

Agenda

- Motivación
- VAR y PLS
- VAR-PLS
- Ejemplo
- Resultados
- Conclusiones

Introducción

Abstract:

Phillip Hans Franses (2006) propuso una metodología para realizar pronósticos conjuntos de manera óptima a través de una representación autorregresiva de orden p , h pasos adelante. “Mínimos Cuadrados Parciales (PLS) Autorregresivo” ($PLSAR(h, p)$). Nosotros generalizamos esta propuesta para lo que llamaremos $PLSVAR(h, p)$

Este modelo se encuentra situado entre un $AR(p)$ que pronosticó todos los h pasos adelante y también para diferentes modelos AR en diferentes horizontes, de tal forma que formuló esos tres modelos para realizar h pronósticos, los cuales podemos representar de la siguiente manera:

$$AR(p): y_{T+h} = \hat{\mu} + \hat{p}_1 y_{T+h-1} + \hat{p}_2 y_{T+h-2} + \cdots + \hat{p}_p y_{T+h-p} \quad (1)$$

$$AR_h(p): y_{t+h} = \hat{\mu} + \hat{p}_{1,h} y_t + \hat{p}_{2,h} y_{t-1} + \cdots + \hat{p}_{p,h} y_{t-p} \quad (2)$$

$$PLSAR(h, p): \hat{Y} = XB_{PLS} \quad (3)$$

El modelo $AR(p)$ es la forma clásica de realizar los h pasos hacia adelante cuyos parámetros son estimados generalmente por Mínimos Cuadrados Ordinarios (OLS). El $AR_h(p)$ es una alternativa al caso anterior, esto debido a que OLS minimiza la suma cuadrada de $\hat{\varepsilon}_t$ pero no garantiza que lo sea para h errores a futuro. También, realizarlo de esta manera permite contar con diferentes modelos para diferentes horizontes de pronóstico y asimismo para series de tiempo estacionarias, el pronóstico de un $AR(p)$ converge rápidamente a la media incondicional (obviamente la rapidez depende directamente de $h \geq p$).

Para más detalles acerca de la literatura de este tipo de modelos, se puede consultar a Pesaran & Pick (2010), Marcellino, Stock & Watson (2004), Carreiro, Kapetorios & Marcellino (2010), Tiao & Xu (1993) entre muchos otros.

Ahora, es claro que existen correlaciones adyacentes entre las series de tiempo, la cual estos dos modelos no explotan. Dicho en otras palabras, se conoce que y_t y y_{t-j} están correlacionadas, más aún y_{T+h} y y_{T+h-j} también lo están, de esta manera, una alternativa viable es predecir conjuntamente $(y_{t+h}, y_{t+h-1}, y_{t+h-2}, \dots, y_{t+1})$ con la ayuda de $(y_t, y_{t-1}, y_{t-2}, \dots, y_{t-p})$, siendo así PLS una técnica atractiva dado la naturaleza de la problemática en que nos situamos (más adelante detallaremos cómo se realiza el procedimiento).

Así, Franses propone colocar la información de $(y_t, y_{t-1}, y_{t-2}, \dots, y_{t-p})$ en una matriz X (predictores) y $(y_{t+h}, y_{t+h-1}, y_{t+h-2}, \dots, y_{t+1})$ en Y (predichos), realizando finalmente el ejercicio de regresión a través de PLS, de tal manera que el proceso de construcción de variables latentes y cargas asociadas contengan la información relevante que tiene X en Y.

En el ejercicio empírico realizado compara los modelos (1), (2) y (3) y fue realizado para la serie del índice de producción industrial de Estados Unidos mediante dos procesos:

- Muestreo recursivo: Se realiza el pronóstico de horizonte h trabajando con $T_1 - j$, donde $j \geq h$ realizando $T_1 + h - j$. Así se reincorpora a $T_1 - j$ los $t = h$ datos originales y se realiza el siguiente $T_2 + h - j$ hasta así hasta que $T_n + h - j = T$, restimando el modelo en cada recursión.
- Campana de tiempo fija: Trabajar con una $T_i - j$ fija de tal manera que para cada pronóstico de horizonte h se reincorpore la $T_i + 1$ real y se pierda $t = 1$, estimando el modelo cada ocasión que se realice el pronóstico.

Para este ejemplo, se estimó un $AR(5)$ donde la componente $k = p$ obtenida por PLS resulta equivalente al $AR_h(5)$ (Hank & Friedman (1991)) de tal manera que el modelo PLS autorregresivo planteado fue un $PLSAR(5,4)$.

Los resultados mostraron que para la muestra recursiva, el $PLSAR(5,4)$ fue superior 4 veces que los otros dos modelos (la componente 1) y para la campana de tiempo fija, el $AR(5)$ fue mejor en 3 de los 5 horizontes de pronóstico, siendo el $PLSAR(5,4)$ superior en las restantes 2, esto considerando la raíz cuadrada del error cuadrático medio.

Lo anterior implica que un modelo PLS planteado como un fenómeno autorregresivo puede llegar a ser competitivo respecto a modelos clásicos existentes en la literatura.

Ahora, lo que proponemos en este trabajo son extensiones naturales que se devengan de lo realizado por Franses, las cuales resumimos en los siguientes puntos:

1. Extensión multivariada dada la flexibilidad de los modelos Vectores Autorregresivos (VAR), generando un modelo denominado VAR-PLS
2. Introducir a este concepto variables determinísticas (dummies, tendencias, etc.) y exógenas
3. Construcción de intervalos de predicción vía Bootstrap
4. Construir un modelo VAR con capacidad predictiva y asimismo pronosticar con el VAR-PLS observando el grado de competitividad que tiene éste respecto a un modelo enfocado a predecir.

Estos puntos serán desarrollados en los siguientes temas, no obstante en la siguiente sección se da una ligera revisión de los elementos importantes que rescataremos de un modelo VAR y la construcción del ejercicio de PLS.

En la siguiente sección se agruparán las similitudes que dan cabida al VAR-PLS, después explicaremos a detalle el ejemplo para el caso de la predicción del índice nacional de precios al consumidor a través de un planteamiento clásico de la fuente inflacionaria, explicando los resultados obtenidos y finalizando con las conclusiones.

Vectores Autorregresivos y Mínimos Cuadrados Parciales

Un proceso $VAR(p)$ es definido como:

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + C D_t + u_t \quad (4)$$

A_i : Matrices de coeficientes para $i = 1, 2, \dots, p$

ε_t : Proceso ruido blanco con matriz de covarianza positiva definida $E[u_t u_t'] = \Sigma_u$

C : Matriz de potenciales regresores determinísticos

D_t : Vector columna de regresores determinísticos apropiados

Notemos que el $VAR(p)$ puede ser definido como un $VAR(1)$ de la forma

$$Y_t = A Y_{t-1} + V_t \quad (5)$$

$$Y_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}$$

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}$$

$$V_t = \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Si los valores propios de A son menores que uno, entonces el proceso $VAR(p)$ es estable.

Una parte importante de la especificación del modelo está dada en encontrar el orden p del VAR a estimar. El procedimiento general consiste en ordenar los $p = 0, \dots, p_{max}$ y elegir el valor de p que minimiza algún criterio de selección. La selección del criterio para el $VAR(p)$ tienen la forma

$$IC(p) = \ln |\hat{\Sigma}(p)| + C_T \cdot \varphi(K, p) \quad (6)$$

Donde $\hat{\Sigma}(p) = T^{-1} \sum_{t=1}^T \hat{u}_t' \hat{u}_t$, C_T es una secuencia indexada por el número de realizaciones de T y $\varphi(K, p)$ es la función de penalización por la longitud del modelo $VAR(p)$. Los cuatro criterios

de información más utilizados son los criterios de Akaike (AIC), Schwarz-Bayesiano (BIC), Hannan-Quinn (HQ) y el Error Final de Predicción (FPE):

$$AIC(p) = |\hat{\Sigma}(p)| + \frac{2}{t} p K^2 \quad (7a)$$

$$BIC(p) = |\hat{\Sigma}(p)| + \frac{\ln T}{t} p K^2 \quad (7b)$$

$$HQ(p) = |\hat{\Sigma}(p)| + \frac{2\ln \ln T}{t} p K^2 \quad (7c)$$

$$FPE(p) = \left(\frac{T+p^*}{T-p^*} \right)^K \det(\hat{\Sigma}(p)) \quad (7d)$$

El criterio AIC sobreestima asintóticamente el orden con probabilidad positiva, mientras que el BIC y HQ estima consistentemente el orden bajo ciertas condiciones favorables si el verdadero valor de p es menor o igual que p_{max}

Notemos que al igual que el caso univariado, podemos con (4) pronosticar recursivamente de la siguiente manera:

$$y_{T+h|T} = A_1 y_{T+h-1} + \cdots + A_p y_{T+h-p} + CD_{T+h} \quad (8)$$

La estimación de A_i es generalmente realizada por OLS y haciendo $\text{vec}(\hat{A}) = \begin{bmatrix} \hat{A}_1 \\ \vdots \\ \hat{A}_p \end{bmatrix}$ y bajo ciertas condiciones de generalidad de comportamiento estacionario y ergodicidad en los modelos VAR (Hamilton (1994), Lutkepohl (1991) entre otros), $\text{vec}(\hat{A})$ es consistente y distribuido asintóticamente con matriz de covarianza:

$$\widehat{\text{avar}}(\text{vec}(\hat{A})) = \hat{\Sigma} \otimes (Z'Z)^{-1} \quad (9)$$

Donde

$$\hat{\Sigma} = \frac{\sum_{t=1}^T \hat{\varepsilon}_t' \hat{\varepsilon}_t}{T-K} \quad (10)$$

$\hat{\varepsilon}_t = Y_t - \hat{A}' Z_t = Y_t - \hat{A}' Y_{t-1}$ es el residual de mínimos cuadrados en el tiempo t

El i -ésimo elemento de $\text{vec}(\hat{A})$ es asintóticamente normalmente distribuido con errores estándar dados por la raíz cuadrada de los elementos de la diagonal de $\hat{\Sigma} \otimes (Z'Z)^{-1}$. Entonces, las pruebas t son válidas asintóticamente para los coeficientes estimados.

Una situación de interés radica en la presencia de una o más raíces unitarias en las y_j . Esto ha dado pie a toda una teoría económica, que se basa en la conceptualización de modelar el comportamiento de largo plazo y analizar la dinámica temporal de un conjunto de series. En otras palabras, el concepto de cointegración y corrección del error.

Definamos estos dos conceptos importantes que formalizan las ideas expresadas anteriormente:

Cointegración: Las componentes del vector y_t se dice que son cointegradas de orden d, b , denotado por $y_t \sim CI(d, b)$ si a) todos los componentes de y_t son $I(d)$; y b) existe un vector $\beta (\neq 0)$ tal que $z_t = \beta'y_t \sim I(d - b)$, $b > 0$. El vector β es llamado vector de cointegración

Corrección de error: El vector bivariado $y_t = (y_{1t}, y_{2t})'$ con vector de cointegración $\beta = (1, -\beta_2)'$, entonces $\beta'y_t = y_{1t} - \beta_2 y_{2t} \sim I(0)$ existe una representación de corrección de error de

$$\Delta y_{1t} = \alpha_1 + \gamma_1(y_{1t-1} - \beta_2 y_{2t-1}) + \sum_{i=1}^K \psi_{1,i} \Delta y_{1t-i} + \sum_{i=1}^K \psi_{2,i} \Delta y_{2t-i} + u_{1t} \quad (11a)$$

$$\Delta y_{2t} = \alpha_2 + \gamma_2(y_{1t-1} - \beta_2 y_{2t-1}) + \sum_{i=1}^L \xi_{1,i} \Delta y_{1t-i} + \sum_{i=1}^L \xi_{2,i} \Delta y_{2t-i} + u_{2t} \quad (11b)$$

Notemos que el modelo $VAR(p)$ puede expresarse como:

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + CD_t + u_t \quad (12)$$

$$\begin{aligned} \Gamma_i &= -(A_{i+1} + \dots + A_p), \quad i = 1, \dots, p-1 \\ \Pi &= -(I - A_1 - A_2 - \dots - A_p) \end{aligned}$$

Y asimismo:

$$\Delta y_t = \Pi y_{t-p} + \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + CD_t + u_t \quad (13)$$

$$\begin{aligned} \Gamma_i &= -(I - A_1 - A_2 - \dots - A_i), \quad i = 1, \dots, p-1 \\ \Pi &= -(I - A_1 - A_2 - \dots - A_p) \end{aligned}$$

La primera representación es conocida como Modelo de Vector de Corrección de Error (VECM) transitorio y la segunda VECM de largo plazo.

De esta manera, la matriz Π tiene los siguientes puntos de interés:

1. $rk(\Pi) = n$, todas las n combinaciones deben ser estacionarias; el VECM es un modelo VAR en niveles
2. $rk(\Pi) = 0$, no existe una combinación lineal estacionaria tal que Πy_{t-1} sea estacionaria, excepto la solución trivial, i.e. es un modelo $VAR(p-1)$ en primeras diferencias
3. $0 < rk(\Pi) < n$, en este caso $\Pi = \alpha\beta'$ con dimensión (nxr) y $\beta'y_{t-1}$ es estacionaria. Cada columna de β representa una relación de largo plazo

Si el objetivo es pronosticar, aún en el caso de variables integradas y cointegradas, hacerlo mediante la representación VAR es bastante conveniente (ver Lutkepohl 2006) y dado que el objetivo de este trabajo es obtener un método alternativo de pronóstico, especificaremos un planteamiento apropiado de predicción, no sin antes detallar en qué consiste la técnica de PLS.

Mínimos Cuadrados Parciales

El método de PLS puede ver desde diferentes ópticas y desde la perspectiva de interés en este trabajo, podemos relacionar la expresión del Vector Autorregresivo (expresión 5) con la forma de un modelo lineal generalizado de la forma clásica:

$$Y = XB + U \quad (14)$$

Donde Y es $N \times k$, X es una matriz $n \times N$, B es una matriz de $(N+1) \times k$ y U es una matriz de $N \times k$. Consideremos Y la matriz de variables dependientes, X de independientes, B los parámetros que establecen la relación entre X y Y y U como el término de error.

Aunado a esto y bajo la concepción que plantea Franses, de explotar las correlaciones existentes entre $(y_t, y_{t-1}, y_{t-2}, \dots, y_{t-p})$ y $(y_{t+h}, y_{t+h-1}, y_{t+h-2}, \dots, y_{t+1})$, la diferencia entre (5) y (14) radicará únicamente que mientras en (5) lo hacemos a través de OLS en (14) a través PLS, lo que lleva consigo un proceso de descomposición que maximiza la covarianza existente una matriz $X = Y'_{t-1}$ con $Y = Y'_t$

De esta forma, el procedimiento básico consiste en maximizar $\text{cov}(X\alpha, Y\beta)^2$ s. a.

$$\alpha'(S^*_{xx} + \lambda_x)\alpha = 1 \quad \& \quad \beta'(S^*_{yy} + \lambda_y)\beta = 1 \quad (15)$$

Dónde:

$$S^*_{xx} = (1 - \lambda_x)S_{xx} \quad \& \quad S^*_{yy} = (1 - \lambda_y)S_{yy}$$

Nótese que $(X\alpha, Y\beta)$ representa una combinación lineal de las variables pudiendo ser más o menos importante alguna de éstas (generalidad) y maximizar la covarianza o covarianza al cuadrado es indistinto pues no interesa el signo sino la maximización. Finalmente S_{xx} y S_{yy} son las matrices muestrales de varianza y covarianza.

Notemos que si $\lambda_x = 0$ y $\lambda_y = 1$ entonces se asegura que cada una de las direcciones encontradas está descorrelacionadas y:

$$\beta'\beta = 1 \quad \& \quad \alpha'S_{xx}\alpha = 1$$

Siendo ésta la versión que tradicionalmente se maneja en la literatura

Alternativas a estos valores dan soluciones a problemas específicos de PLS o Correlación Canónica (CCA)

Continuando con el problema, maximizando (14) obtenemos:

$$\mathcal{L} = (\alpha'S_{xy}\beta)^2 - \gamma\{\alpha'(S^*_{xx} + \lambda_x)\alpha - 1\} - \mu\{\beta'(S^*_{yy} + \lambda_y)\beta - 1\}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = 2S_{xy}\beta - 2\gamma(S^*_{xx} + \lambda_x)\alpha = 0 \rightarrow S_{xy}\beta = \gamma(S^*_{xx} + \lambda_x)\alpha$$

$$\frac{\partial \mathcal{L}}{\partial \beta} = 2S_{yx}\alpha - 2\mu(S^*_{yy} + \lambda_y)\beta = 0 \rightarrow S_{yx}\alpha = \mu(S^*_{yy} + \lambda_y)\beta$$

Multiplicando por α' la primera ecuación y por β' la segunda tenemos que:

$$\alpha'S_{xy}\beta = \gamma\alpha'(S^*_{xx} + \lambda_x)\alpha = \gamma$$

$$\beta'S_{yx}\alpha = \gamma\beta'(S^*_{yy} + \lambda_y)\beta = \mu$$

Dado que las expresiones anteriores tienen las mismas componentes entonces $\gamma = \mu$, por lo que haciendo esto evaluando en la segunda ecuación de primer orden tenemos que:

$$(S^*_{yy} + \lambda_y)^{-1} \gamma^{-1} S_{yx} \alpha = \beta$$

Y lo anterior a la primera ecuación de primer orden:

$$\begin{aligned} S_{xy}(S^*_{yy} + \lambda_y)^{-1} \gamma^{-1} S_{yx} \alpha &= \gamma(S^*_{xx} + \lambda_x) \alpha \\ (S^*_{xx} + \lambda_x)^{-1} S_{xy}(S^*_{yy} + \lambda_y)^{-1} \gamma^{-1} S_{yx} \alpha &= \gamma^2 \alpha \end{aligned}$$

Evaluando en términos de S_{xx} y S_{yy} tenemos que:

$$\{(1 - \lambda_x)S_{xx} + \lambda_x\}^{-1} S_{xy} \{(1 - \lambda_y)S_{yy} + \lambda_y\}^{-1} S_{yx} \alpha = \lambda w$$

Donde λ son los valores propios asociados a la matriz y w el vector propio

$$\{(1 - \lambda_x)S_{xx} + \lambda_x\}^{-1} S_{xy} \{(1 - \lambda_y)S_{yy} + \lambda_y\}^{-1} S_{yx}$$

Similarmente para β obtenemos que:

$$\{(1 - \lambda_y)S_{yy} + \lambda_y\}^{-1} S_{yx} \{(1 - \lambda_x)S_{xx} + \lambda_x\}^{-1} S_{xy} \beta = \lambda q$$

De esta manera obtenemos los scores para X y Y :

$$\begin{aligned} t &= Xw = Ew \\ u &= Yq = Fq \end{aligned}$$

Normalizando los scores $t = t / \sqrt{t't'}$

Los scores de Y no son necesarios para realizar la regresión, pero generalmente se guardan para fines interpretativos. De esta manera E y F son iniciadas con X y Y respectivamente. De tal forma obtenemos los loadings para X y Y :

$$\begin{aligned} p &= E't \\ q &= F't \end{aligned}$$

Con esto las matrices son “desinfladas” quitando el efecto relacionado a la variable latente anterior como:

$$\begin{aligned} E_{n+1} &= E_n - tp' \\ F_{n+1} &= F_n - tpq' \end{aligned}$$

Este proceso se repite hasta que E_{n+1} y $F_{n+1} = 0$

Agrupamos cada w, t, p y q en matrices obtenemos:

$$R = W(P'W)^{-1}$$

Así realizamos la regresión de Y en X , usamos los scores T calculando los coeficientes de regresión como en (14) pre-multiplicando con la matriz R ($T = XR$) obteniendo finalmente:

$$Y = XB + U = XB_{PLS} + U \quad (16)$$

Donde $B_{PLS} = R(T'T)^{-1}T'Y = RQ'$

En la siguiente sección agruparemos los conceptos del VAR y de PLS, así como la idea fundamental de Franses para dar cabida al VAR-PLS con fines predictivos.

VAR-PLS

Al igual cómo lo propone Franses trabajaremos con matrices en estructura de rezagos según cada una de las variables, sin embargo, en nuestro caso agruparemos en la matriz X la

información observada en el pasado como en $X = Y_{t-1} = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{bmatrix}$ y en Y lo observado en tiempo

t $Y = Y_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}$ estimando el modelo (16) a través del proceso de construcción de variables latentes. Adicionalmente en nuestro trabajo se pueden introducir efectos de variables exógenas

o determinísticas colocadas en C , modificando la composición de $X = Y_{t-1}^* = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \\ D_t \end{bmatrix}$ y por ende

de la matriz de coeficientes $A^* = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p & C \\ I & 0 & \cdots & 0 & 0 & 0 \\ 0 & I & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 & 0 \end{bmatrix}$

Finalmente proponemos utilizar la representación que da cabida a un modelo VAR, de tal manera que sea el proceso generador de datos y la en sí, la metodología VAR, la que dictamine el orden del proceso autorregresivo que será utilizado en la regresión PLS, lo anterior para un modelo que tiene como fuente de inflación variables tradicionales en la economía.

Ejemplo

El modelo considerará, cómo fuente principal de la dinámica inflacionaria mexicana, el crecimiento o variación de las condiciones monetarias del país. Al respecto existe una gran discusión, incluso hoy en día, sobre si existe una relación de largo plazo entre el fenómeno monetario y el traspaso inflacionario que tiene éste, no obstante, tiende a haber un consenso de que la inflación, en el largo plazo, es un fenómeno netamente monetario.

Por citar tan sólo un ejemplo, Woodford (2007) argumenta que no es apropiado considerar, bajo una concepción Neo Keynesiana, el crecimiento monetario y su relación con la inflación ya que afirma que son sólo las tasas de interés las que juegan un roll en la dinámica inflacionaria. Sin embargo Nelson (2008) realiza una interpretación distinta a esta argumentación, indicando que aunque el objetivo inflacionario está ligado a las tasas de interés es producto de la política que toma el banco central para controlarlas y esto es precisamente intervención en el mercado de dinero, es decir, crecimiento monetario.

Claro está, en este trabajo no abordaremos tal discusión, sino que se mostraremos las propiedades empíricas del modelo desarrollado en términos del error de predicción fuera de muestra.

Para ello, proponemos utilizar las siguientes variables de enero de 2000 a febrero de 2012:

p: El Índice Nacional de Precios al Consumidor

m0: Billetes y monedas en circulación

r: tasa de interés interbancaria a 28 días

y: Indicador Global de la Actividad Económica

La idea fundamental es relacionar empíricamente la variable de precios, que a su vez es una función de la tasa de inflación (inflación mensual, inflación interanual, acumulada, etc.) con el resto de las variables permitiendo las relaciones multivariadas existentes, generando así un $VAR(p)$ de rango completo y/o un $VECM$ para el caso cointegrado.

Como el objetivo es construir un modelo de pronóstico robusto que a su vez sea comparado con la técnica de VAR-PLS se realiza el siguiente procedimiento:

- Se eliminaron las 24 observaciones finales con el fin de obtener una campana de tiempo de largo horizonte que sea comparada según cada uno de los modelos.
- Para el VAR-PLS, se realiza el procedimiento tradicional de seleccionar el rezago óptimo a través de algún criterio de información, en este caso a fue el Error de Final de Predicción (FPE) y para observar que tal proceso generador de datos sea consistente con la teoría económica, se estima el $VAR(p)$ denotando su característica estocástica, es decir, si es cointegrado o no, obteniendo en su caso, los coeficientes de largo plazo.
- Posteriormente, para ese p óptimo con una especificación constante y sin variables determinísticas, se procede a estimar el VAR-PLS con $h = 24$, computando el error fuera de muestra para cada una de las componentes existentes $pK = 48$
- Para el modelo $VAR(p)$, se tienen 4 variables, 4 posibles especificaciones del VAR (ninguna, constante, tendencia, constante y tendencia), 11 posibles variables dummies estacionales (una para cada mes). Combinado todas de las variables, con las

especificaciones y variables determinísticas posibles nos da un total de $VAR_j(p)$ $j = 1, \dots, 484$ posibles a estimar.

- Se estiman cada uno de los modelos y se seleccionan los que minimizaron 7 criterios de error fuera de muestra, es decir, para esto se eliminaron otras $h = 24$ observaciones, con el fin computar los estadísticos de error (Hyndman & Koehler 2006):
 - MAPE: Media del porcentaje de error absoluto
 - MdAPE: Mediana del porcentaje de error absoluto
 - RMSPE: Media de la raíz cuadrada del porcentaje de error al cuadrado
 - RMdSPE: Mediana de la raíz cuadrada del porcentaje de error al cuadrado
- Adicionalmente se trabajó con un modelo benchmark (Autorregresivo de orden 1) para el cual se calcularon los $i = 1, \dots, h$ ($h = 24$) pronósticos generando el siguiente estadístico:

$$\text{test} = \frac{Y_{t+i} - Y_{t+i, VARj(p)}^f}{Y_{t+i} - Y_{t+i, AR(1)}^f}$$

Posteriormente los siguientes 3 estadísticos:

- MRAE: Media absoluta del estadístico test
- MdRAE: Mediana absoluta del estadístico test
- GMRAE: Media geométrica absoluta del estadístico test
- De esta manera se obtienen 7 modelos, los cuales se integran en un solo número obtenido del cuantil 50% para cada uno de los horizontes de pronóstico.

Es importante señalar que para el VAR-PLS podemos obtener hasta pK componentes, y para este ejercicio se estiman cada una de ellas sólo para fines comparativos, puesto que es uno de los objetivos de este trabajo observar la capacidad predictiva respecto a métodos recursivos de pronóstico como el que se utiliza con el modelo VAR siendo indistinta la componente seleccionada.

Intervalo de predicción: VAR-PLS

Uno de los objetivos de este trabajo radica en la construcción del intervalo de predicción para el modelo VAR-PLS. Vale la pena destacar que el procedimiento para su realización es similar a lo propuesto por Pascual, Ruíz y Fresoli (2011), Kim (2001), entre otros, cuyo procedimiento debe de adecuarse según lo que implica el cálculo de las matrices de coeficientes de regresión a través de PLS.

El procedimiento lo resumimos a continuación:

1. Estimamos $Y_t = X_t \hat{\beta}_{PLS} + \hat{U}_t$
2. Realizamos $\hat{U}_t^* = \hat{U}_t - \bar{\hat{U}}_t$ para un orden aleatorio de tales residuos centrados
3. Con los p valores iniciales $Y_0 = \{Y_p, \dots, Y_1\}$ y con lo obtenido en 1) y 2) generamos Y_t^* a través de la siguiente manera:

$$Y_t^* = X_t \hat{\beta}_{PLS} + \hat{U}_t^* \quad t = 1, \dots, n - p \tag{17}$$

4. Obtenemos \hat{Y}_{T+h}^* con (17) repitiendo el paso 2 a 4 $n = 1, \dots, N$ ocasiones
5. Finalmente, para cada una de las n^{th} variables y los N conjunto de pronósticos obtenemos:

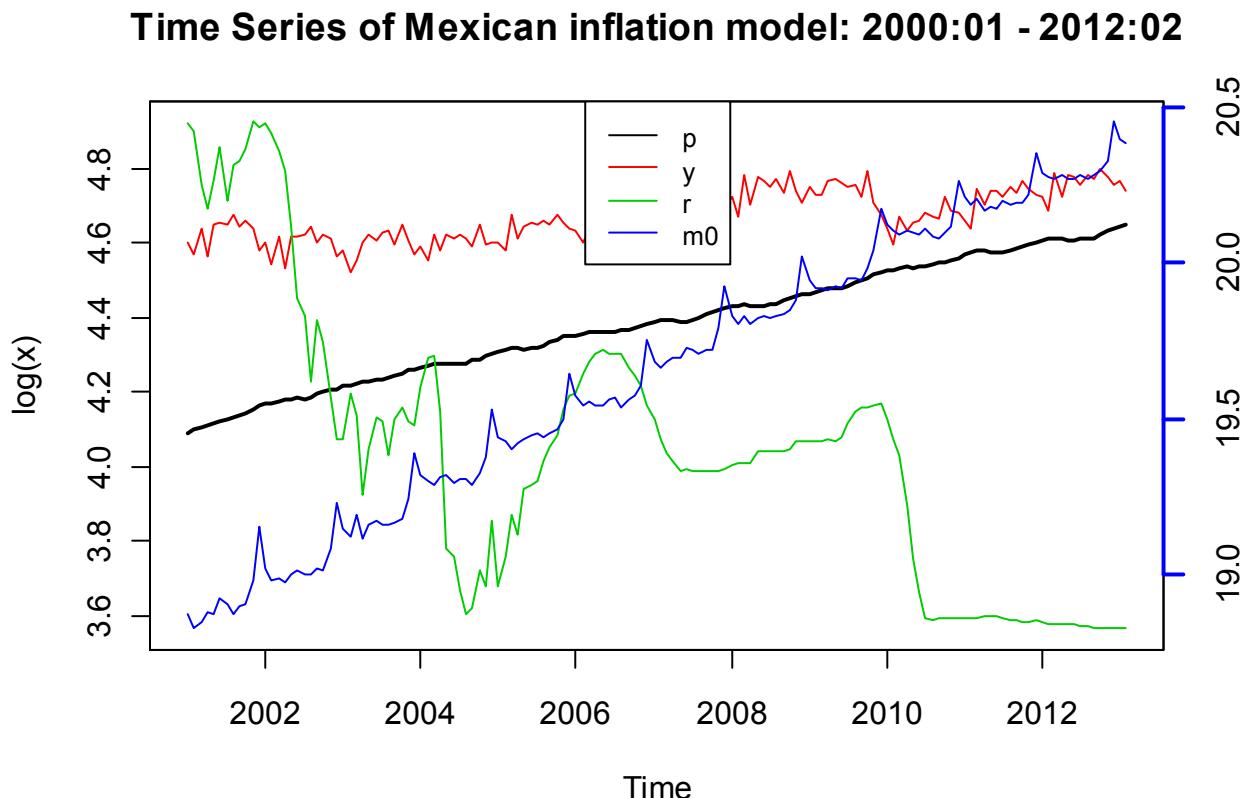
$$CI_{T+h} = \{y_{n,T+k} | y_{n,T+k} \in [q_B^*(\tau), q_B^*(\tau - 1)]\}$$

Donde $q_B^*(\tau)$ es el τ^{th} percentil de $G_{n,B}^*(x) = \#(y_{n,T+k}^{*(b)} \leq x) / N$

Resultados

El primer paso fue especificar el modelo VAR y por ende el VAR-PLS apropiado, de tal manera que se analizará la naturaleza existente entre las series de tiempo. En el siguiente gráfico se muestran de manera descriptiva los principales comportamientos.

[Gráfico 1]



Podemos apreciar que la serie de precios lleva una clara tendencia creciente en niveles, el índice monetario presenta una estacionalidad característica en todo el periodo de tiempo, la producción económica con tendencia de menor pronunciación que la serie de precios la cual que a partir del 2010 exhibe cierta recuperación respecto a los niveles observados en la primera mitad del gráfico. La tasa de interés claramente ha tenido un periodo de estabilidad a partir del segundo semestre del 2010.

De esta manera, se especificó el orden p del VAR teórico a través del criterio de Error Final de Predicción el cual fue de 12, asimismo, se realizó la prueba de Johansen para denotar la presencia o no de relaciones de largo plazo.

Los resultados obtenidos mostraron que al 1%, 5% y 10% existe una ecuación de cointegración, la cual está dada por la siguiente expresión:

$$p + 9.68 - 0.43m - 0.89y + 0.10r = 0$$

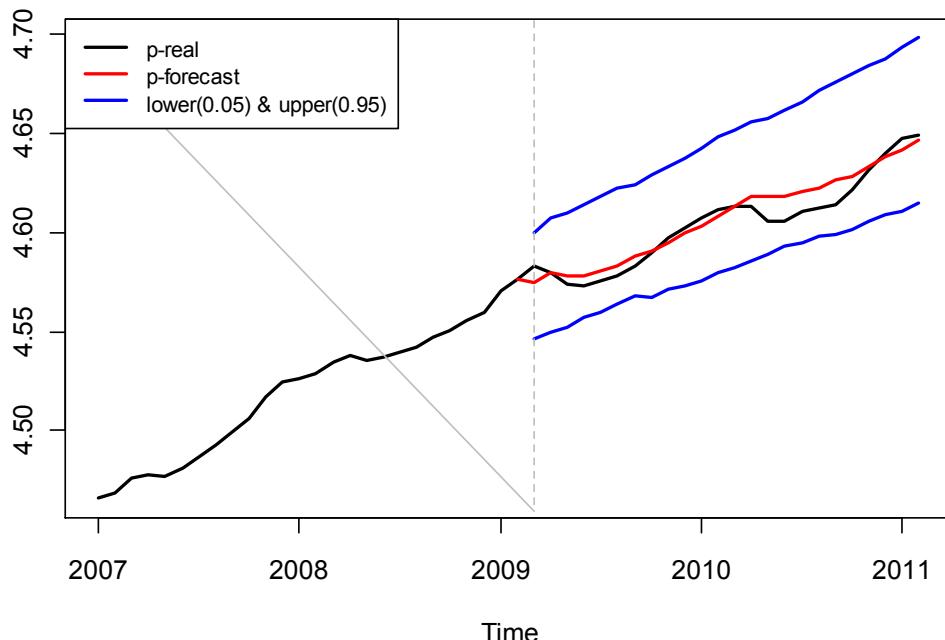
Lo anterior es congruente con la realidad ya que establece el traspaso inflacionario está impulsado por el crecimiento monetario, exceso de demanda y la reducción del costo del dinero. No obstante existe un debate teórico en relación a la especificación que analiza la inflación desde una perspectiva monetaria, los resultados son coherentes y se expresará que tal relación, dada por la estructura de rezagos, es apropiada desde una perspectiva estadística como econométrica.

De esta manera se estima el $VAR(12) - PLS(24,j)$ y se realizan los $pK = 48$ modelos, computando para cada serie el $MAPE$ para fines comparativos para un horizonte de pronóstico de 24 datos.

Vale la pena destacar que el número de componentes que maximizan la variabilidad, podría ser 2 (89.78%) y con 15 prácticamente se explica el 100% de la variación total. Para fines visuales, utilizamos tal componente para visualizar el comportamiento de pronóstico respecto al observado real de precios.

[Gráfico 2]

Forecast series p: VAR(12)-PLS(h=24,k=15)



Podemos observar que prácticamente, el valor real es el mismo que el pronóstico obtenido con un error absoluto promedio de 0.11%. Se observan los intervalos de predicción vía Bootstrap que “atrapan” en cada uno de los casos el valor real de la serie.

Para términos netamente económicos, ésta aproximación puede resultar competitiva si nuestro objetivo es pronosticar la inflación mexicana, teniendo como precios, el caso más generalizado el comparativo de inflación que se deseé realizar.

Al ser un modelo multivariado se obtienen el resto de los pronósticos de las series, sin embargo dado que el objetivo es predecir la serie de precios, no se considera relevante para este trabajo observar a detalle el comportamiento, no obstante es válido comentar que pronosticó (con esta misma componente) **con un error de 0.23% para m0, para y 0.83% y finalmente con un 5.84% la tasa de interés.**

Para el Integral-VAR, los $VAR_j(p)$ óptimos fueron los siguientes:

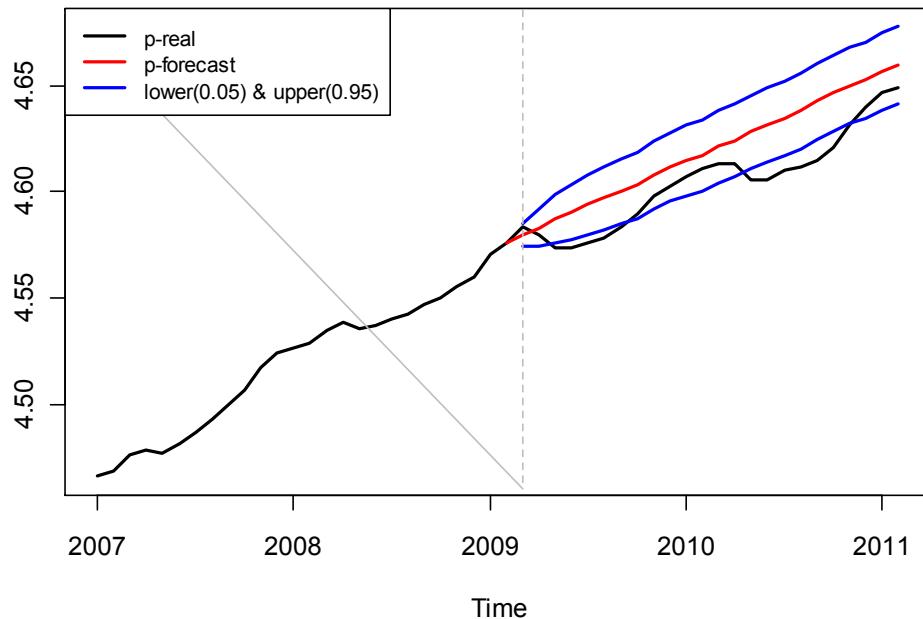
[Cuadro 1:] Modelos VAR óptimos

Criterio	MAPE	MdAPE	RMSPE	RMdSPE	MRAE	MdRAE	GMRAE
Estadístico	0.29	0.30	0.33	0.30	0.24	0.22	0.18
Rezagos	2	2	2	2	2	2	3
Estacionalidad	10	10	10	10	4	11	7
Especificación	Constante y tendencia						

Integrando los pronósticos y el intervalo de predicción con su respectivo cuantil del 50% para cada modelo y horizonte de pronóstico obtenemos lo siguiente:

[Gráfico 3]

Forecast series p: Integral-VAR



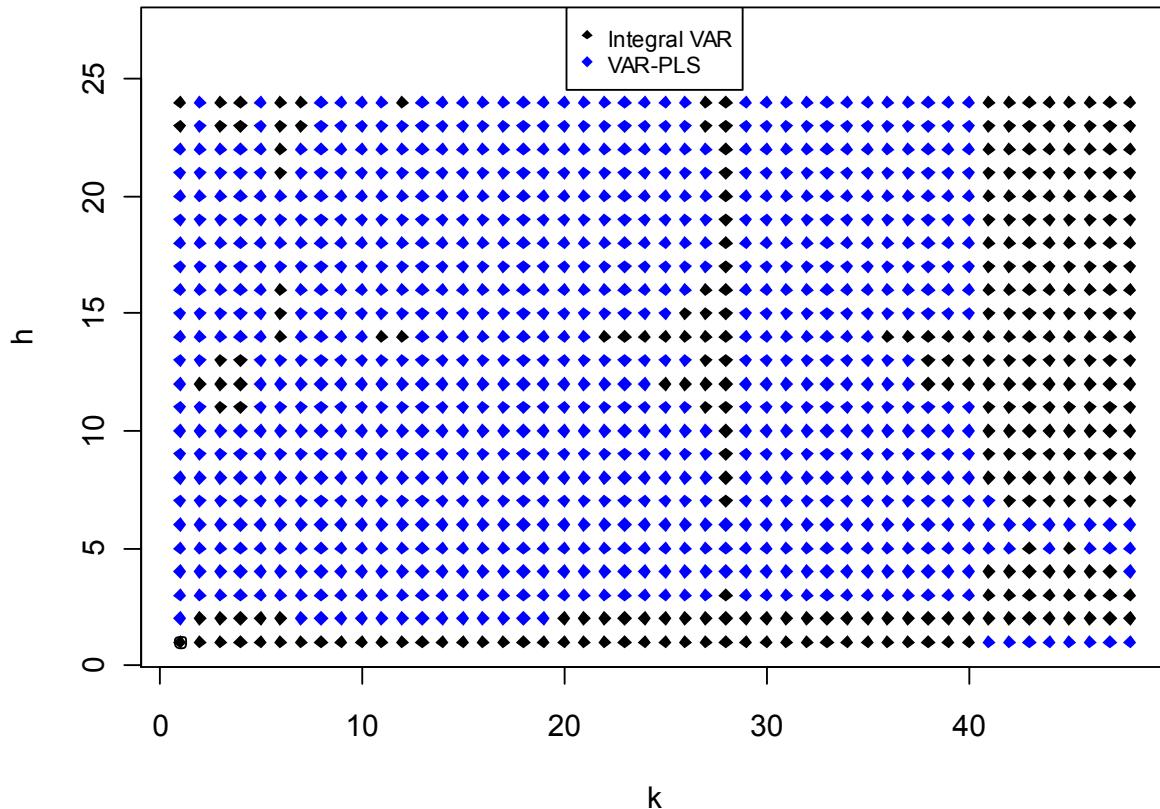
Aunque en términos numéricos el error fue de 0.32%, a detalle se observan grandes diferencias respecto al real observado, sobreestimando la inflación en cada uno de los casos. Lo anterior puede deberse a que alguna de las variables puede estar incrementando el error, esto se prevé dado que incluso para el modelo óptimo “pasado”, el MAPE de construcción fue de 0.29%, **más grande que el real observado en el PLS-VAR**, situación que podría llevar a replantear la especificación haciendo variar las variables que se pronostican en conjunto con los precios, situación recurrente en la práctica.

No obstante, se decidió realizarlo de esta manera para hacer comparativo las variables que intervienen en los pronósticos para ambas especificaciones.

El siguiente gráfico muestra cómo se comportó **cada componente respecto al Integral VAR**:

[Gráfico 4]

Integral VAR vs VAR-PLS (Comparasion with MAPE)



Resulta muy interesante que en promedio, **un 73.52% de las veces, fueron superior las componentes** y hace sentido que las componentes más lejanas dejan de tener efectividad dado a que son éstas las que tienen menor explicación de la variabilidad observada.

En otras palabras, la metodología VAR-PLS parece ser atractiva respecto su competidor inmediato, **que es un VAR enfocado a predecir**.

Conclusiones

En este trabajo se presentó una metodología alterna para realizar pronósticos multivariados mediante la perspectiva de explotar un conjunto de series de tiempo y_{t+j} con y_t , aprovechando la naturaleza que tiene por construcción la técnica de PLS, específicamente al momento de plantear un modelo lineal generalizado, situación que se presenta para el caso de un VAR.

Para este caso en particular, se planteó un modelo que considera como principales fuentes inflacionarias un agregado monetario (billetes y monedas), tasas de interés y una variable que ingreso y se realizaron dos metodologías de pronóstico; la primera de ellas un VAR-PLS cuyo orden p está dado por el criterio de FPE de un VAR, estimando en este caso, todas y cada una de las pK componentes posibles. Finalmente, los intervalos de predicción se hicieron vía

Bootstrap. El objetivo es pronosticar $h = 24$, evaluando el porcentaje de error al finalizar el ejercicio.

El pronóstico con la metodología VAR se realizó como un proceso recursivo variando la estacionalidad y la especificación del VAR, estimando 484 modelos y seleccionando los 7 que optimizaron criterios específicos; integrando finalmente los pronósticos, denominando a tal integración como pronóstico el pronóstico de un *Integral VAR*

Para el caso del VAR-PLS se trabajó con $T - h$ datos, para el caso de la selección de los VAR con $T - 2h$, estimando hasta $T - h$ para posteriormente, evaluar los modelos hasta $T + j - h$, para $j = 1, 2, \dots, h$.

Los resultados obtenidos mostraron que el orden autorregresivo del VAR-PLS fue de 12 y con 15 componentes (explicamos hasta casi el 100% de la variabilidad) el error de pronóstico, escogiendo esta última para realizar los pronósticos, obteniendo para la serie de precios un error de 0.11% y para el Integral VAR un 0.32%.

Ahora, visualizando cada una de las componentes, a partir de la componente 40 el Integral VAR tiene mejor comportamiento, pero para el resto el VAR-PLS es más certero, esto representa en promedio un 73.52% de las veces.

Lo anterior representa un punto de partida que indica una posibilidad latente de que el VAR-PLS es una técnica atractiva de pronóstico multivariado, ya que fue puesto a competir con un modelo recursivo que fuera de muestra, optimizó diversos criterios, siendo indistinta casi en su totalidad la componente seleccionada que se utilizó para comparar el error de pronóstico.

Quedan líneas futuras a trabajar como el incluir en el ejercicio de cointegración, la técnica de PLS, denotando las regularidades teóricas que esto implica, así como dentro de la perspectiva empírica, podemos construir modelos PLS-VAR que integren cada una de las componentes y/o poner en competencia la metodología VAR-PLS con VAR que mezclen las variables que intervienen dentro de un pronóstico.