Comunicaciones en Estadística
Junio 2009, Vol. 2, No. 1

Estimación de observaciones faltantes en series de tiempo usando métodos multivariados con restricciones

Time Series Missing Data Imputation Using Restricted Multivariate Methods

María Ana Velásquez Gallo[a]
hanwenzhang@usta.edu.co

Jorge Martínez Collantes[b]
hugogutierrez@usantotomas.edu.co

Abstract

Para tener éxito en el modelamiento de cualquier fenómeno físico, es importante disponer de una fuente de información completa. En este documento se plantea la reconstrucción o estimación de observaciones faltantes en series de tiempo múltiples, siguiendo la metodología de Guerrero (1989) de pronósticos con restricciones las cuales vienen dadas por información adicional relacionada con el modelo, que tiene en cuenta su información histórica y su estructura de correlación, minimizando el error cuadrático medio, y asumiendo además, que no se presentó un cambio de estructura en el segmento faltante, lo cual se verifica mediante una prueba de bondad de ajuste. El pronóstico óptimo se restringe a conservar la estructura inicialmente identificada, de manera que empalmen los pronósticos obtenidos con la información posterior disponible, y con esto, no se vean afectados los parámetros, la parte estocástica, y la parte deterministica del modelo.

***Key words*:** Series de tiempo multiples, Restricciones lineales, Pruebas de compatibilidad, Cambios estructurales..

Resumen

Para tener éxito en el modelamiento de cualquier fenómeno físico, es importante disponer de una fuente de información completa. En este documento se plantea la reconstrucción o estimación de observaciones faltantes en series de tiempo múltiples, siguiendo la metodología de Guerrero (1989) de pronósticos con restricciones las cuales vienen dadas por información adicional relacionada con el

[a]Docente. CIEES, Universidad Santo Tomás
[b]Director. Centro de Investigaciones y Estudios Estadísticos (CIEES)

modelo, que tiene en cuenta su información histórica y su estructura de correlación, minimizando el error cuadrático medio, y asumiendo además, que no se presentó un cambio de estructura en el segmento faltante, lo cual se verifica mediante una prueba de bondad de ajuste. El pronóstico óptimo se restringe a conservar la estructura inicialmente identificada, de manera que empalmen los pronósticos obtenidos con la información posterior disponible, y con esto, no se vean afectados los parámetros, la parte estocástica, y la parte deterministica del modelo.

***Palabras clave*:** Series de tiempo multiples, Restricciones lineales, Pruebas de compatibilidad, Cambios estructurales..

1 INTRODUCCIÓN

Las series de tiempo son una herramienta útil para la planificación futura del comportamiento de eventos o fenómenos medidos en el tiempo. Siendo estos de carácter natural, climático, agrícola, mercantil, económico, social, entre otros, situaciones que inciden en el diario vivir, o en la toma de decisiones de una comunidad determinada.

Desde este punto de vista, es de gran importancia trabajar con las herramientas adecuadas para el modelamiento de dichos fenómenos. Gran parte del éxito de un buen modelamiento depende de la calidad y disponibilidad de una fuente información, que además sea "completa".

El propósito de este articulo, es analizar tres técnicas de estimación de datos faltantes para la reconstrucción de series temporales múltiples, en primera instancia haciendo uso de la teoría de pronósticos con restricciones desarrollada por Guerrero (1989), que contempla la imposición de restricciones sobre la función de pronostico mediante una combinación lineal, que puede incluir una componente aleatoria, el cálculo de estos pronósticos se obtiene a partir de los clásicos pronósticos de un modelo ARIMA, que posteriormente es desarrollada por Guerrero y Peña (1998) que hacen uso de información adicional por medio de restricciones, en segundo lugar se trabaja el problema de la reconstrucción de las series de tiempo con la metodología de Damsieth (1976), quien propuso la estimación de los valores faltantes por medio de una combinación lineal entre un pronóstico retrospectivo y uno prospectivo y por último se retoma el trabajo de Nieto (1984), que obtuvo los pronósticos restringidos haciendo uso del filtro de Kalman(1960). Que a partir de la representación canónica o markoviana de un modelo de estado-espacio de un sistema dinámico[1], resume toda la información necesaria del presente y el pasado para la predicción de valores futuros. Procedimiento que se calcula mediante un filtro que opera a través de un mecanismo de predicción y corrección, para diagnosticar un nuevo estado a partir de una estimación previa, a la que se le añade

[1] Que cambian permanentemente sus parámetros

un término de corrección proporcional al error de predicción. Técnica que ha sido ampliamente trabajada en la literatura estadística por diversos autores.

En la sección dos se hace un breve recuento de la teoría estadística para la construcción de modelos VAR, continuando con el desarrollo de la técnica para calcular pronósticos restringidos cuando no se presentó un cambio de estructura en el horizonte de pronóstico y su respectiva prueba de hipótesis. Posteriormente, en la sección cuatro, se presentan resultados para el cálculo de dichos segmentos faltantes, cuando en el modelo se presentó un cambio estocástico o determinístico en la serie, aunque en las simulaciones no se considera este caso y finalmente se encuentran los resultados de las simulaciones, hechas en el software estadístico R(versión 2.4) y SAS(versión 9.0).

2 MODELO VAR

El proceso vectorial autoregresivo $AR(p)$ es un modelo estocástico asociado a un conjunto $Z_t = \{Z_{1t}, Z_{2t}, Z_{3t}, \ldots, Z_{nt}\}'$ $n-dimensional$ constituido por n series de tiempo, el cual tiene en cuenta las correlaciones conjuntas, individuales y a diferentes rezagos del tiempo para las n series.

Dicho modelo esta dado por:

$$\Phi_p(B)Z_t = a_t \qquad a_t \sim N(0, \Sigma) \tag{1}$$

con

$$\Phi_p(B) = [I - \Phi_1(B) - \cdots - \Phi_p(B)]$$

Este proceso es estacionario si las raíces están fuera del circulo unidad, que en términos prácticos quiere decir que las distribuciones de probabilidad de los vectores Z_t y Z_{t+l}. son invariantes en el tiempo, para $l = 0, 1, 2, 3, \ldots$

De modo que se puede identificar una distribución asociada al modelo $AR(p)$, con el cálculo de su esperanza y su varianza, a partir de la obtención de su primer y segundo momento, que existen y son finitos, por medio de:

$$E(Z_t) = \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

$$\Gamma(k) = Cov(Z_t, Z_{t+k}) = \begin{bmatrix} \gamma_{11}(k) & \cdots & \gamma_{1n}(k) \\ \vdots & \ddots & \vdots \\ \gamma_{n1}(k) & \cdots & \gamma_{nn}(k) \end{bmatrix}$$

Con matriz de correlación igual a:

$$\rho(k) = D^{-1/2}\Gamma(k)D^{-1/2} = \begin{bmatrix} \rho_{11}(k) & \cdots & \rho_{1n}(k) \\ \vdots & \ddots & \vdots \\ \gamma_{n1}(k) & \cdots & \gamma_{nn}(k) \end{bmatrix}$$

con $D = diag(\Gamma(0))$ De donde se extrae el conocimiento de las correlaciones[2] entre las componentes del proceso Z_t

El proceso vectorial $AR(p)$ de (1) puede expresarse en su forma de promedio móvil, mediante:

$$Z_t = [\Phi_p(B)]^{-1}a_t = \sum_{s=0}^{\infty} \Psi_s a_{t-s} \tag{2}$$

Bajo la condición de que $\sum_{s=0}^{\infty} \Psi_s a_{t-s}$ sea absolutamente convergente, obteniéndose las matrices Ψ_i por medio de la ecuación

$$\Phi_p(B)[I - \Psi_1 B^1 - \cdots - \Psi_p B^p - \cdots - \cdots] = I$$

resultando:

$$\Psi_i = \begin{cases} I & \text{para } i = 0 \\ \Phi_1\Psi_{i-1} + \cdots + \Phi_p\Psi_{i-p} & \text{para } i \geq p \\ 0 & \text{e.o.c.} \end{cases}$$

Con las a_s, variables independientes idénticamente distribuidas ($a_s \sim N(0, \Sigma_a)$)

Teniendo presente la forma en que se puede representar un proceso $AR(p)$, su matriz de covarianza, se define como:

$$\begin{aligned} \Gamma(k) &= E(Z_t; (Z'_t - Z'_{t-1}\Phi_1 - \cdots - Z'_{t-p}\Phi_p)) \\ &= E(Z_t; \sum_{s=0}^{p} \Phi_s Z_{t-s} + a_t) \\ &= \sum_{s=0}^{p} \Gamma(k-s)\Phi'_s \qquad k = 0, 1, 2, 3, \ldots \end{aligned} \tag{3}$$

Y su matriz de correlación estimada es:

$$\hat{\rho}(k) = [\hat{\rho}_{ij}(k)] \qquad \hat{\rho_{ij}}(k) = \frac{\sum\limits_{t=1}^{n-k}(Y_{it} - \bar{Y}_i)(Y_{jt} - \bar{Y}_j)}{\sum\limits_{t=1}^{n-k}(Y_{it} - \bar{Y}_i)\sum_{t=1}^{n-k}(Y_{it} - \bar{Y}_i)} \tag{4}$$

A esta matriz de correlación se le asocia una varianza aproximada[3] de $\frac{1}{n}$

$$Var(\hat{\rho}_{ij}(k)) \approx \frac{1}{n}$$

De tal forma, que el orden del modelo se identifica a partir del número de matrices de correlación significativas[4], o en las que sus componentes en valor absoluto son mayores a $\frac{2}{\sqrt{n}}$.

[2]Herramienta fundamental para la identificación del modelo
[3]Bartlett 1966
[4]Box y Tiao (1981)

3 PRONÓSTICOS RESTRINGIDOS

La idea de esta técnica es obtener pronósticos para segmentos faltantes en un proceso $n - dimensional$ ajustado a un modelo $AR(p)$ con la información disponible, asumiento que el proceso no cambio la estructura identificada, de manera que este sea restringido a conservar dicha estructura.

Esta metología es comunmente usada en económia, siendo las restricciones identidades económetricas. Designamos a $Z = [Z_1', \cdots, Z_n]'$ como el vector que contiene la información disponible de la serie múltiple[5]. $Z_F = [Z_{n+1}' \cdots Z_{n+h}']'$ el vector de valores faltantes a ser estimados, con un horizonte de pronóstico $h \geq 1$, empezando desde $h = n + 1$

El pronóstico lineal optimo[6] basado en la información histórica contenida en el modelo $AR(p)$ esta dado por la esperanza condicional $E(Z_{n+h_1} \mid Z)$ para $h_1 = 1, 2, \cdots, h$, donde:

$$Z_t - E(Z_{n+h} \mid Z) = \sum_{s=0}^{h-1} \Psi_s a_{n+h-s} \tag{5}$$

con $Z_{n+h} = E(Z_{n+h} \mid Z)$ para $h \leq 0$. Al que se le asocia un error cuadrático medio de:

$$ECM(E(Z_F \mid Z)) = E((Z_F - E(Z_F \mid Z))(Z_F - E(Z_F \mid Z))) \quad = \Psi(I_h \otimes \Sigma_a)\Psi'$$

Con

$$\Psi = \begin{bmatrix} \psi_0 & 0 & \cdots & 0 \\ \psi_1 & \psi_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \psi_{h-1} & \psi_{h-1} & \cdots & \psi_0 \end{bmatrix}_{(kh)\times(kh)}$$

Donde $\psi_0 = I_k$, y ψ_i son las matrices identificadas en el modelo.

Para esta metodología, empezamos suponiendo que existe un vector $Y = (Y_1, \cdots, Y_m)$ de información adicional que se relaciona con el proceso. Esta información es un conjunto de m restricciones estocásticas linealmente independientes, a ser cumplidas en el horizonte de pronósticos, que además pueden estar dadas por la forma en que se encuentren distibuidos los r segmentos faltantes. Dichas restricciones vienen dadas por :

$$Y = CZ_F + U \text{ Con } U \sim N(0, \Sigma_U);\ U \perp Z \tag{6}$$

Con C, una matriz de restricciones conocida de dimensión $m \times (kh)$, donde el número de restricciones debe ser inferior a la cantidad de series por el número

[5] Donde hay contenidas k series
[6] Mínimo error cuadrático medio

de pronósticos ($m < kh$). El vector de ruido U[7] que captura la diferencia entre la matriz CZ_f de información adicional y las verdaderas observaciones dee ser ortogonal al conjunto de información disponible, de modo que las restricciones sean insesgadas.

Luego el pronóstico optimo restringido se obtiene mediante:

$$\hat{Z}_F = E(Z_F \mid Z) + A[Y - CE(Z_F \mid Z)] \tag{7}$$

Con

$$A = \Psi(I_h \otimes \Sigma_a)\Psi' C' [C\Psi(I_h \otimes \Sigma_a)\Psi' C' + \Sigma_U]^{-1}$$

Y un error cuadrático medio de:

$$ECM(\hat{Z}_F) = (I_{hk} - AC)\Psi(I_h \otimes \Sigma_a)\Psi' \tag{8}$$

Satisfaciendo asimismo la siguiente igualdad.

$$ECM[E(\hat{Z}_F \mid Z)] = E(\hat{Z}_F) + AC\Psi(I_h \otimes \Sigma_a)\Psi'$$

4 PRONÓSTICOS RESTRINGIDOS CON LA PRESENCIA DE UN CAMBIO ESTRUCTURAL IDENTIFICADO EN EL MODELO

En algunas series de tiempo, es común encontrar la presencia de outliers, y en la mayoría de estudios, a estos se les aplica técnicas univariadas para ser removidos, sin tener en cuenta, que en un proceso múltivariado, un outlier de una componente puede ser causado por otras componentes, y el impacto generado por este, no es captado en toda su totalidad por los métodos univariados, ya que estos métodos no tienen la capacidad de combinar información a lo largo de las componentes de la serie. De manera que la detección de outliers es un factor de gran importancia en el modelamiento de series múltiples, ya que de esto depende la válidez y futuras inferencias que se hagan acerca del modelo. La presencia de outliers es un cambio estructural que se debe a perturbaciones endógenas o exógenas en los valores observados, ocasionando sesgo en el valor de los parámetros y arrojando pronósticos bastante deficientes o pobres, al no ser modelados adecuadamente. Tsay (2000) generaliza al caso múltivariado los cuatro tipos de cambios estructurales, para cambios deterministicos, mediante la siguiente expresión:

$$Z_t^* = Z_t + \alpha(B)\omega\varepsilon_t^{(h)} \tag{9}$$

$$\alpha(B) = \begin{cases} \Phi(B) & \text{para un oulier innovativo} \\ I & \text{para un oulier aditivo} \\ (1-B)^{-1}I & \text{para un cambio de nivel} \\ [D(\delta)]^{-1} & \text{para un cambio temporal} \end{cases}$$

[7]proviene de una condición especifica del problema que se este analizando

En donde $D(\delta)$ es una matriz diagonal de dimensión $k \times k$ con elementos $(1 - \delta_1 B), \cdots, (1 - \delta_k B)$ con $0 < \delta_i < 1$, $\omega = (\omega_1, \cdots, \omega_t)$ es el tamaño del impacto inicial del outlier en la serie Z_t, y $\varepsilon_t^{(h)}$ es una variable indicadora para el subindice de tiempo h.

$$\varepsilon_t^{(h)} = \begin{cases} 1 & t = h \\ 0 & \text{e.o.c.} \end{cases}$$

Y para cambios estócasticos, [8] los cambios estructurales se expresan mediante:

$$Z_t^* = Z_t + \alpha(B)\varsigma_t S_t^{(h)} \tag{10}$$

$$\alpha(B) = \begin{cases} \Phi(B) & \text{para un cambio innovativo en la varianza} \\ I & \text{para un cambio aditivo en la varianza} \end{cases}$$

Donde ς_t es una secuencia de vectores aleatorios i.i.d.$N(0; \Sigma_\varsigma)$ independientes del ruido del modelo Z_t, donde Σ_ς es la matriz de covarianza que perturba el proceso, y $S_t^{(h)}$ es una variable indicadora que se define como:

$$S_t^{(h)} = \begin{cases} 1 & t \geqslant h \\ 0 & \text{e.o.c.} \end{cases}$$

Para un cambio determinístico descrito por (9), el horizonte de pronóstico está dado por:

$$Z_{F,D} = Z_F + D_F$$

Donde $Z_{F,D} = (Z_{n+1}^*, \cdots, Z_{n+h}^*)$ es un vector de dimensión $(kh)x1$ que contiene los valores a ser pronósticados, y $D_F = (d_{n+1}, \cdots, d_{n+h})$ es un vector de la misma dimensión de $Z_{F,D}$, que contiene el cambio determinístico. Siendo $d_t = \alpha(B)\omega\varepsilon_t^{(h)}$

En forma similar, para un cambio estócastico descrito por(10) el horizonte de pronóstico está dado por:

$$Z_{F,V} = Z_F + V_F$$

Donde $Z_{F,V} = (Z_{n+1}^*, \cdots, Z_{n+h}^*)$ es un vector de dimensión $(kh)x1$ contiene los valores a ser pronósticados, y $V_F = (v_{n+1}, \cdots, v_{n+h})$ también tiene la misma dimensión de $Z_{F,V}$, que contiene el cambio estocástico. Teniéndose que $v_t = \alpha(B)\varsigma_t S_t^{(h)}$

Finalmente se obtiene una expresión que combina cambios determinísticos y estocásticos, a través de:

$$Z_{F,D,V} = Z_F + D_F + V_F$$

De manera que el pronóstico calculado con la información histórica es:

$$E(Z_{F,D,V} \mid Z) = E(Z_F \mid Z) + D_F$$

[8] ver Guerrero V. (2006)

Al que se le asocia un Error Cuadrático Medio de:

$$ECM(Z_{F,D,V} \mid Z) = \Psi(I_h \otimes \Sigma_a)\Psi' + (I_h^* \otimes \Sigma_\varsigma)$$

En donde $I^* = diag(0, \cdots, 0, 1, \cdots, 1)$ es una matriz diagonal de dimensión hxh con las primeras $h_1 - 1$ posiciones correspondientes a ceros, y h_1 fué la observación donde se presentó el cambio estructural, con $1 \leqslant h_1 \leqslant h$.

La restricción viene dada de la misma forma que (6). Encontrándose como resultado el cálculo del pronostico óptimo a través de:

$$\hat{Z}_{F,D,V} = E(Z_{F,D,V} \mid Z) + A[Y - CE(Z_{F,D,V} \mid Z)] \tag{11}$$

Con:

$$A = \Sigma_{F,D,V} C'[C\Sigma_{F,D,V} C' + \Sigma_U]^{-1}$$

Y un Error Cuadrático Medio asociado a:

$$ECM(\hat{Z}_{F,D,V}) = (I_{Hk} - AC)\Sigma_{F,D,V} \tag{12}$$

Para más detalles acerca de la identificación de outliers en series de tiempo multiples, puede consultarse Tsay R. (2000) y Guerrero V. (2006).

5 PRUEBA DE HIPÓTESIS PARA UN CAMBIO ESTRUCTURAL

Para que el pronóstico calculado en (7) sea realmente óptimo, se debe probar que las fuentes de información disponibles, Y y $CE(Z_F \mid Z)$ son realmente compatibles, para este propósito se hace uso de una prueba de bondad de ajuste, que no es más que una generalización de la propuesta de Guerrero (1989).

Siendo el conjunto de hipótesis simple a verificar:

$$\begin{cases} H_o : Y \text{ es compatible con } CE(Z_F \mid Z) \\ H_a : Y \text{ no es compatible con } CE(Z_F \mid Z) \end{cases}$$

Obteniéndose como resultando la aceptación de la compatibilidad entre la información adicional del modelo Y con la información del modelo $CE(Z_F \mid Z)$, si la distancia entre dichos vectores es "cercana a cero".

Luego la diferencia entre estos dos vectores es calculada como:

$$d = Y - CE(Z_F \mid Z) = C\Psi a_F + U$$

De manera que

$$d \sim N(0; C\Psi(I_h \otimes \Sigma_a)\Psi' C' + \Sigma_U)$$

Asi que

$$w = d'[C\Psi(I_h \otimes \Sigma_a)\Psi'C' + \Sigma_U]^{-1}d \sim \chi^2(m)$$

$$\begin{aligned} w_{calculado} &= d'[C\Psi(I_h \otimes \Sigma_a)\Psi'C' + \Sigma_U]^{-1}d \\ &\leq \chi^2_{(m)(\alpha)} \end{aligned} \tag{13}$$

Entonces, con un nivel de significancia del $100\alpha\%$, no se rechaza que Y sea compatible con $CE(Z_F \mid Z)$, si $w_{calulado}$ de (8) es menor o igual que el $\alpha -$ ésimo percentil superior de una distribución χ^2 con m grados de libertad.

Cuando no se acepta H_o, y son declaradas incompatibles las fuentes de información, puede considerarse que se presentó un cambio estructural en el horizonte de pronóstico, lo cual afecta:

El calculo de los parámetros

La parte deterministica del modelo

La parte estocástica del modelo

6 ESQUEMA DE PRONOSTICOS RESTRINGIDOS

La figura 1 presenta un diagrama de flujo, que resume el procedimiento para calcular pronósticos restringidos.

7 FILTRO DE KALMAN

Wei (1989) define el estado de un sistema como el mínimo conjunto de información del presente y del pasado, tal que el comportamiento del futuro puede describirse completamente por el conocimiento del estado presente del sistema y las entradas futuras. Esta característica se tiene por la propiedad markoviana de un proceso estocástico de esperanza condicional, que nos dice que dado un estado presente, el futuro es independiente del pasado. Luego la representación estado-espacio[9] de un sistema es la base fundamental para el cálculo de pronósticos. El filtro de kalman (1960) es un algoritmo recursivo, que se va actualizando en cada iteración, pronosticando un nuevo estado a partir de la estimación anterior, al que se le agrega un término de corrección proporcional al error de predicción. Este procedimiento cumple con la condición de Gauss-Markov, que afirma que entre todos los estimadores lineales insesgados, los estimadores de mínimos cuadrados tienen el mínimo error cuadrático medio.

[9] Conocida también como representación markoviana

Figure 1: Diagrama de Pronósticos Restrigidos

La finalidad de este mecanismo es estimar y predecir el movimiento de una variable no observable mediante la medición de su efecto, contaminado por un ruido, a través de otras variables, para hacer inferencias sobre el vector de estado.

Si el sistema es lineal[10] e invariante[11], el espacio de estados se representa como:

$$\begin{cases} Y_t = AY_{l-1} + GX_{t+1} \text{ Ecuación de medición} \\ Z_t = HY_t \text{ Ecuación de transición} \end{cases} \tag{14}$$

Donde

Y_t es el vector de estados de dimensión $1 \times k$

X_t es el vector de entrada al sistema de dimensión $n \times 1$

Z_t: Vector de salida de dimensión $m \times 1$

A es la matriz de transición de dimensión $k \times k$

[10]Si y solo si la combinación lineal de un sistema de entradas $aX_{1t} + bX_{2t}$ produce la misma combinación lineal de las salidas $aY_{1t} + bY_{2t}$ para algunas constantes a y b.

[11]Si las características del sistema no cambian con el tiempo, es decir, si una entrada X_t produce una salida Y_t, entonces la entrada X_{t-t_0} producirá la salida Y_{t-t_0}.

G es la matriz de entrada de dimensión $k \times n$

H es la matriz de observaciones de salida de dimensión $m \times k$

Existiendo dos grupos de ecuaciones: (1) las de actualización del tiempo o predicción[12] y (2) las de actualización de los datos observados[13]

Como se tiene que X_t y Z_t son procesos estocásticos, entonces la representación del espacio de estados es:

$$\begin{cases} Y_t = AY_{l-1} + Ga_{t+1} \text{ Ecuación de medición} \\ Z_t = HY_t + b_t \text{ Ecuación de transición} \end{cases} \tag{15}$$

Con $a_t = X_t - E(X_{t+1} \mid X_t : t < n)$, G, H, y A matrices fijas y conocidas, y

$$\begin{cases} a \sim N(0; \Sigma) \\ b \sim N(0; \Omega) \\ \begin{pmatrix} a \\ a \end{pmatrix} \sim N \begin{pmatrix} \Sigma & 0 \\ 0 & \Omega \end{pmatrix} \end{cases}$$

Teniendo como resultado el cálculo del l-ésimo pronóstico hacia delante dado por el conjunto de ecuaciones detalladas abajo:

$$\hat{Y}_t(l) = E\{Y_{t+1} \mid Y_j; j \leq t\} = A\hat{Y}_t(l-1) = A \cdot A\hat{Y}_t(l-1) = \cdots = A^l\hat{Y}_t$$

$$\hat{Z}_t = E\{Z_{t+1} \mid Z_j; j \leq t\} = H\hat{Y}_t(l) = HA^l\hat{Y}_t$$

$$\hat{Y}_t = E\{Y_t \mid Y_j; j \leq t\} = Y_t$$

Kalman propuso este filtro en 1960, y sus desarrollos surgieron en el ámbito de la ingeniería. Durante la década de los 60, Kalman fue líder en el desarrollo de una teoría rigurosa de los sistemas de control. Entre sus contribuciones geniales esta la formulación y el estudio de la mayoría de las nociones fundamentales del estado. Que hoy en día son herramientas estándares en la investigación. Durante los años 70 Kalman desempeñó un papel importante en la introducción de técnicas algebraicas y geométricas en el estudio de los sistemas de control lineares y no lineales. En la década de los 80, su trabajo se centro en un acercamiento las fundamentaciones de la estadística, el modelamiento económico y econométrico.

8 DIGRAMA DEL FILTRO DE KALMAN

Igualmente, para el filtro de Kalman, presentamos un diagrama resumen, del funcionamiento de esta tecnica. Ver figura 2.

[12] Ecuación de medición
[13] Ecuación de transición

Figure 2: filtro de kalman

9 RESULTADOS

Por medio de la simulación de un proceso $Z_t = (Z_{2t}, Z_{2t})$ autoregresivo estacionario multivariado, de orden dos, $AR(2)$, con 200 observaciones en cada componente. Se ajustó un modelo vectorial ARIMA correspondiente a la figura 3 y descrito por el modelo (1):

$$\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0.07 \\ 0 & -0.03 \end{bmatrix} B - \begin{bmatrix} 0 & -0.061 \\ 0 & -0.060 \end{bmatrix} B^2 \right) \begin{bmatrix} Z_{1t} \\ Z_{2t} \end{bmatrix} = \begin{bmatrix} \alpha_{1t} \\ \alpha_{2t} \end{bmatrix}$$

con

$$\begin{bmatrix} \alpha_{1t} \\ \alpha_{2t} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} ; \begin{bmatrix} 1.006 & -0.067 \\ -0.067 & 0.891 \end{bmatrix} \right)$$

[14]

[14] con un criterio de Akaike de -0.09

Figure 3: Simulación del proceso vectoriar $Z_t = (Z_{1t}, Z_{2t})$

Modelo 1: Modelo ARIMA que ajusta la simulación.

Para este modelo se asumió la presencia de segmentos con información perdida, y para reconstruirlos se proponen las siguientes restricciones:

Una restricción Y_1, asumiendo una incertidumbre nula , es decir $U = 0, \Sigma_U = 0$ y

$$Y_1 = -2Z_{1(n+2)} + Z_{1(n+3)} - 2Z_{2(n+2)} + Z_{2(n+3)}$$

Luego

$$c_1^* = \begin{bmatrix} 0 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$c_1^{**} = \begin{bmatrix} 0 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$c_1 = \begin{bmatrix} c_1^* & c_1^{**} \end{bmatrix}$$

Con n=138, para estimar las observaciones

$$Zt = (Z_{1(138)}, \cdots, Z_{1(148)}, Z_{2(138)}, \cdots, Z_{2(148)})$$

Una restricción Y_2, asumiendo una incertidumbre conocida con $U = 0, \Sigma_U \neq 0$ y

$$Y_1 = -0.2Z_{1(n+2)} + 0.1Z_{1(n+3)} - 0.02Z_{2(n+2)} + 0.01Z_{2(n+3)}$$

Luego

$$c_2^* = \begin{bmatrix} 0 & -0.2 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$c_2^{**} = \begin{bmatrix} 0 & -0.02 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$c_2 = \begin{bmatrix} c_2^* & c_2^{**} \end{bmatrix}$$

Con n=180, para estimar las observaciones

$$Zt = (Z_{1(180)}, \cdots, Z_{1(189)}, Z_{2(180)}, \cdots, Z_{2(189)})$$

Dos restricciones de la forma $Y = (Y_1, Y_2)$, sin asumir incertidumbre, como en el caso 1.

$$Y_i = \begin{cases} -0.01Z_{1(n+2)} - 0.1Z_{2(n+1)} & \text{para } i = 1 \\ 0.1Z_{2(n+2)} & \text{i=2} \end{cases}$$

La recostrucción de los segmentos se hace aplicando el modelo ARIMA identificado, filtro de Kalman y la metodología de restricción.

La figura 4 presenta la estimación obtenida por las tres metodologías para el caso 1, en este caso el error cuadrático medio asociado a los pronósticos obtenidos con la restricción de c_1 resultó menor que los obtenidos con ARIMA y Kalman, sin embargo sus valores estimados se encuentran bastante alejados de los valores observado, de manera que no es suficiente que no rechaze que el horizonte de pronóstico mantenga la misma estructura del modelo identificado. Realmente es de gran importancia que exista una relación entre la restricción y el modelo que descrive el comportamiento de las series simultaneamente.

Figure 4: Pronósticos restringidos asociados a c_1 para Z_{1t}

En contraste, la restricción c_2(caso 2) en donde se asume la presencia de una incertidumbre conocida para la información adicional de la restricción, podemos darnos cuenta, que al igual que la restricció c_1 también tiene la agradable característica de presentar un error cuadratico medio pequeño, y menor con respeto al error cuadratico medio de ARIMA Y KALMAN, lo que se traduce en una mayor precisión de los pronósticos, esto debido a que la información adicional captura muy bien la relación con el modelo ARIMA que describe el comportamiento de las series.

Para el tercer caso se encontró, que mas de dos restricciones inducen a rechazar la hipótesis de conservación de la estructura del modelo en el horizonte de pronóstico, ya que la matriz de restricciones c se hace más sensible, cuando sus entradas son mayores a tres en valor absoluto, y el horizonte de pronóstico es muy grande, para este caso solo es posible calcular los pronósticos restringidos para un horizonte de tamaño dos, tres y cuatro. A partir del quinto, se rechaza la hipótesis de

Figure 5: Pronósticos restringidos asociados a c_1 para Z_{2t}

Figure 6: Pronósticos restringidos asociados a c_2 para Z_{1t}

conservación de la estructura del modelo (1). Siendo bastante complicado adaptar la restricción impuesta al horizonte de cinco pronósticos.

10 CONCLUSIONES

Repitiendo el mismo procedimiento de simulación, para procesos vectoriales $Z_t = (Z_{2t}, Z_{2t})$ que se ajustaron a modelos autorregresivos de orden uno y tres, variando la cantidad de restricciones y el horizonte de pronósticos, se encontró que:

> El éxito de la técnica de pronósticos restringidos en términos de precisión y confiabilidad, depende de que tan buena sea la relación entre la información disponible y el modelo ajustado. Además de no rechazar la prueba de

Figure 7: Pronósticos restringidos asociados a c_2 para Z_{2t}

compatibilidad, es importante evaluar hasta cuantos pronósticos se pueden pronósticar adecuadamente con la restricción, ya que esta se vuelve sensible después de un determinado horizonte de pronostico.

La predicción de los pronósticos restringidos puede ser engañosa, en la medida en que no se tenga en cuenta la condición anterior. Ya que el hecho de no rechazar la prueba de compatibilidad no implica que la información contenida en la restricción verdaderamente nos este conduciendo hacia la estimación adecuada, como el ejemplo de la restricción c_2 ya expuesto.

El hecho de aumentar el horizonte de pronósticos bajo una misma restricción, hace menos factible precisar la matriz c adecuada que no rechace la hipótesis de ruptura estructural, luego a mayor horizonte de pronósticos, mayor sensibilidad habrá en la matriz c, lo cual se ve a través de los fuertes rechazos por valores críticos extremadamente grandes que asume la prueba χ^2. Luego entre más simples sean las restricciones, mayor núemro de pronósticos podra calcularse.

Otro motivo que dificulta el establecimiento de la matriz c, es cuando el número de restricciones excede el orden del modelo autorregresivo, puesto que además de rechazar la hipótesis nula a verificar, en algunos casos genera sistemas singulares.

References

[1] Nieto F. (1989). Reconstruccion de una serie de tiempo censurada usando filtros de Kalman

[2] Guerrero, V. M. (1989). Optimal conditional ARIMA forecasts. Journal of Forecasting, 8, 215-229.

[3] Lutkepohtl H. (1991). Introduction to Multiple Time Series Análisis Berlin: Springer-Verlag.

[4] Guerrero, V. M. (1991). ARIMA forecasts with restrictions derived from a structuralchange. International Journal of Forecasting, 7, 339-347.

[5] Wei W. (2006). Time Series Analysis Univarite and Multivariate Methods. Canada United States:Addison Wesley

[6] Peña D. (2001). A Course in Time Serie Analisys. New York: Jhon Wiley & Sons

[7] Guerrero V, Senra E. (2001) Restricted Forecasting With a VEC Model: Validating the Feasibility of Economic Targets

[8] Castañeda M. (1994). Reconstrucción de Series de tiempo univariadas mediante el enfoque de pronósticos con restricciones

[9] Peter J. (1998). Introduction to Time Series and Forecasting New York Springer-Verlag.

[10] Tsay R, Peña D. (2000). Outliers in multivariate time series. Biometrika 4 789-804.

[11] Solera A. (2003). El filtro de kalman. Banco Central de Costa Rica

[12] Guerrero V, Gomez N. "Restricted VAR forecasts of economic time series with contemporaneous constraints". Documentos de Trabajo, Departamento de Estadística. ITAM, México

[13] Guerrero, V, Peña D. (2003). Combining multiple time series predictors: a useful inferential procedure. Journal of Statistical Planning and Inference, 116, 249-276.

[14] Guerrero V, Gomez N. (2006). Restricted forecasting with VAR models: Analysis of test for joint compatibility betwen restrictions and forecasts. International. Journal of Forecasting Vol 22 N° 4 pp 751-770.