

Modelos de series de tiempo aplicados a los expedientes de la Comisión de Derechos Humanos del Distrito Federal

Time series models applied to the records of the Human Rights Commission of the Federal District

Marco Jair Guerrero Quintana*
Sonny Alberto Medina Jiménez**

Resumen

El presente documento, detalla el ajuste de modelos de series de tiempo ARIMA, para la realización de pronósticos sobre el histórico de expedientes acumulados en la Comisión de Derechos Humanos del Distrito Federal (CDHDF). El propósito es comprender el comportamiento de los expedientes acumulados en el tiempo y pronosticarlos.

Abstract

The document presented here, details the adjustment of ARIMA (Autoregressive Moving Average) time series models that allow to forecast the historic accumulation of files in the Mexico City Human Rights Commission. The purpose is to comprehend the behaviour of these, throughout time and to forecast the number of accumulated files.

Palabras clave:

- Modelos econométricos
- Derechos humanos
- Áreas básicas del Derecho

Keywords:

- Econometric Models
- Human Rights
- Basic Areas of Law

JEL: C5, R15, K1

Introducción

Los modelos autoregresivos integrados de medias móviles, ARIMA por sus siglas en inglés, son los modelos de series de tiempo aditivas más populares. Son una generalización de los modelos autoregresivos de medias móviles (ARMA), útiles cuando la serie de tiempo a modelar no es estacionaria a causa de una marcada tendencia.

En la primera sección analizaremos a detalle la serie de tiempo Expedientes Acumulados en la CDHDF, misma que está compuesta por las series de expedientes acumulados en cada una de las cinco Visitadurías Generales. En la segunda sección se tratará brevemente la teoría de los modelos ARIMA, enfatizando en los requerimientos estadísticos que debe poseer la serie para un

*Director de Evaluación y Estadística de la Comisión de Derechos Humanos del Distrito Federal.

Profesor de la Facultad de Economía UNAM.

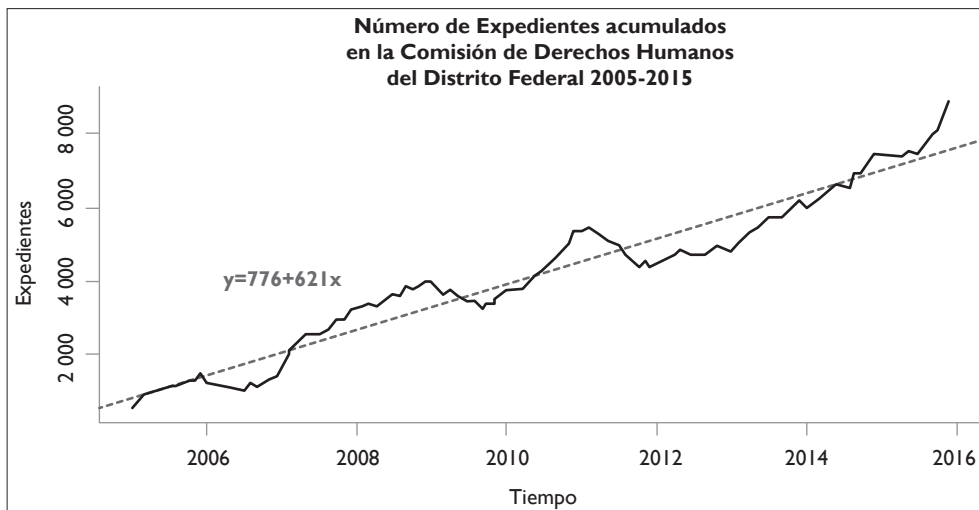
**Analista de la Dirección de Evaluación y Estadística de la Comisión de Derechos Humanos del Distrito Federal.

Conocer el número de expedientes acumulados en la Comisión durante un periodo mensual, o para cualquier periodo determinado, es un proceso estocástico compuesto de la adición del número de expedientes acumulados en cada una de las cinco Visitadurías Generales.

Sea X_t : el número de expedientes acumulados en la CDHDF.

$X_t = \sum_{(i=1)}^5 X_t^{(i)}$ donde $X_t^{(i)}$ es el número de expedientes acumulados en la i -ésima Visitaduría General. El comportamiento de la serie X_t puede ser observado en la figura 1.

Figura 1
Expedientes acumulados en la CDHDF



Si una serie es considerada aditiva, cualquier modelo de serie de tiempo adecuado a ésta debe estar conformado por las componentes de: tendencia, estacionalidad y aleatoria. De tal forma que si X_t es aditiva, debe ser de la forma:

$$X_t = T_t + S_t + E_t$$

Donde T : representa la tendencia, S : la estacionalidad y E : la componente aleatoria del proceso.

En la figura 2 se puede observar la descomposición automatizada de la serie X_t en las componentes de las series de tiempo aditivas.

En la figura 3 podemos observar las gráficas de cada una de las series $X_t^{(i)}$, el número de expedientes acumulados por visitaduría. Es importante hacer notar que no hay similitudes en varianza en general si por pares la primera Visitaduría y la segunda visitaduría; la tercera Visitaduría y la cuarta Visitaduría.

Figura 2
Descomposición de la serie X_t en sus componentes aditivas

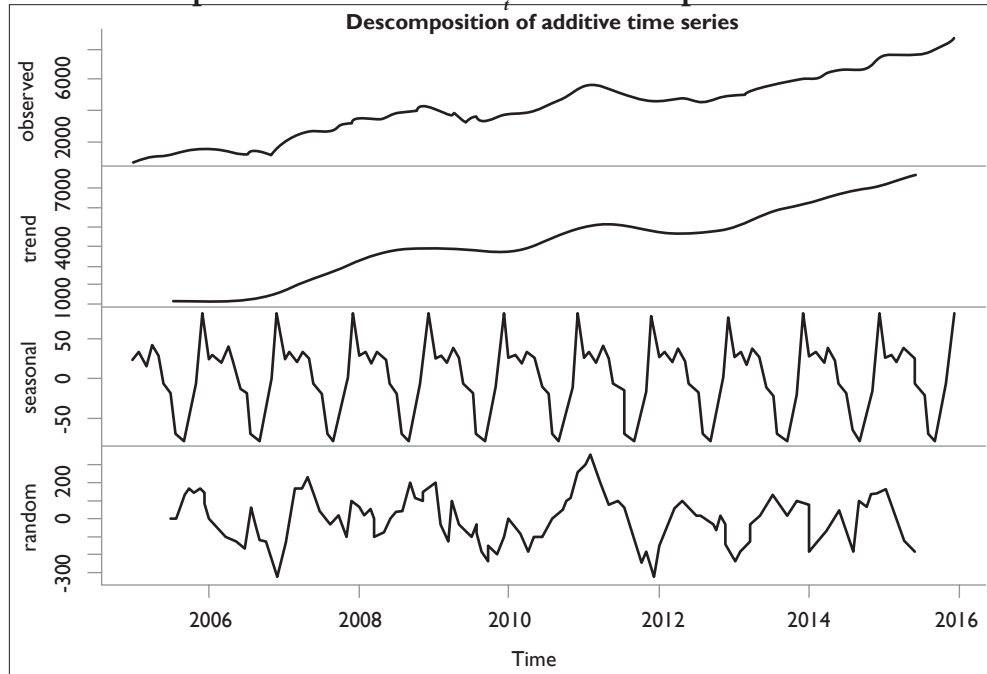
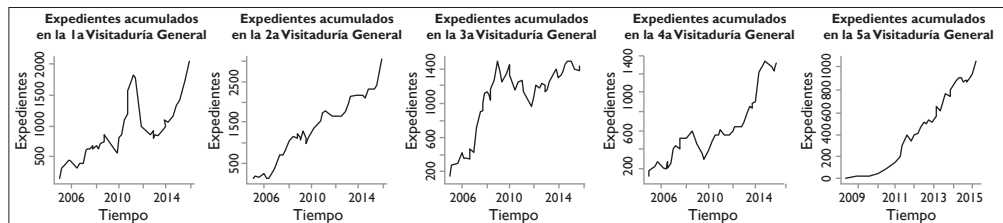


Figura 3
Expedientes acumulados en cada una de las Visitadurías Generales



2. Modelos ARIMA

Los procesos ARIMA son los más populares para el modelado de procesos econométricos, financieros, meteorológicos e incluso biológicos. Son una generalización de los procesos ARMA (Auto Regresivos y de Medias Móviles), usados cuando la serie de tiempo presenta una tendencia, que la vuelve un proceso no estacionario.

La función de autocovarianzas (ACVF) es un criterio eficiente para el ajuste de un modelo de series de tiempo. La ACVF teórica está definida como la covarianza entre x_t y x_{t+h} .

Definición: La función de autocovarianza muestral es definida como:

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x})$$

Se observa que $\hat{\gamma}(-h) = \hat{\gamma}(h)$, esto es, la función de autocovarianza es simétrica.

Consideremos x_t observaciones de una serie de tiempo y w_t un ruido blanco; es decir, w_t es una sucesión de variables aleatorias no correlacionadas, con media cero y varianza finita.

Definición: Un modelo autorregresivo de orden p , denotado como $AR(p)$, tiene la forma:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t$$

La función de autocorrelación parcial (facp) de un proceso $AR(p)$ (Autorregresivo de orden p) es 0 para valores de $h > p$.

Definición: Un promedio móvil de orden q , $MA(q)$ es definido como:

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q}$$

La función de autocovarianzas de un proceso $MA(q)$ es 0 para valores de $h > q$.

Definición: Una serie de tiempo $\{x_t; t = 0, \pm 1, \pm 2, \dots\}$ es un $ARMA(p, q)$ si es estacionaria y si:

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \dots + \theta_q w_{t-q}$$

Con $\Phi_p \neq 0$ y $\theta_q \neq 0$. Los parámetros p y q son llamados orden autorregresivo y orden de promedio móvil respectivamente. Asumimos que w_t es un ruido blanco gaussiano con media cero y varianza finita σ_w^2 . Cuando $q=0$ el proceso es llamado autorregresivo de orden p , $AR(p)$ y cuando $p=0$ el proceso es llamado promedio móvil de orden q , $MA(q)$.

Si suponemos que las observaciones $\{x_1, \dots, x_n\}$ no son necesariamente generadas por una serie de tiempo estacionaria, buscaremos una transformación que genere una nueva serie con las características deseadas para permitir el ajuste de un modelo $ARMA$, lo anterior es frecuentemente logrado mediante la diferenciación de la serie. Sin embargo, otra forma de tratar los datos, sin necesidad de aplicar transformaciones, es ajustar un modelo $ARIMA$.

Un ejemplo de la utilidad de los modelos $ARIMA$ es el modelado de un proceso con deriva. Si X_t es un proceso con deriva, i.e. $X_t = \delta + X_{t-1} + w_t$, entonces diferenciando el modelo $Y_t = \nabla_{x_t}$ obtenemos una serie de tiempo estacionaria, como en este ejemplo, en muchas otras situaciones una serie de tiempo puede ser pensada como la suma de dos componentes: una tendencia (no estacionaria) y una componente estacionaria con media cero. Los modelos $ARIMA$ son modelos que contemplan la diferenciación de los datos.

Definición: Un proceso x_t es un llamado $ARIMA(p,d,q)$ si el proceso:

$$\nabla_{x_t}^d = (1 - B)_{x_t}^d$$

Es un $ARMA(p,q)$ causal, en general escribiremos el modelo como:

$$\phi(B)(1-B)^d x_t = \theta(B)w_t$$

Otros modelos de interés

Existen otros modelos de series de tiempo más sencillos, pero igualmente eficientes al momento de generar pronósticos, el modelo en el que estamos interesados por su utilidad en el presente documento es el proceso Naïve Raw Drift, que significa proceso ingenuo de deriva bruta.

Las predicciones bajo el método Naïve, son simplemente el valor de la última observación, es decir $y_{T+h} = y_T$ con y_T el último valor observado.

El método Naïve Raw Drift es una variación del método Naïve, que permite a los pronósticos incrementarse en el tiempo, considerando un valor cambiante en el tiempo (llamado deriva), calculado como el cambio promedio en los datos históricos. De esta manera,

$$y_{t+h} = y_t + \frac{h}{T-1} \sum_{t=2}^T (y_t - y_{t-1}) = y_t + h \left(\frac{y_T - y_1}{T-1} \right)$$

3. Metodología

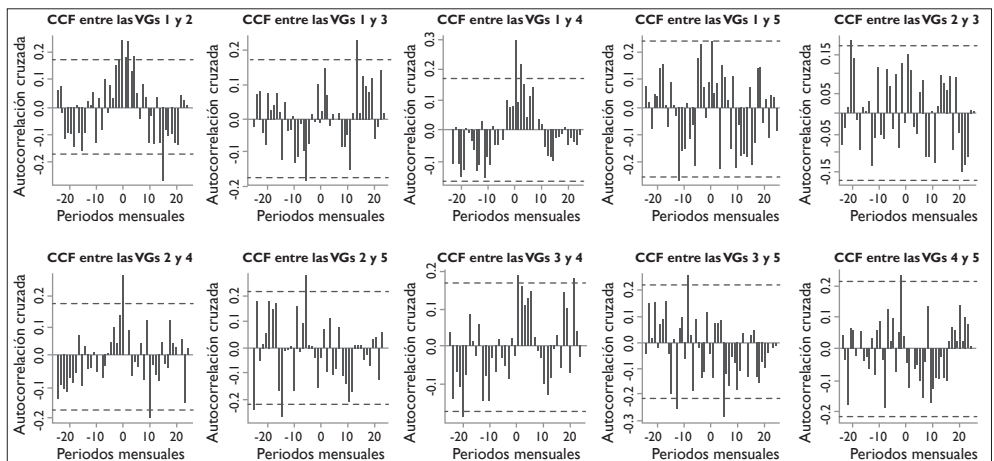
Como se pudo ver en la primera sección, la serie de tiempo X_t : número de expedientes acumulados en la CDHDF está compuesta por cinco componentes, las series que reflejan el número de expedientes acumulados en cada una de las cinco Visitadurías Generales. El ajuste de un modelo adecuado al comportamiento de X_t puede hacerse en forma agregada o desagregada:

- La primera opción implica ajustar un modelo de series de tiempo a la serie X_t ,
- La segunda opción propone ajustar modelos individuales, uno a cada uno de las $X_t^{(i)}$, de tal manera que al poder replicar el comportamiento de cada una de ellas, la serie X_t pueda ser modelada con la suma de los procesos en los modelos individuales.

La forma de modelado desagregada es preferible, el comportamiento de las series (como puede observarse en la gráfica 3) no es el mismo en cada una de las Visitadurías Generales. Los modelos individuales replicarán o se ajustarán con mayor fidelidad al comportamiento de cada una de las series. Sin embargo, el modelado desagregado depende de una hipótesis adicional a las del agregado, la independencia mutua de los $X_t^{(i)}$.

Para mostrar dicha independencia, se hace uso de la función de correlación cruzada (ccf), es un criterio que no determina si dos procesos son independientes, sí refleja la posible no correlación de estos. Habiendo demostrado la no correlación de las series de tiempo, procedemos al ajuste de los modelos individuales.

Figura 4
Función de Correlación Cruzada (CCF) entre Visitaduras Generales

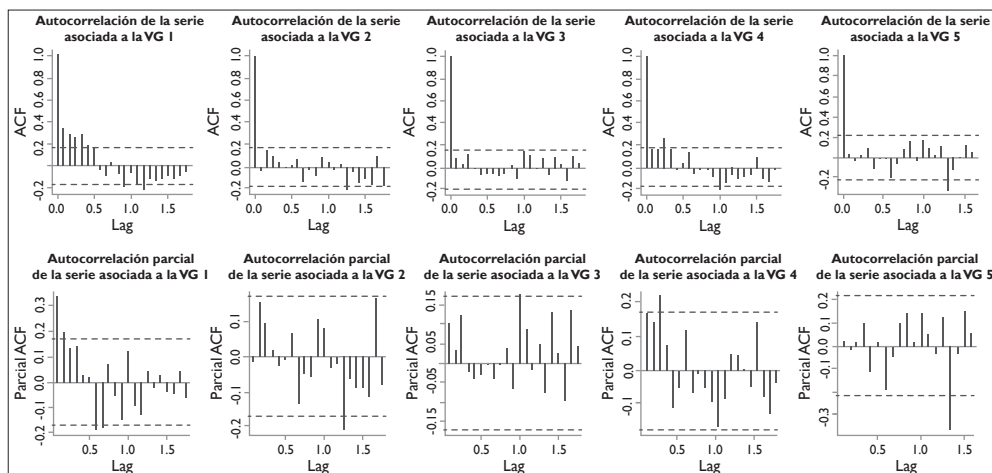


Ajuste del Modelo

Para realizar el ajuste de los modelos individuales, debemos retirar la componente de tendencia de cada una de las series, a fin de tener series de tiempo estacionarias. La forma de hacerlo es utilizando el operador ∇ , es decir, diferenciando.

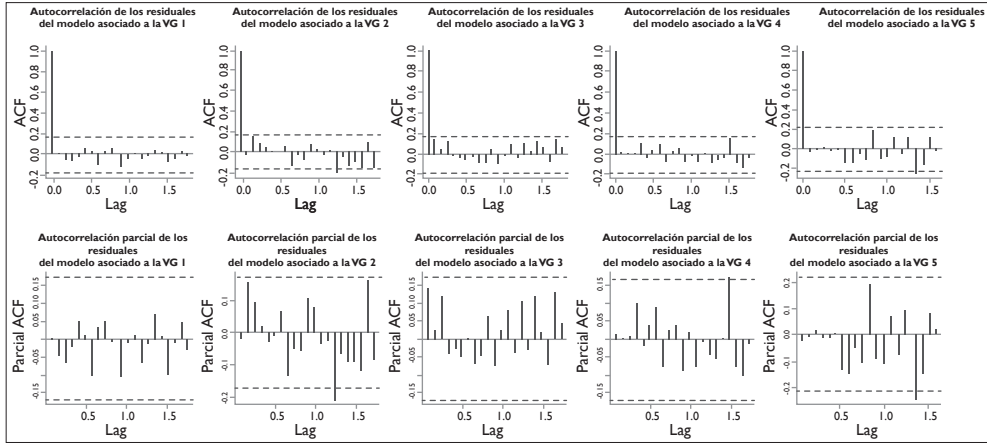
Una vez realizada la diferenciación, se grafican las funciones de autocorrelación y autocorrelación parcial, figura 5. En las gráficas de estas funciones se observa que la serie posee un comportamiento estacionario y algunas características adicionales que facilitan el ajuste de los modelos. Véanse las características de las funciones de autocorrelación y autocorrelación parcial de los modelos *AR* y *MA* explicados en la sección anterior.

Figura 5
Funciones de autocorrelación y autocorrelación parcial por cada visitaduría general



Se busca que los residuales del modelo sean no correlacionados y despreciables, para lo cual se analizan las funciones de autocorrelación y autocorrelación parcial. Los resultados pueden observarse en la figura 6.

Figura 6
Funciones de autocorrelación y autocorrelación
parcial de los residuos de los modelos individuales



Implementación del Modelo

Para la selección de los modelos individuales definitivos se compararon diversos procesos, usando los siguientes criterios: AIC, BIC, Pruebas de razón de verosimilitudes entre modelos anidados, Pruebas de Wald sobre los parámetros del modelo y Principio de Parsimonia.

Los modelos ajustados fueron los siguientes:

- Primera Visitaduría General, $ARIMA(8,1,8)$

$$\phi_8(B)\nabla X_t^{(1)} = \theta_8(B)w_{-t}$$

- Segunda Visitaduría General, Raw Drift Model

$$X_t^{(2)} - \delta = X_{t-1}^{(2)}$$

- Tercera Visitaduría General, $SARIMA(0,1,0)(1,0,2)_{12}$

$$\Phi(B^{12})\nabla X_t^{(3)} = \Theta_2(B^{12})w_t$$

- Cuarta Visitaduría General, $SARIMA(0,1,3)(0,0,1)_{12}$

$$\nabla(X_t^{(4)} - \delta) = \theta_3(B)\Theta(B^{12})w_t$$

- Quinta Visitaduría General, $SARIMA(4,1,0)(0,1,1)_{12}$

$$\phi_4(B)\nabla\nabla_{12}X_t^{(5)} = \Theta(B^{12})w_t$$

4. Pronósticos

Los pronósticos a un año pueden consultarse en la tabla 1. Los intervalos de confianza para los pronósticos de la serie de tiempo fueron realizados a un nivel de 90% de confianza.

Figura 7
Pronósticos de la serie de tiempo *Número de expedientes acumulados en la CDHDF a un año*

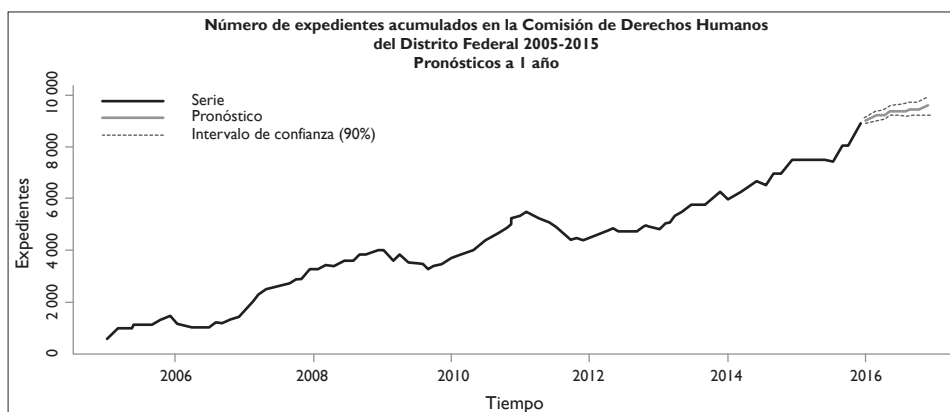


Tabla I
Pronósticos del número de expedientes acumulados en la CDHDF para el 2016

Mes	Media	Límite Inferior	Límite Superior
Enero	8 992	8 904	9 081
Febrero	9 100	8 975	9 226
Marzo	9 206	9 051	9 361
Abril	9 236	9 053	9 418
Mayo	9 339	9 132	9 545
Junio	9 363	9 136	9 591
Julio	9 370	9 123	9 617
Agosto	9 397	9 132	9 663
Septiembre	9 429	9 147	9 711
Octubre	9 468	9 170	9 766
Noviembre	9 518	9 205	9 831
Diciembre	9 585	9 258	9 912

5. Conclusión

Se ajustaron modelos de series de tiempo individuales adecuados al comportamiento del número de quejas acumuladas en cada una de las Visitadurías Generales de la CDHDF. Se presentaron pronósticos sustentados en regiones de confianza para el comportamiento de las mismas. Se mostró la validez de cada uno de los modelos mediante al análisis de residuales.

Contar con un modelo adecuado del comportamiento de las series, nos permite establecer cotas de comportamiento que sirvan como puntos de referencia para la detección de comportamientos anormales en el número de expedientes acumulado, de tal forma que se determine la llegada a valores atípicos que reflejen que el número de expedientes registrados ha aumentado significativamente, o bien que el número de expedientes concluidos es menor al esperado. De la misma manera será fácil determinar si el número de quejas es bajo, proporcionando fundamentos para la toma de decisiones.

Referencias

- Brockwell Perter J. and Richard A. Davis (2002), *Introduction to Time Series and Forecasting*, Second Edition, Springer.
- Robert H. Schumway and David S. Stoffer (2010), *Time Series Analysis and Its Applications. With R Examples*, Third edition. Springer.
- Ruey S. Tsay (2010), *Analysis of Financial Time Series*, Third Edition, Wiley.
- Haugh D. Larry (1976), Checking the Independence of Two Covariance-Stationary Time Series: A Univariate Residual Cross-Correlation Approach, *Journal of the American Statistical Association*, vol. 71, núm. 354 (jun.), pp. 378-385.
- Guerrero M. Victor (2004), *Análisis estadístico de series de tiempo económicas generadas con datos oficiales*, INEGI.
- E. P. Box, M. Jenkins and C. Reinsel (1994), *Time Series Analysis, Forecasting and Control*, Third Edition, Wiley.