MODELOS ARIMA

Nicolás Chávez Quisbert

Nicolás Chávez Quisbert tiene el título de licenciado en Estadística, obtenido en la Universidad Mayor de San Andrés; realizó estudios de Postgrado en Demografía en el CIDES-UMSA. Catedrático en la Universidad Mayor de San Andrés y en la Escuela Militar de Ingeniería (EMI). Actualmente es profesor en la Carrera de Ingeniería de Sistemas de la Universidad Católica Boliviana.

INTRODUCCION

El problema de la predicción en cualquier campo es una tarea compleja que exige en algunas ocasiones la cuidadosa utilización de las técnicas estadístico econométricas muy complicadas.

La recolección de información, en particular de tipo económicas ya sea de personas, empresas o países, se la realiza con fines de análisis para posteriormente llevar a cabo la planeación y la toma de decisiones. El proceso de toma de decisiones lleva aparejada una cierta evaluación del futuro, que se pueden resumir en seis etapas.

1. Planeamiento del problema.
2. Valoración de la situación actual.
3. Consideración de las alternativas de futuros.
4. Formulación de objetivos.
5. Elección de políticas.
6. Control de realización del proceso.

Es lógico suponer que no se pueden tomar decisiones de políticas de un determinado proyecto, sin considerar la evaluación futura de todos aquellos elementos que lo condicionan. Por lo tanto la consideración del futuro en cualquier campo es ineludible para la toma de decisiones.

Por otra parte las predicciones implicadas no tienen que ser únicas ni permanentes, además no solo debe considerarse en general la existencia de predicciones alternativas, sino la necesidad de una revisión permanente de las predicciones.

La predicción es, en si misma una información referida al futuro de suma importancia, la cual requiere un análisis minucioso. Estos elementos son los que me incentivaron a realizar el estudio de series de tiempo con Modelos ARIMA "Autorregresivos Integrados de Medias Móviles", basados en la teoría de procesos estocásticos.

OBJETIVO

Es mostrar el enfoque complejo del análisis de Series de Tiempo univariantes sugerido por Box y Jenkins, que en la actualidad es de gran utilidad en el campo estadístico para determinaciones de proyecciones, de un sin fin de variables de tipo económico social demográfico y de diferentes campos, los cuales son susceptibles a cambios en función al tiempo, a su vez detectar los efectos de dichos cambios.

Ya que la mayor parte de los procedimientos univariados para pronosticar son totalmente automáticos y estáticos, en el sentido que solamente un conjunto de pronósticos pueden obtenerse de una serie de datos, y estos a su vez no son cambiantes. Sin embargo el procedimiento de Box y Jenkins envuelve un elemento subjetivo que le permite a uno escoger entre una gran variedad de modelos, y realizar un control de calidad de las proyecciones en forma sencilla. Esa gran versatilidad es a su vez la fuerza del enfoque de Box y Jenkins, ya que como se mencionó anteriormente, las predicciones no tienen que ser únicas ni permanentes, esto facilita en lo posible toma de decisiones y políticas a adoptar.

La ventaja de poder escoger entre una gran familia de modelos en vez de estar restringidos a un modelo es clara.

Procesos Autorregresivos y de Medias Móviles (ARMA).-

Definición.- A la combinación formada por procesos autorregresivos y de medias móviles se conoce como proceso ARMA. Llamado también proceso mixto, si este contiene p términos autorregresivos y q términos de medias móviles, se dice que es de orden ARMA (p,q) el cual se lo define de la siguiente manera.

$$\phi(B)\ \overline{Z}_t = \theta(B)\ a_t \qquad \text{con} \quad \overline{Z}_t = \overline{Z}_t - \mu$$

donde

$\phi(B)$ y $\theta(B)$ son polinomios de retraso de orden p y q respectivamente, $\mathbf{a}_t$ es un proceso de ruido blanco, y μ es el nivel del proceso Z_t

Estacionariedad e Invertibilidad.

Como los modelos ARMA solo pueden ser aplicados a series que no muestren tendencia, es bueno considerar el problema de la estacionariedad e invertibilidad.

si

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \theta_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

son los polinomios de retraso, entonces

el proceso $\overline{Z}$ es estacionario, si las raíces de la ecuación característica

$$\phi(X) = 1 - \phi_1 X - \phi_2 X^2 - \dots - \phi_p X^p$$

se encuentra fuera del círculo unitario.

El proceso $\overline{Z}$ es invertible, si las raíces de la ecuación característica

$$\theta(X) = 1 - \theta_1 X - \theta_2 X^2 - \dots - \theta_q X^q$$

se encuentran fuera del círculo unitario.

La importancia del proceso ARMA reside en el hecho de que una serie de tiempo estacionaria puede ser a menudo descrita por un modelo ARMA que contenga menos parámetros que un proceso MA ó AR por si solo

Teorema.- En un modelo estacionario autorregresivo y de medias móviles ARMA (p,q) de la forma:

$$\overline{Z}_t = \phi_1\ \overline{Z}_{t-1} + \phi_2\ \overline{Z}_{t-2} + \dots + \phi_p\ \overline{Z}_{t-p} + a_t - \theta_1 a_{t-2} - \dots - \theta_q a_{t-q}$$

con $\overline{Z}_t = Z_t - \mu$

donde μ es el nivel del proceso Z_t y a_t es un proceso de ruido blanco.

La esperanza matemática, la varianza, la función de autocovarianza y la función de autocorrelación son:

$E(\overline{Z}) = 0$

$$Y_0 = \phi_1 Y_1 + \phi_2 Y_2 + \ldots + \phi_p Y_p + \sigma_{a^2} - \theta_1 E(a_{t-1}\overline{Z}_t) - \ldots - \theta_q E(a_{t-q}\overline{Z}_t)$$

$$\begin{aligned} Y_k &= \phi_1 E(\overline{Z}_{t-1}\overline{Z}_{t-k}) - \phi_2 E(\overline{Z}_t\overline{Z}_{t-k}) + \ldots + \phi_p E(\overline{Z}_{t-p}\overline{Z}_{t-k}) \\ &\quad + E(a_t\overline{Z}_{t-k}) - \theta_1 E(a_{t-1}\overline{Z}_{t-k}) - \ldots - \theta_q E(a_{t-q}\overline{Z}_{t-q}) \text{ con } k \geq 0 \end{aligned}$$

Demostración:

Esperanza matemática

$E(\overline{Z}_t) = E(Z_t - \mu) = E(Z_t) - \mu = \mu - \mu = 0$

VARIANZA

$$\begin{aligned} Y_0 &= E(\overline{Z}^2) = E((\phi_1\overline{Z}_{t-1} + \theta_2\overline{Z}_{t-2} + \ldots + \phi_p\overline{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q})\overline{Z}_t) \\ &= \phi_1 Y_1 + \phi_2 Y_2 + \ldots + \phi_p Y_p + \sigma_a^2 - \theta_1 E(a_{t-1}\overline{Z}_t) - \theta_2 E(a_{t-2}\overline{Z}_t) - \ldots - \theta_q E(a_{t-q}\overline{Z}_t) \end{aligned}$$

donde

$E(a_{t-i}\overline{Z}_t) = E(a_{t-i-1}\overline{Z}_{t-1}) = \ldots$ para $i = 0, \pm 1, \pm 2, \ldots$

$E(a_{t-1}\overline{Z}_t) = \phi_1 E(\overline{Z}_{t-1}a_{t-1}) + \ldots + \theta_i E(\overline{Z}_{t-i}a_{t-1}) - \theta_i\sigma_a^2$

para $i = 1, 2, 3, \ldots, \max(p,q)$

si $q<p \Rightarrow \theta_{q+1} = \theta_{q+2} = \ldots = \theta_p = 0$

si $p<q \Rightarrow \phi_{q+1} = \phi_{q+2} = \ldots = \phi_p = 0$ y además $E(a_t Z_t) = \sigma_a^2$

$$\begin{aligned} Y_k &= E(\overline{Z}_t\overline{Z}_{t-k}) = E[(\phi_1\overline{Z}_{t-1} + \ldots + \phi_p\overline{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q})\overline{Z}_{t-q}] \\ &= \phi_1 E(\overline{Z}_{t-1}\overline{Z}_{t-k}) + \ldots + \phi_p E(\overline{Z}_{t-p}\overline{Z}_{t-k}) + E(a_t\overline{Z}_{t-k}) - \theta_1 E(a_{t-1}\overline{Z}_{t-k}) - \ldots \\ &\quad \ldots - \theta_q E(a_{t-q}\overline{Z}_{t-k}) \end{aligned}$$

donde

$E(a_{t-i}\overline{Z}_{t-k}) = 0$ si $k > i$ por ser independientes

si $k > q$ se tiene que

$$Y_k = \phi_1 Y_{k-1} + \phi_2 Y_{k-2} + \ldots + \phi_p Y_{k-p}$$

si $k < q \Rightarrow Y_k$ tendrá parámetros $\theta_k, \theta_{k+1}, \ldots, \theta_q$

Analizando la función de autocorrelación ρ_k

$$\rho_k = \frac{Y_k}{Y_0}$$

Si $k>q$ se tiene que:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p}$$

Si $k<q$ se tiene que:

ρ_k involucrará a los parámetros $\theta_k, \theta_{k+1}, \theta_{k+2} \ldots, \theta_q$

Entonces podemos concluir que si $q<p$ la función de autocorrelación estará determinada tan solo por la ecuación en diferencia

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)\, \rho_k = \theta \quad \text{si} \quad k>p$$

Sujeta a las p condiciones iniciales dadas por $\rho_1, \rho_2, \ldots, \rho_p$

Si $p <= q$ habrá q autocorrelaciones iniciales $\rho_1, \rho_2, \ldots, \rho_q$ que no sigan el comportamiento general de la función de autocorrelación, pero las autocorrelaciones $\rho_{q+1}, \rho_{q+2}, \ldots$ se comportarán de acuerdo con la ecuación en diferencia

$$(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)\, \rho_k = 0 \qquad \text{con} \qquad k>q$$

y las p condiciones iniciales $\qquad \rho_{q-p+1}, \rho_{q-p+2}, \ldots, \rho_q$

Procesos Autorregresivos Integrados de Medias Móviles ARIMA (p,d,q).-

Como ya se dijo, la mayor parte de las series no son estacionarias. Para ajustar una serie no estacionaria es necesario eliminar la fuerte variación no estacionaria. Si la serie de tiempo observada no es estacionaria en la media, lo cual implica que la serie presenta en su comportamiento una tendencia de tipo polinomial no determinista (no estacionariedad

homogénea), entonces podemos tomar las diferencias de las series de tal forma de obtener una serie estacionaria.
Si suponemos que $\overline{Z}_t$ es un proceso original no estacionario, con tendencia polinomial adaptivo de orden d entonces, se puede obtener un nuevo proceso $\omega_t = \nabla_d \ \overline{Z}_t$ tal que es estacionario el cual da origen a un proceso autorregresivo integrado de medias móviles.

Modelos Puramente Estacionales SARIMA (P,D,Q).-

Los modelos SARIMA captan el comportamiento puramente estacional de una serie, en forma similar, como hemos visto, se realiza para la componente regular o no estacional. Una serie con influencia solamente por la componente estacional puede ser descrito por un modelo SARIMA (P,D,Q), el cual se lo representa de la siguiente manera:

$$\phi_p (B^s) \nabla_S^D (Z_{t-u}) = \theta_Q (B^s) a_t$$

donde la constante μ es el nivel del proceso original Zt.

$$\phi_p (B^s) = 1 - \phi_s B^s - \phi_{2s} B^{2s} - \ldots - \phi_{ps} B^{ps}$$

Es un polinomio autorregresivo estacional de orden P

$$\phi_Q (B^s) = 1 - \theta_s B^s - \theta_{2s} B^{2s} - \ldots - \theta_{Qs} B^{Qs}$$

Es un polinomio de promedios móviles estacional de orden Q.
$\mathbf{a_t}$: Es un proceso de ruido blanco.

Como es de esperar en la práctica no siempre se presentan series con componente regular únicamente, o afectadas por la estacionalidad solamente, sino por el contrario, generalmetne se presentan series afectadas por ambas componentes, tendencia regular y estacionalidad. En este sentido Box y Jenkins (1970) propone un modelo denominado multiplicativo, el cual puede explicar el comportamiento de una serie afectada por ambas componentes.

Modelo Multiplicativo ARIMA (p,d,q) * SARIMA (P,D,Q).-

El modelo multiplicativo sugerido por Box y Jenkins para explicar a series con ambos efectos, parte del siguiente hecho.
Sea el modelo estacional:

$$\phi_P (B^s)\, \nabla_S^D (Z_{t-\mu}) = \theta_Q (B^s)\, \alpha_t$$

donde α es un proceso no de ruido blanco, sino mas bien generado por un proceso ARIMA (p,d,q) de la forma :

$$\phi_P (B)\, \nabla^D \alpha_t = \theta_q (B)\, a_t$$

Con $\mathbf{a}_t$ proceso de ruido Blanco.
De la expresión anterior despejando α_t se tiene que:

$$\alpha_t = \frac{\theta_k (B)\, a_t}{\phi_p (B)\, \nabla^d}$$

reemplazando la expresión del modelo multiplicativo:

$$\phi_P (B)\, \phi_P (B^S)\, \nabla^d\, \nabla_S^D (Z_{t-\mu}) = \theta_q(B)\, \theta_Q (B^S)\, a_t$$

De todo el estudio realizado se puede concluir que los modelos ARIMA son de gran utilidad en muchos campos. Entre las desventajas que se puede mencionar están la aplicabilidad a series con un número grande de observaciones, de por lo menos 30 datos para series afectadas solamente por la componente regular, y como mínimo 50 datos en series con estacionalidad, lo cual no permite ser aplicable a muchas variables económicas de nuestro país, por no contarse con la cantidad necesaria de datos.

Y entre las recomendaciones que se puede hacer es que siempre que se desee utilizar los modelos ARIMA, aplicar el operador diferencia un número óptimo de veces para no perder mucha información, y en casos donde sea necesario realizar transformaciones elegir la más sencilla o adoptar la logarítmica para no entrar en problemas de cálculo e interpretación.

En la etapa de modelación se recomienda ensayar en primera instancia los más simples, en caso de no satisfacer condiciones teóricas recién tomar modelos más complejos.
En la parte que corresponde a las proyecciones, realizar extrapolaciones para tiempos futuros no muy lejanos, ya que a fechas más lejanas como se pudo ver existirá menor precisión en las estimaciones.

Sin más comentarios se da fin a este trabajo realizado.