos.**

Validación del modelo

La validación cruzada para series temporales es un poco diferente de la validación cruzada que se utiliza en otros tipos de modelos estadísticos o de aprendizaje automático. En las series temporales, no podemos dividir los datos de forma aleatoria porque las observaciones están ordenadas en el tiempo y la estructura temporal puede ser importante.



Criterio de Información de Akaike (AIC):

El AIC es un estimador de la calidad relativa de los modelos estadísticos para un conjunto de datos dado. Se basa en dos principios: **la bondad del ajuste del modelo** y **la simplicidad (o parsimonia)** del modelo. Un modelo con menor AIC no solo se ajusta bien a los datos, sino que también evita ser innecesariamente complejo.

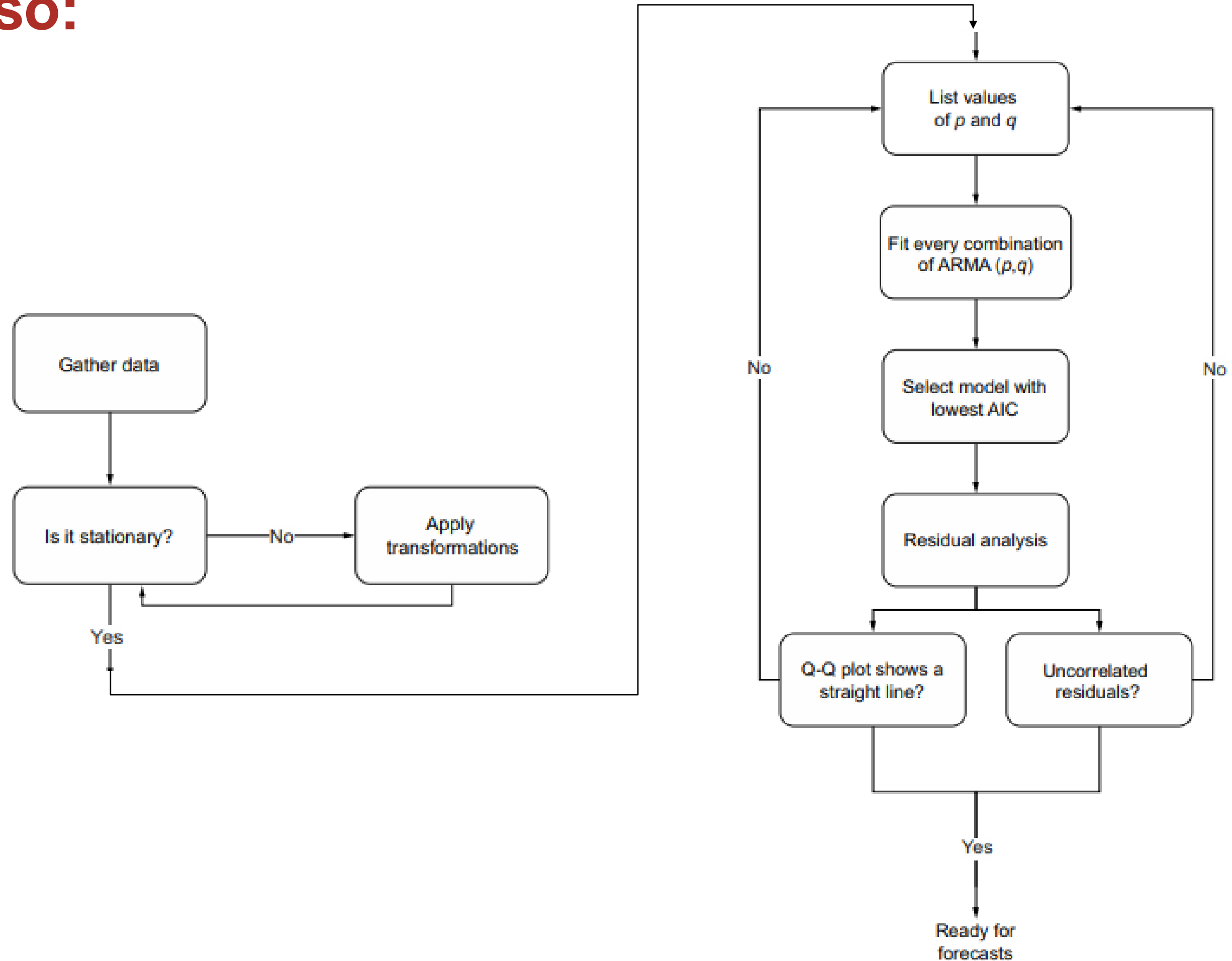
$$AIC = 2k - 2 \ln(L)$$

Donde **k** es el número de parámetros estimados en el modelo, y **ln(L)** es el logaritmo natural del máximo valor de la función de verosimilitud para el modelo.

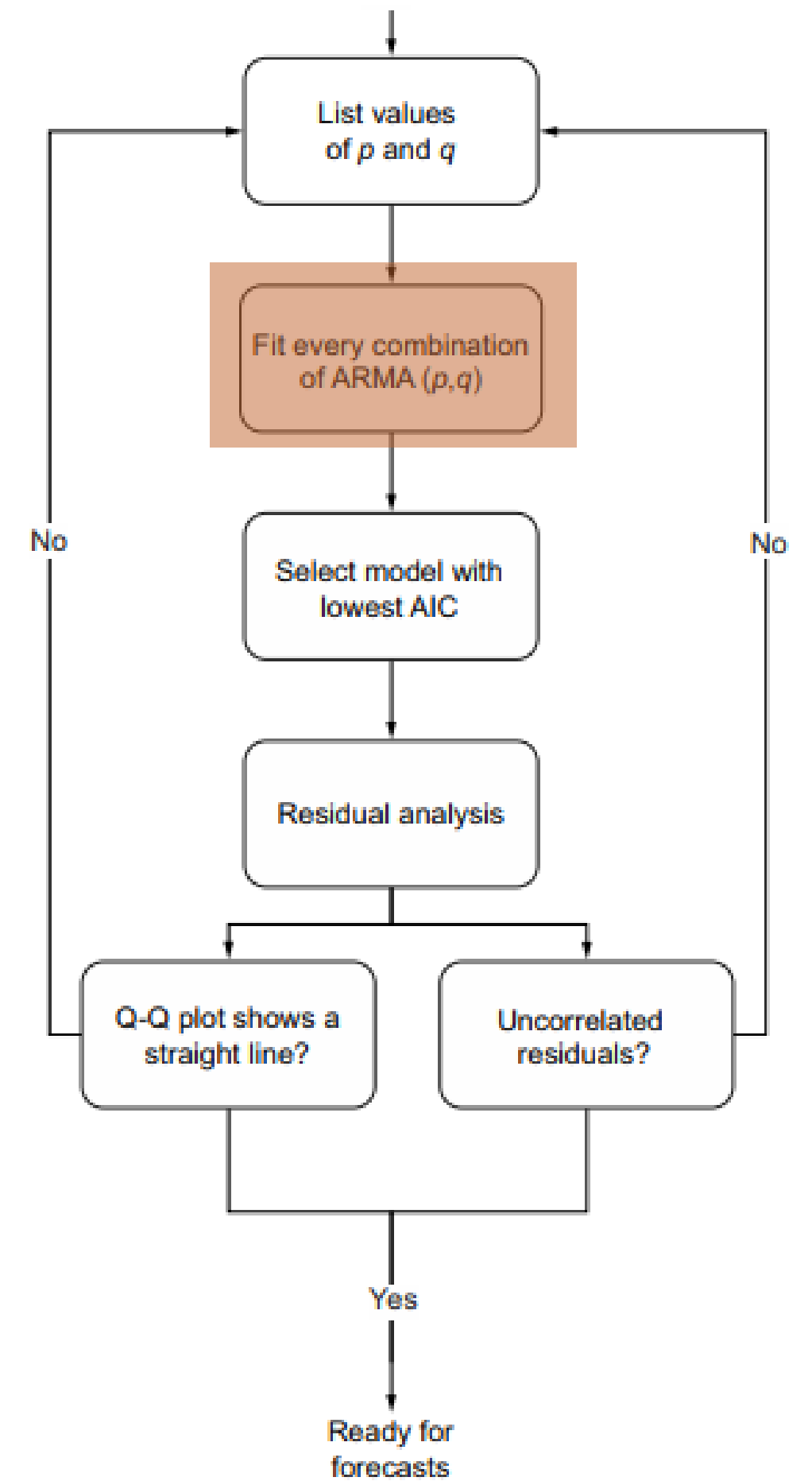
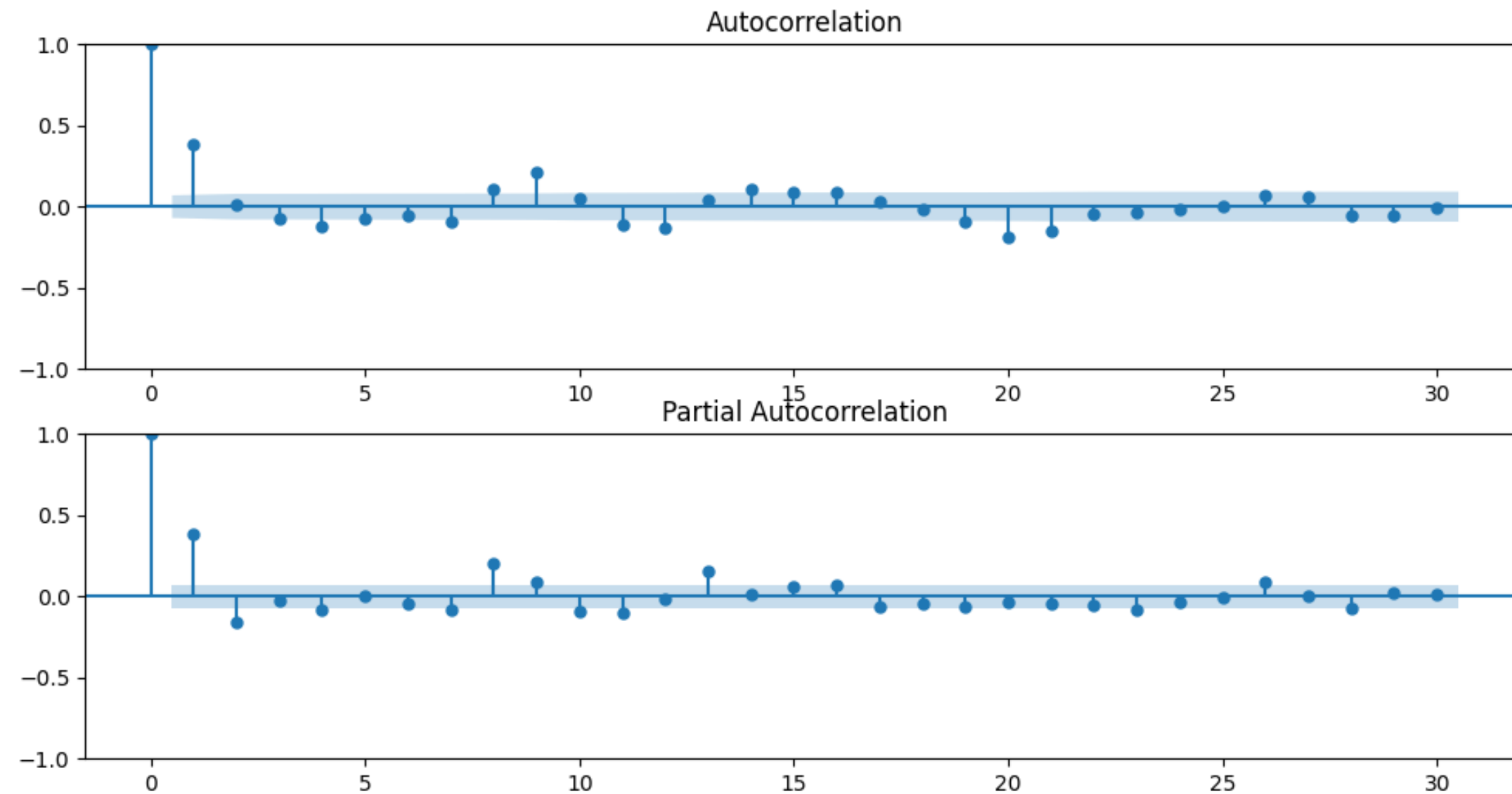
Cuando comparamos modelos, el modelo con el menor valor de AIC es generalmente el preferido.

El AIC permite comparar modelos que no son anidados, es decir, modelos que pueden tener diferentes formas o variables.

Proceso:



Proceso:



Proceso:

$$AIC = 2k - 2 \ln(L)$$

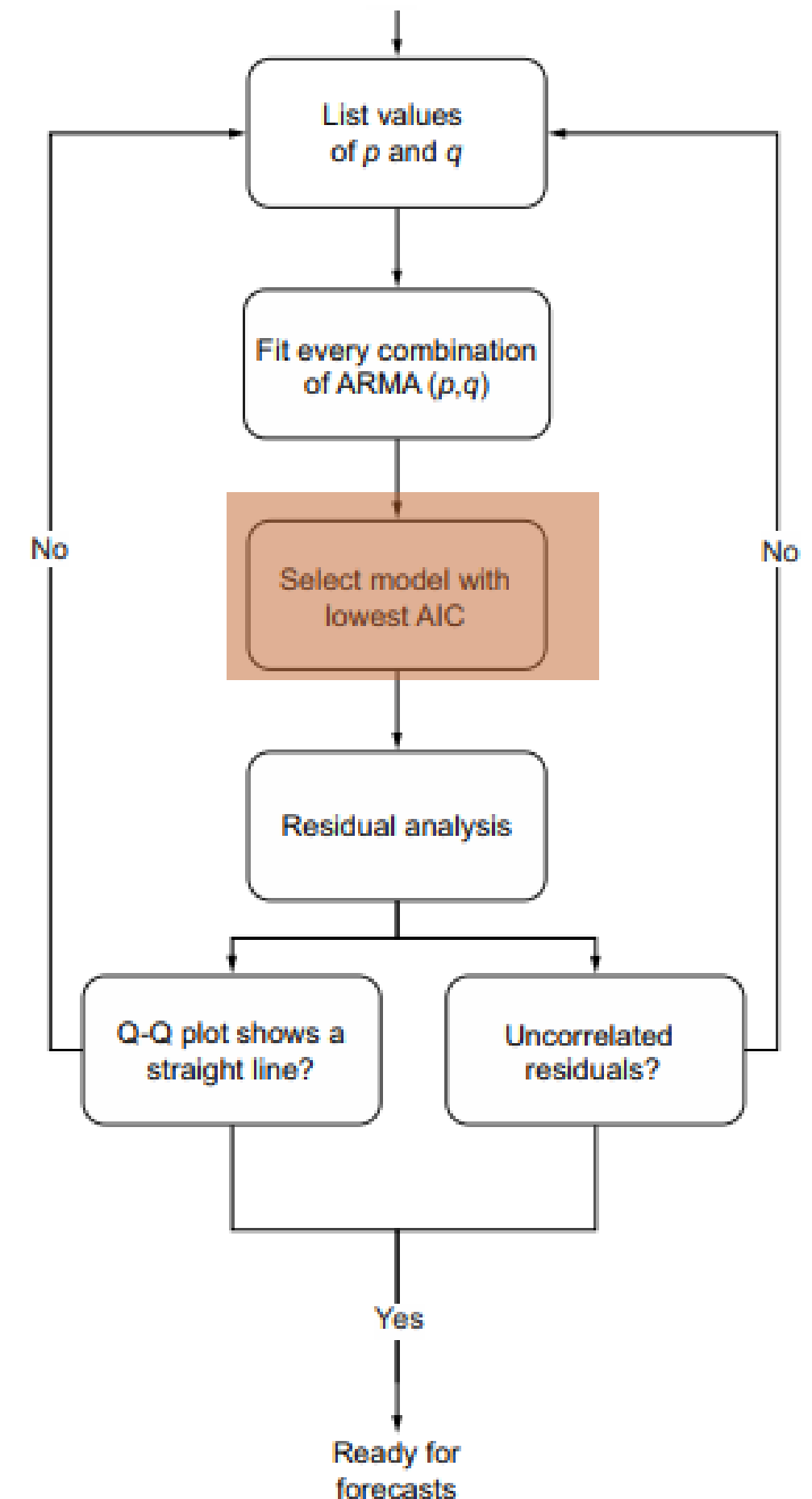
El AIC cuantifica la cantidad relativa de información perdida por el modelo. Cuanta menos información se pierda, menor será el valor del AIC y mejor será el modelo.

El número de parámetros estimados k está directamente relacionado con el orden (p,q) de un modelo ARIMA(p,d,q). Para modelo más complejo puede penalizar la puntuación del AIC: a medida que el orden (p,q) aumenta, el número de parámetros k aumenta,

La función de **verosimilitud** mide la bondad de ajuste de un modelo. Se puede ver como lo opuesto a la función de distribución

¿Qué tan probable es que mis datos observados provengan de un modelo ARMA(p,q)?

Un modelo con un valor más alto de la función de verosimilitud (**indicando un mejor ajuste**) y menos parámetros (**indicando simplicidad**) tendrá un valor más bajo de AIC y será generalmente preferido.



La verisimilitud

Estimar los parámetros del modelo de manera que maximicen la probabilidad de observar los datos dados. La fórmula exacta de la función de verosimilitud puede variar dependiendo de la estructura

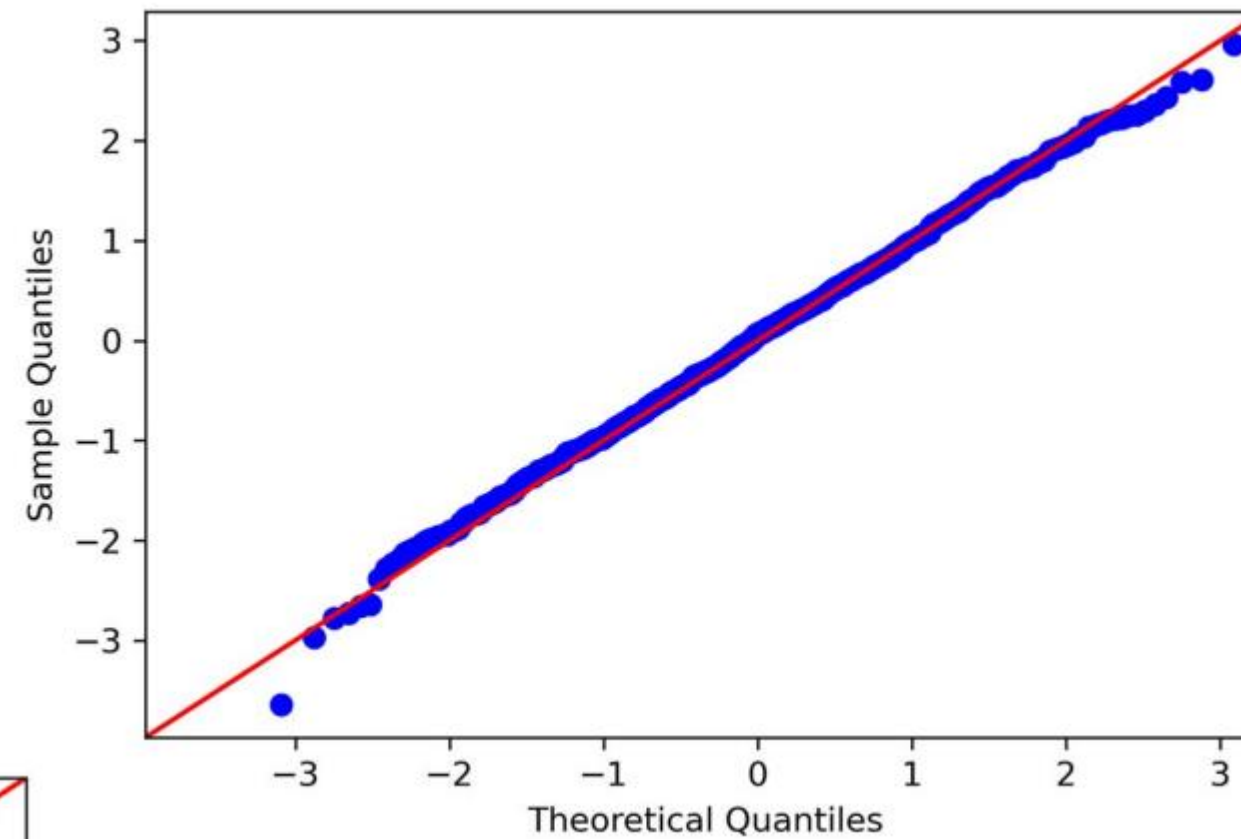
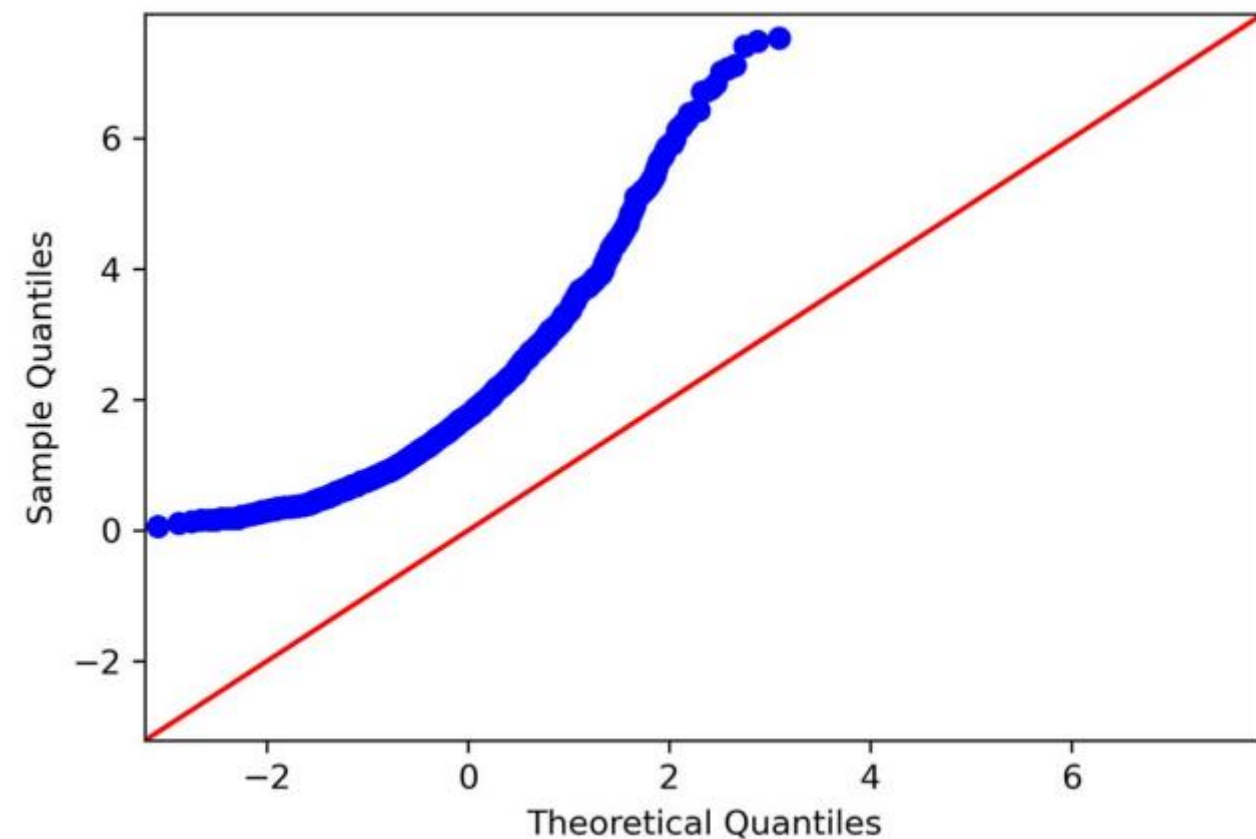
Se hace la suposición de que los residuos o errores del modelo siguen una distribución normal. En términos de probabilidades, la función de verosimilitud se expresa como el producto de las probabilidades de cada observación individual, dada la estructura del modelo.

$$L(\theta|y) = \prod_{i=1}^n P(y_i|\theta)$$

- θ son los parámetros del modelo. Para ARIMA sería p y q
- $P(y_i|\theta)$ es la probabilidad de observar y_i dado el modelo con parámetros θ .

Proceso:

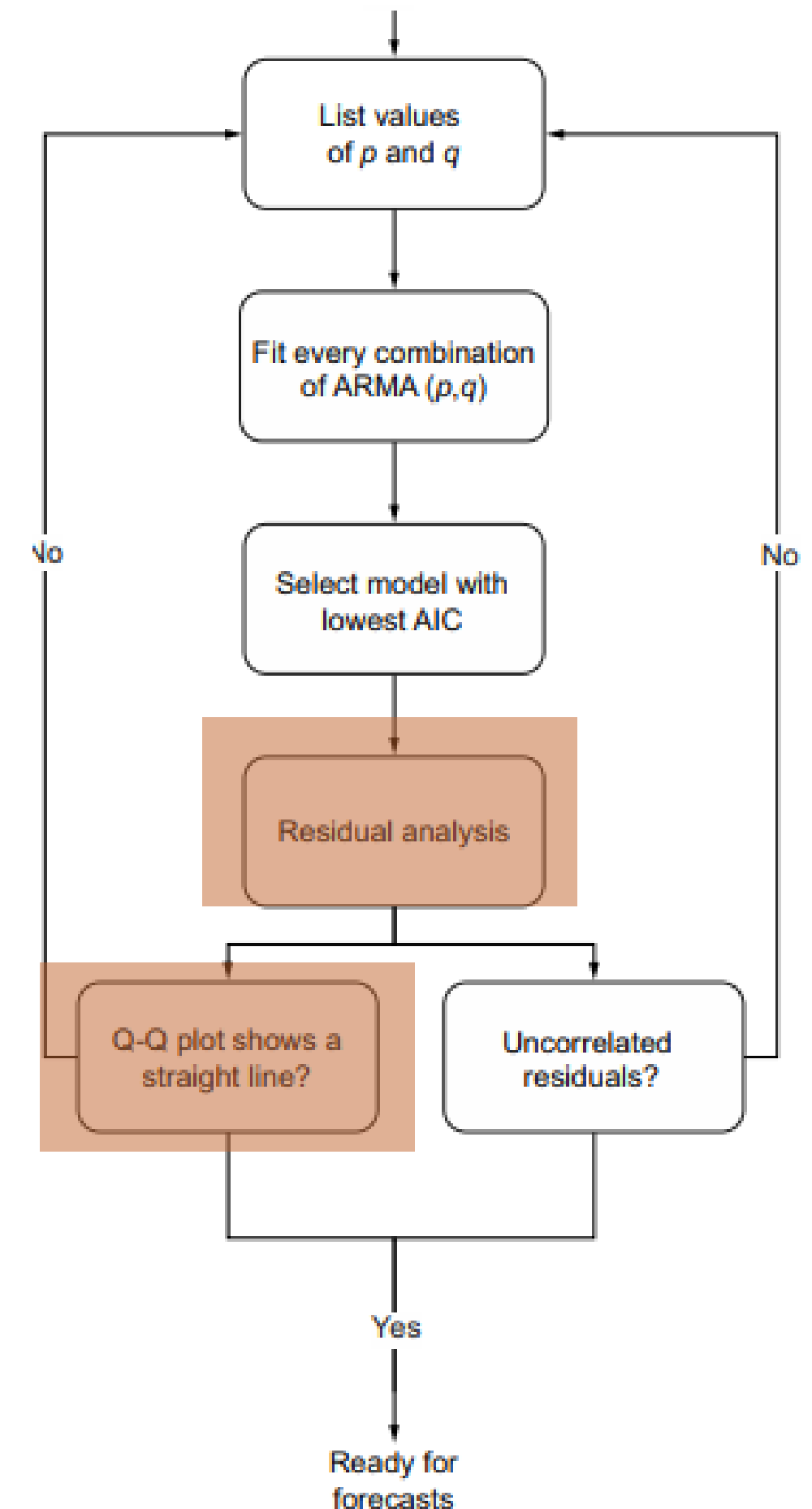
El gráfico de cuantiles-cuantiles (Q-Q) de residuos evaluar si una serie de residuos sigue una distribución normal.



Alineación con la Línea Recta: sugiere que los residuos tienen una distribución normal.

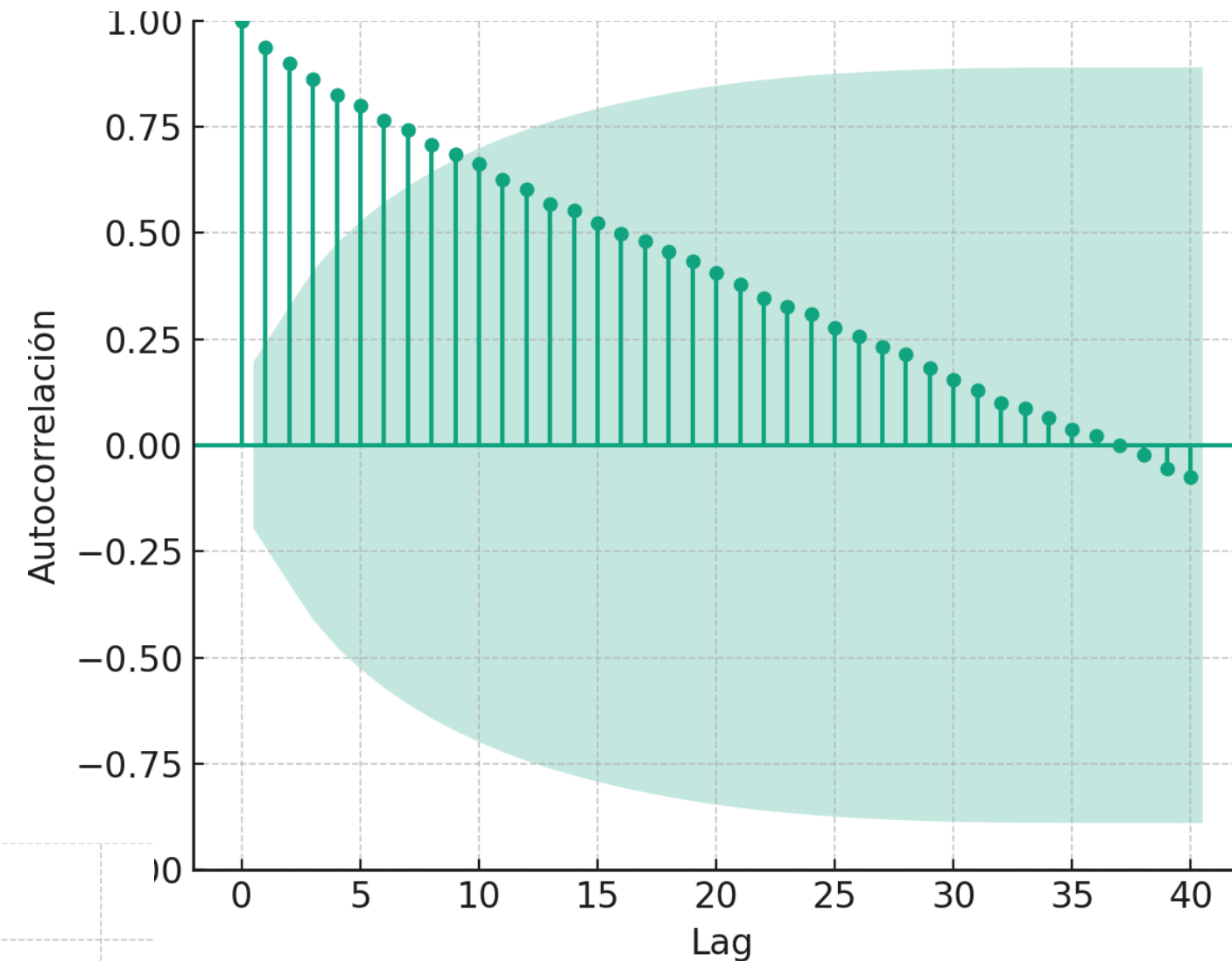
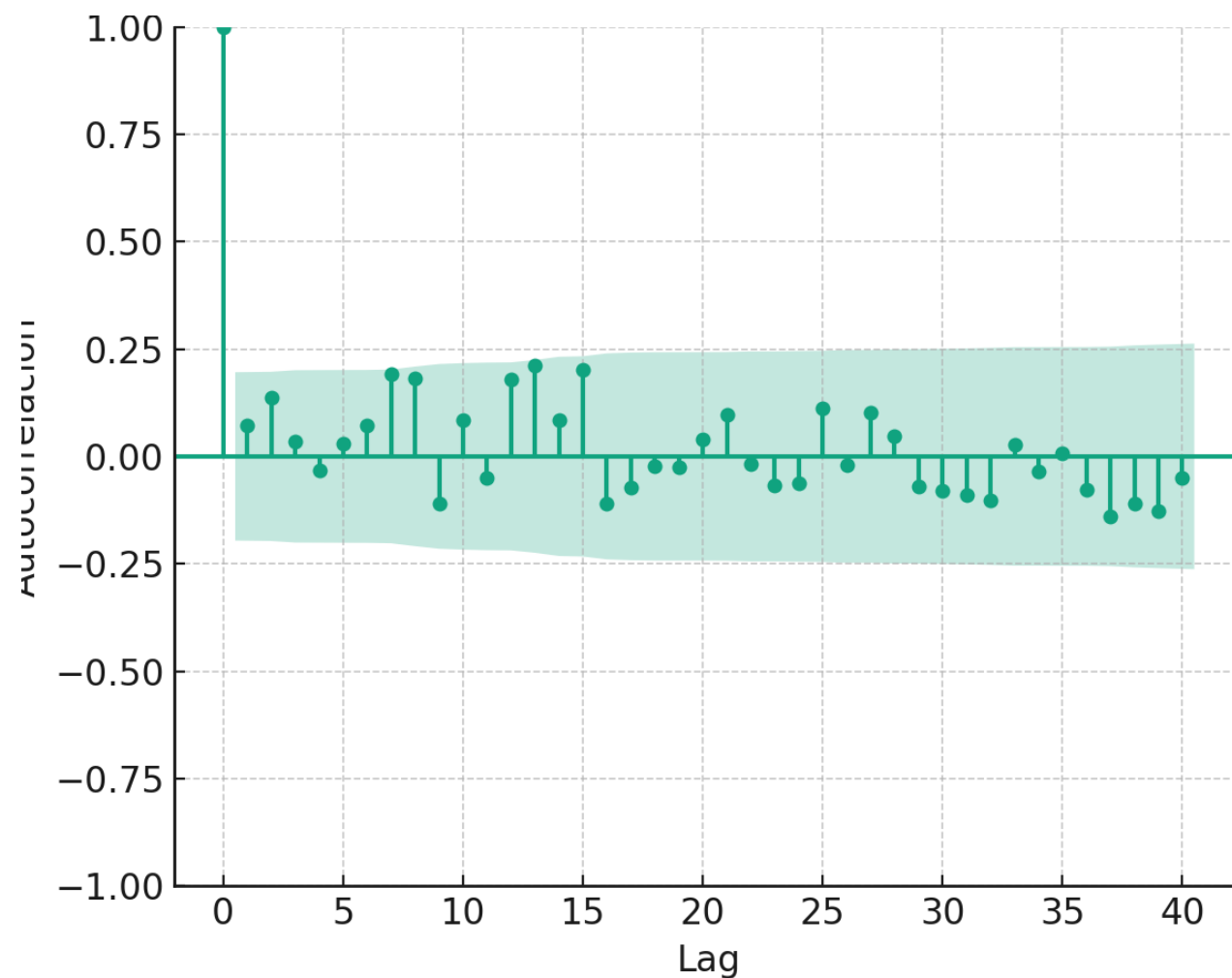
Desviaciones de la Línea: Si los puntos se desvían significativamente de la línea recta, indica que los residuos no siguen una distribución normal.

Colas Pesadas o Ligeras: Por ejemplo, si los puntos forman una curva "S" que se desvía de la línea en los extremos, puede indicar colas pesadas o ligeras en la distribución de residuos

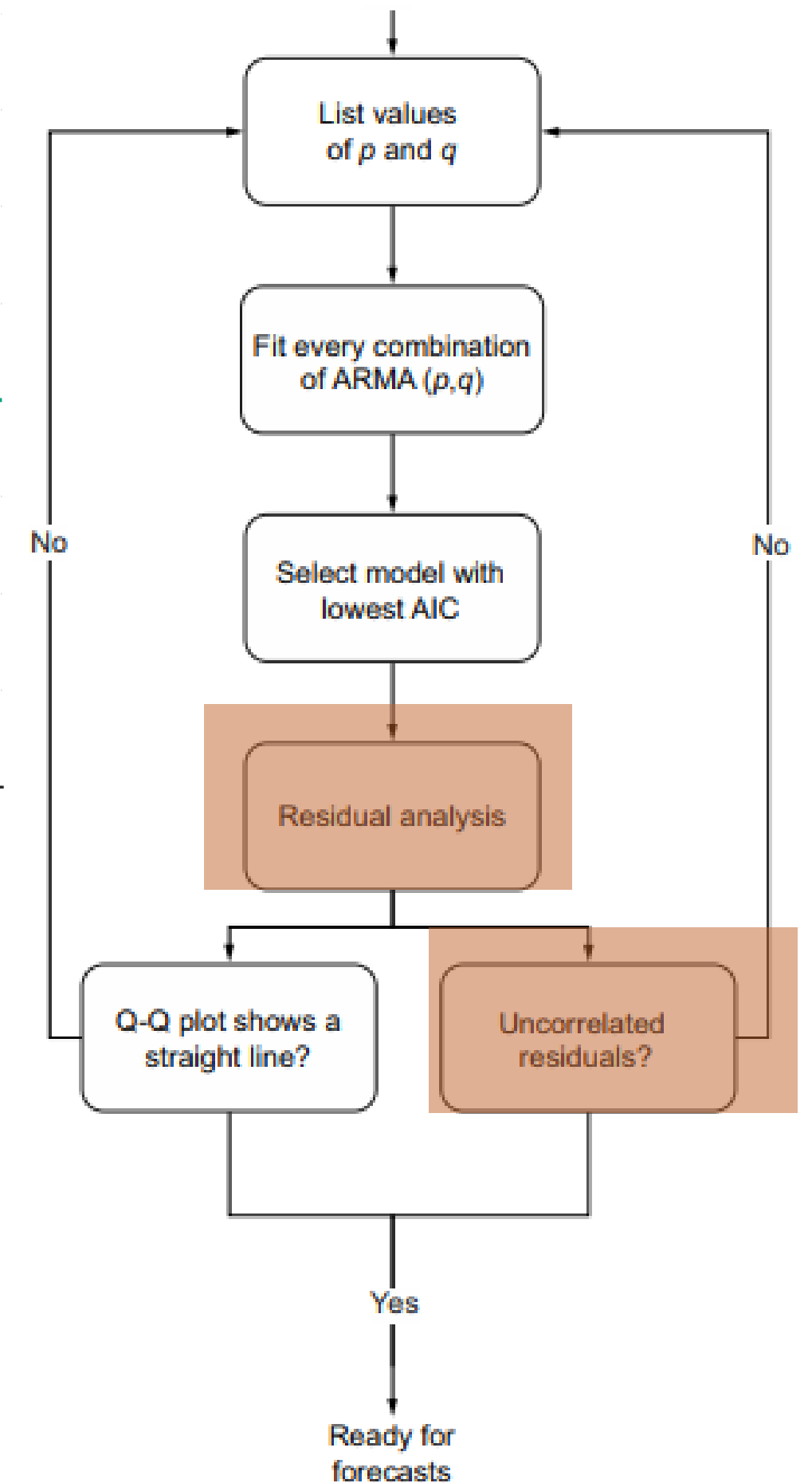


Proceso:

Indica una autocorrelación significativa en los residuos, lo que sugiere que los residuos están influenciados por sus valores pasados



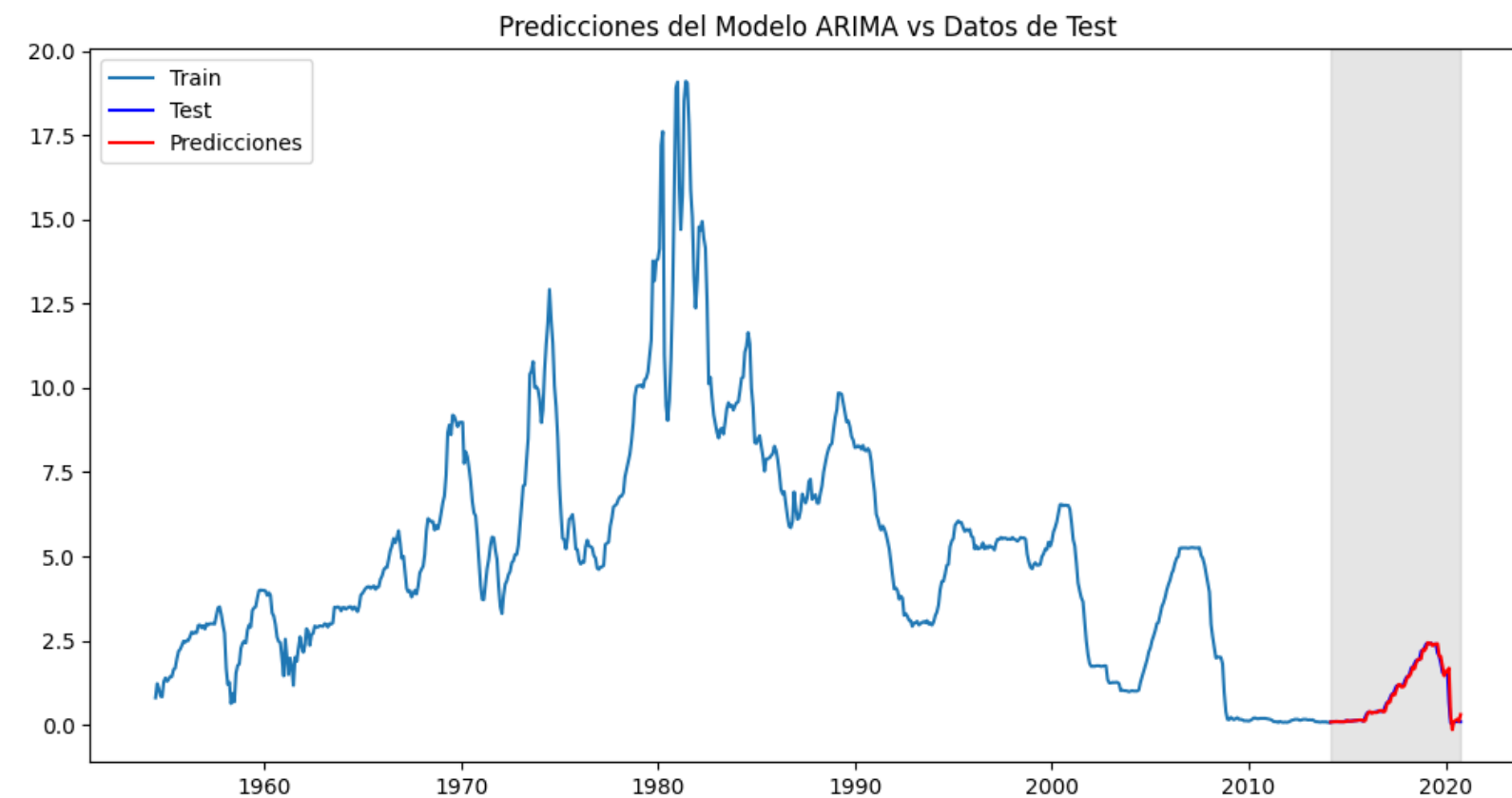
se mantienen dentro de los intervalos de confianza, lo que sugiere que cualquier autocorrelación observada es probablemente debido al azar y no es significativa. Esto indica que los residuos son independientes



Predicciones Fuera de la Muestra

Los modelos ARIMA dependen en gran medida de los datos históricos. Cuando haces predicciones más allá de tus datos actuales, el modelo utiliza sus propias predicciones como "datos históricos". Esto puede aumentar la incertidumbre o el error en las predicciones a medida que te alejas más del último punto de datos conocido

Actualización de Modelos: Es una buena práctica actualizar periódicamente tus modelos con nuevos datos para mantener su relevancia y precisión..



Incertidumbre Aumentada: A medida que el horizonte de predicción se extiende, la incertidumbre asociada con las predicciones tiende a aumentar

Ejercicio

La presencia de un patrón estacional, con picos en los meses de verano (junio, julio y agosto) y una disminución al principio y al final del año, sugiere la necesidad de un modelo que no solo capture las tendencias y autocorrelaciones, sino también la estacionalidad de los datos. ¿Qué hacemos?

