Modelos Arima

Metodología econométrica basada en modelos dinámicos que utiliza datos de series temporales.

Concepto

El modelo Arima es una metodología econométrica basada en modelos dinámicos que utiliza datos de series temporales. La metodología utilizada en los modelos ARIMA fue descrita inicialmente por el estadístico George Edward Pelham Box y el estadístico e ingeniero Gwilym Meirion Jenkins en 1970 en su libro: Análisis de series temporales. Predicción y control *(Time Series Análisis: Forecasting and Control).*

ÍNDICE

Consideraciones generales

Para trabajar con modelos ARIMA es necesario tener en cuenta una serie de conceptos básicos tales como: proceso estocástico, ruido blanco, sendero aleatorio y estacionariedad.

Un proceso estocástico es una sucesión de variables aleatorias ($\{Y_t\}$, t = -∞,...-1, 0, 1,..., ∞) que dependen de un parámetro, en el caso de las series temporales este parámetro es el tiempo.

Un ruido blanco es una sucesión de variables aleatorias que se caracterizan por tener una esperanza constante e igual a cero, igual varianza, y además, son independientes a lo largo del tiempo (covarianza es cero).

Un sendero aleatorio es un proceso estocástico que se caracteriza porque su primera diferencia es un ruido blanco.

Un proceso estocástico es débilmente estacionario o estacionario en un sentido amplio, si se cumple que su media y su varianza son constantes para cualquier período de tiempo y las covarianzas entre dos variables solo dependen del lapso de tiempo que transcurre entre ellas.

Etapas en la elaboración de un modelo arima

Identificación

Para identificar cual es el proceso ARIMA que ha generado una determinada serie temporal es necesario que los datos sean estacionarios, es decir, no pueden presentar tendencia creciente o decreciente (si presentan tendencia habría que diferenciar la serie porque la serie no es estacionaria en media), ni tampoco pueden presentar fluctuaciones de diferente amplitud. Si la dispersión no se mantiene constante entonces la serie no es estacionaria en varianza y habría que transformarla siendo, la transformación logarítmica la más habitual.

Una vez que la serie es estacionaria es necesario obtener las funciones de autocorrelación simple y parcial muestrales para determinar el proceso ARIMA(p,d,q) más adecuado que haya podido generar la serie estacionaria.

En los modelos ARIMA(p,d,q), p representa el orden del proceso autorregresivo, d el número de diferencias que son necesarias para que el proceso sea estacionario y q representa el orden del proceso de medias móviles.

Para identificar un proceso autorregresivo de orden p, es necesario que la función de autocorrelación simple no se anule y decrezca de forma exponencial o sinusoidal hacia cero, mientras que, la función de autocorrelación parcial tenga solo p coeficientes distintos de cero. La expresión general de un AR(p) es una combinación lineal de p valores pasados de la variable y un ruido blanco actual,

$Y_t = \varphi_1 Y_{t-1} + \ldots + \varphi_p Y_{t-p} + a_t$ o bien $(1 - \varphi_1 L - \ldots - \varphi_p L^p) Y_t = a_t$

Escrito de forma compacta: $\varphi_t (L) Y_t = a_t$

Donde a_t es un ruido blanco y L es el operador de retardos. Los procesos autorregresivos serán siempre invertibles y para que sean estacionarios es necesario que las p raíces del polinomio característico estén fuera del círculo unidad.

Si el proceso es de medias móviles de orden q, el comportamiento de las funciones de autocorrelación simple y parcial será el contrario de los procesos autorregresivos, es decir, la función de autocorrelación simple tendrá q coeficientes distintos de cero y la función de autocorrelación parcial no se anula y decrece de forma exponencial o sinusoidal hacia cero. La expresión general de un MA(q) muestra como la variable Y_t es una combinación lineal de ruidos blancos,

$Y_t = a_t + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}$ o bien $(1 - \theta_1 L - \ldots - \theta_q L^q) a_t = Y_t$

Escrito en forma compacta: $\theta_q(L) a_t = Y_t$

Donde a_t es ruido blanco. Los procesos de medias móviles son siempre estacionarios y para que sea invertible es necesario que las q raíces del polinomio estén fuera del círculo unidad.

Sin embargo, si el proceso es un ARMA, ni la función de autocorrelación simple y ni la parcial se anulan y su identificación es más complicada que la de un AR o un MA. Así, los procesos ARMA o modelos mixtos se obtienen de combinar los procesos AR y MA. La expresión general de un ARMA(p,q) es,

$Y_t = \varphi_1 Y_{t-1} + \ldots + \varphi_p Y_{t-p} + a_t + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}$

O bien $(1 - \varphi_1 L - \ldots - \varphi_p L^p) Y_t = (1 - \theta_1 L - \ldots - \theta_q L^q) a_t$

Escrito en forma compacta: $\varphi_p (L) Y_t = \theta_q (L) a_t$

Donde a_t es ruido blanco. Los procesos ARMA(p,q) son estacionarios si las p raíces del polinomio de la parte autorregresiva están fuera del círculo unidad y será invertible si las q raíces del polinomio de parte de las medias móviles están también fuera del círculo unidad.

Estimación

Después de identificar el proceso que ha generado los datos de una determinada serie temporal es necesario estimar los parámetros de los que depende. Si se supone que de forma genérica que el proceso es estacionaria es invertible y que la serie temporal ha sido generada por un ARIMA(p,d,q), entonces, la transformación estacionaria $W_t = (1-L)^d Y_t$ será un ARMA(p,q):

$W_t = \varphi_1 W_{t-1} + \ldots + \varphi_p W_{t-p} + a_t + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}$

donde los parámetros que hay que estimar son: $\varphi_1, \ldots, \varphi_p, \theta_1, \ldots, \theta_p$.

Estos parámetros del modelo ARMA se pueden estimar por máxima verosimilitud asumiendo que la distribución la serie, por ejemplo, es Normal. En este caso el logaritmo de la función de verosimilitud es,

$$\text{Log } L = -\frac{T}{2}\log(2\pi) - \frac{(T-1)}{2}\sigma_a^2 - \frac{1}{2\sigma_a^2}\sum_{t=2}^{T}\left[y_t - E\left[{}^{w_t}/_{W_{t-1}}\right]\right]^2 - \frac{1}{2}\log(\sigma^2) - \frac{(w_1-\mu)^2}{2\sigma^2}$$

donde T es el número de observaciones; σ^2_a es la varianza del ruido; μ es la media de la serie y σ^2 su varianza.

Validación

Después de estimar el modelo es necesario comprobar si se ajusta o no de forma adecuada a los datos observados de la serie temporal objeto de estudio. Para comprobar si el modelo es adecuado se suele realizar lo siguiente: análisis de los parámetros estimados, análisis de los residuos, análisis de la bondad del ajuste y análisis de estabilidad.

a) Análisis de los parámetros estimados. Es necesario contrastar si los parámetros estimados son estadísticamente significativos o no. Para ello la hipótesis nula, para cada uno de los coeficientes, plantea que este es cero frente a la hipótesis alternativa que es distinto de cero. El estadístico del contraste se basa en la distribución t de Student. Si al realizar el contraste alguno de los parámetros no fuese significativo sería necesario eliminarlo del modelo.

(https
2021%

También es necesario comprobar que los parámetros estimados cumplen las condiciones de estacionariedad e invertibilidad. Así, si alguna de las raíces del polinomio de retardos correspondiente a la parte autorregresiva está próxima a uno, es posible que la serie original esté subdiferenciada (sería necesario calcular alguna diferencia más). Sin embargo, si alguna de las raíces del polinomio de retardos correspondiente a las medias móviles está próxima a uno, entonces es posible que el modelo esté sobrediferenciado (sería necesario eliminar alguna diferencia). Por otro lado, si existen raíces comunes en ambos polinomios de retardos, entonces estas se simplificarían y el modelo adecuado sería un ARMA(p-1,q-1).

b) Análisis de los residuos. Uno de los supuestos del modelo ARIMA es que las perturbaciones aleatorias son ruido blanco, sin embargo, como estas son inobservables es necesario calcular los residuos y comprobar si estos son ruido blanco o no. Existen varias formas de comprobar si los residuos son ruido blanco o no, entre ellas se puede destacar: el gráfico de los residuos, la función de autocorrelación simple y parcial estimada de los residuos y el contraste de Portmanteau.

En primer lugar, el gráfico de los residuos nos puede indicar de forma intuitiva si su media es constante e igual a cero, y también, si su varianza es constante. Además, también puede indicar si se producen errores sistemáticos o no, así como, si existen valores atípicos (aquellos que excedan tres veces su desviación típica). Si existiesen valores atípicos sería necesario llevar a cabo un análisis de intervención adecuado que recoja la información proporcionada por dichos valores.

En segundo lugar, se analizarían las funciones de autocorrelación simple (FAC) y parcial (FACP) muestrales de los residuos. De tal forma, que si el modelo está estimado de forma correcta, entonces, todos los coeficientes estimados de la FAC y FACP tienen que ser estadísticamente nulos. Si, por el contrario, la FAC y FACP indicasen que los residuos no son ruido blanco entonces, será necesario identificar el proceso e incorporar esta información en el modelo inicial propuesto con el fin de volver a estimar otra vez el modelo.

Por último, el contraste de Portmanteau o contraste global de significación se utiliza para comprobar si hay ausencia de autocorrelación entre los términos de error, es decir,

H_0: $\rho_1 = \rho_2 = \ldots = \rho_m = 0$

H_1: algún $\rho_k \neq 0$, k = 1,..., m

donde ρ_k es el coeficiente de orden k de la FAC del proceso de ruido blanco. El estadístico Q(m) de Box-Pierce (1970) o el actualizado por Ljung-Box(1976) para realizar este contraste se distribuyen como una χ^2 con (m-p-q) grados de libertad. Si para un nivel de significación α se acepta la hipótesis nula los residuos son ruido blanco, y si se rechaza habrá que volver a estimar el modelo.

c) Análisis de la bondad del ajuste. Para ver la adecuación que existe entre la serie real y la estimada se puede utilizar el coeficiente de determinación o el coeficiente de determinación corregido. Cuanto más próximos estén estos coeficientes a uno mejor será el modelo ajustado.

Sin embargo, si se quieren comparar diferentes modelos, estos coeficientes solo se podrán utilizar cuando se hayan calculado el mismo número de diferencias. En este caso, habrá que utilizar otros métodos alternativos de comparación, tales como, el criterio de información de Akaike (AIC) o el criterio de información de Bayes (BIC). En ambos casos, se elegirá como más adecuado aquel modelo que tenga un menor valor del AIC o del BIC.

d) Análisis de estabilidad. Este análisis, permite establecer si el modelo ARIMA estimado, para el período muestral, es estable también en períodos futuros. Si esto es así se podrán obtener predicciones fiables de la variable objeto de estudio. Para realizar este análisis de estabilidad se puede utilizar el test de estabilidad estructural de Chow.

Predicción

Una vez estimado y validado el modelo ARIMA se puede utilizar para obtener valores futuros de la variable objeto de estudio. Las predicciones obtenidas pueden ser de dos tipos: puntuales o por intervalos. La predicción puntual se obtiene calculando el valor esperado de la variable en el período futuro T+l condicionado al conjunto de información disponible hasta el período T. Y la predicción por intervalos, para un nivel de confianza del 95 %, se obtiene sumando y restando a la predicción puntual la desviación típica del error de predicción multiplicada por el valor tabulado para el 95 % de confianza.

Las características genéricas de las predicciones realizadas con modelos ARIMA son:

a) En los modelos AR(p). La predicción a medida que aumenta el horizonte temporal, tiende a la media del proceso.

b) En los modelos MA(q). Si el horizonte temporal de la predicción es mayor que el orden del proceso entonces, la predicción es igual a la media.

c) En los modelos ARMA(p,q). Para períodos superiores al orden del proceso de las medias móviles, la función de predicción se comporta como la de un proceso AR(p) y, por lo tanto, tiende a la media.

d) En los modelos ARIMA(p,d,q). La predicción ya no tiende a la media, sino que será una línea recta con pendiente igual a la media del proceso que se obtiene al hacer las transformaciones necesarias para que la serie sea estacionaria.

Recuerde que...

- Para trabajar con modelos Arima es necesario tener en cuenta una serie de conceptos básicos tales como: proceso estocástico, ruido blanco, sendero aleatorio y estacionariedad.
- Las etapas en la elaboración de un modelo Arima son: identificación, estimación, validación y predicción.
- Para identificar cual es el proceso Arima que ha generado una determinada serie temporal es necesario que los datos sean estacionarios, es decir, no pueden presentar tendencia creciente o decreciente, ni tampoco pueden presentar fluctuaciones de diferente amplitud.
- Después de identificar el proceso que ha generado los datos de una determinada serie temporal es necesario estimar los parámetros de los que depende. Para comprobar si el modelo es adecuado se suele realizar lo siguiente: análisis de los parámetros estimados, análisis de los residuos, análisis de la bondad del ajuste y análisis de estabilidad.
- Una vez estimado y validado el modelo Arima se puede utilizar para obtener valores futuros de la variable objeto de estudio. Las predicciones obtenidas pueden ser de dos tipos: puntuales o por intervalos.

(https
2021%