



CIMAT

Centro de Investigación en Matemáticas, A.C.

PRONÓSTICOS VÍA VAR-PLS

T E S I S

Que para obtener el grado de
Maestro en Cómputo Estadístico

Presenta

José Antonio García Ramírez

Director de Tesis:

Dra. Graciela González Farías

Autorización de la versión final

Monterrey, N.L., 18 de Enero de 2020

RESUMEN

En la actualidad los pronósticos de series de tiempo proporcionan información que es de ayuda para la toma de decisiones. Las series de tiempo son un tipo de dato que posee distintas propiedades dada la estructura de correlación que posee en el tiempo, que aunado a la interacción con otras variables enriquecen el modelado estadístico.

Algunos de los modelos estadísticos para realizar pronósticos precisos, como los modelos de series de tiempo de alta dimensionalidad, requieren de técnicas computacionales desarrolladas recientemente, entre las cuales se encuentran: técnicas de reducción de dimensionalidad y los métodos de remuestreo; que permiten en términos de complejidad computacional, la estimación de pronósticos usando una mayor cantidad de información.

El presente trabajo propone la metodología VAR-PLS para realizar pronósticos conjuntos hasta h pasos hacia adelante. La metodología parte de un Vector Autorregresivo (VAR) cointegrado y realiza el pronóstico por medio de Mínimos Cuadrados Parciales (PLS) utilizando varios rezagos. Posteriormente se construyen intervalos de confianza de los pronósticos por medio de una adaptación para series de tiempo del método Bootstrap.

Se ilustra la metodología con un ejercicio empírico con otras 16 series de tiempo macroeconómicas para pronosticar el INPC mexicano. Los resultados son comparados con los pronósticos obtenidos a través de un modelo VAR con el fin de de comparar la precisión de la metodología propuesta respecto a un modelo usado comúnmente en la literatura.

El resultado más importante obtenido es que se lograron realizar pronósticos del INPC 12 pasos adelante con una precisión promedio mayor al 99 %.

A mis padres Claudia y Nicolás, por su apoyo incondicional y su enorme esfuerzo al educarme, y a mis hermanos quienes alegran mi vida con sus cariños. A Dios por mi familia y esta vida.

Agradecimientos

Agradezco a la vida por permitirme conocer a la Mtra. Alma Luz quien me indujo en la educación básica y preparatoria hacia las matemáticas. Luego en la facultad de ciencias al Mtro. Lara Aparicio por mostrarme de qué se trata la matemática aplicada. Y más recientemente, desde Acatlán y durante la maestría, doy gracias al Dr. Arturo Erdely por mostrarme que la estadística y el cómputo convergen, aunado a su apoyo y amistad.

En la bella ciudad de Monterrey quiero reconocer a mi asesora y a mi tutor, la Dra. Graciela González Farías y el Dr. Rodrigo Macías, cuya afectuosa y rigurosa guía me permitieron concretar mis estudios de posgrado incluyendo la participación en varios eventos dentro y fuera del país.

Gracias a la unidad Monterrey de CIMAT y quienes trabajan en ella, por su esfuerzo que me permitió seguir con alegría y entusiasmo este desafío.

Quiero reconocer el apoyo de mi país y en particular a CONACYT por la beca que me otorgó para la realización de este programa de maestría con número de becario: 635823 y número de CVU: 859869, además al programa CB-2016-252996 cuyo apoyo me permitió asistir a la universidad de Washington en el verano del 2018.

Finalmente, pero no menos importante, quiero agradecer a mis amigos y compañeros de licenciatura y maestría cuya alegría me permitió pasar los días difíciles: César, Eduardo, Edison, Erick, Hairo, ...

Contenido

1. Introducción	1
2. Fundamentos teóricos	3
2.1. Series de tiempo	3
2.1.1. Estacionariedad y estacionariedad estricta	4
2.1.1.1. Ergodicidad	6
2.1.2. Modelos AR	6
2.1.3. Raíces unitarias	7
2.2. Modelos VAR	8
2.2.1. Contexto del uso de modelos VAR en econometría	12
2.2.2. Cointegración	12
2.2.3. Modelos VEC	13
2.3. Reducción de dimensionalidad y componentes principales supervisados	15
2.3.1. PCA	15
2.3.2. PLS	17
3. Metodología VAR-PLS	23
3.1. Estimación de los h pronósticos conjuntos	25
3.2. Construcción de los intervalos de confianza	26
4. Resultados	28
4.1. Los datos	28
4.2. Aplicación empírica	32
5. Conclusiones	38
Lista de Figuras	39
Lista de Tablas	40
Bibliografía	41

Introducción

Es conocido que en la actualidad, los organismos encargados de recolectar información económica y financiera han ido incrementando tanto el número de observaciones como la cantidad de variables, lo cual conlleva a enfrentar un nuevo paradigma de modelación estadística: ¿cómo aprovechar esa gran cantidad información en beneficio de la sociedad?

Es razonable pensar que utilizando gran cantidad de información, podemos tener mayor certeza sobre el futuro de diversos acontecimientos económicos y financieros. De esta forma, realizar pronósticos de series de tiempo aprovechando la vasta información existente actualmente, puede ayudar a prever crisis económicas, repuntes en la economía, nuevas tendencias, etc.

Para efectuar pronósticos se suele recurrir a modelos estadísticos y econométricos, en particular a modelos estocásticos, ya que se requiere de predicciones que involucren incertidumbre sobre el futuro del hecho a modelar. Estos métodos permiten medir, inclusive, la confianza que puede tenerse en el pronóstico.

Los pronósticos pueden realizarse de manera individual, considerando el pasado de una variable más posibles eventos exógenos, o bien, de manera multivariada, donde los h pronósticos se realizan de manera conjunta tomando en cuenta las diversas variables que integran al modelo.

El objetivo principal del presente trabajo es presentar la metodología VAR-PLS para estimar h pronósticos conjuntos puntuales y también construir su respectivo intervalo de predicción o confianza.

La metodología propuesta se basa en el trabajo de [Frances \(2006\)](#) quien realiza pronósticos de una serie de tiempo h pasos adelante usando una representación autorregresiva de orden p , denotada como AR(p). Con el objetivo de explotar las estructuras de correlación existentes en la serie de tiempo, el autor propone realizar las prediccio-

nes a través de Mínimos Cuadrados Parciales (PLS). La metodología es conocida como PLSAR(h,p).

Es conocido que en los modelos AR(p) las predicciones son secuenciales, de tal forma que el pronóstico obtenido en el primer paso, es utilizado para pronosticar el segundo y así sucesivamente. El pronóstico h pasos adelante converge a la media incondicional del proceso, Frances (2006).

En el segundo escenario, un pronóstico se realiza en el tiempo $t + 1$ y para el tiempo $t + 2$ el valor anterior es utilizado para reestimar el modelo $AR_i(p)$ donde inclusive el orden p puede cambiar. Es decir que se estima un modelo en cada uno de los h horizontes sólo para pronosticar el siguiente valor. De manera general, esta metodología muestra resultados menos efectivos que las predicciones secuenciales, Frances (2006).

En una autorregresión las variables x_t y x_{t-n} están correlacionadas por lo que los pronósticos de x_{t+h} y x_{t-n+h} también deben de estarlo. Lo anterior sugiere un modelo autorregresivo que de manera conjunta pronostique $(x_{t+h}, x_{t+h-1}, x_{t+h-2}, \dots, x_t)$ a partir de $(x_{t-1}, x_{t-2}, \dots, x_{t-p})$. Tal estimación puede realizarse a través de PLS dado que es natural explotar las estructuras de correlación entre las diferentes particiones de las variables, observadas y por pronosticar.

En esta trabajo, se contribuye en las siguientes direcciones:

- Se extiende la idea de Frances (2006) para el caso multivariado, obteniendo de esta forma predicciones h pasos adelante para un conjunto de series de tiempo. Se estiman los intervalos de confianza basados en técnicas de Bootstrap para series de tiempo.

El contenido de éste trabajo se describe como sigue. En el Capítulo 2 se introducen los conceptos estadísticos y computacionales para el planteamiento y desarrollo de la metodología VAR-PLS. En el Capítulo 3 se describe a detalle la metodología VAR-PLS para la obtención de lo pronósticos y sus intervalos de confianza. En el Capítulo 4 se muestran las ilustraciones empíricas usando una base de datos conocida en la literatura, Juselius (2007), y para una aplicación de la inflación mexicana. En ambos casos los resultados son comparados con los obtenidos a través de modelos VAR. Finalmente, en el Capítulo 5 se presentan las conclusiones y líneas futuras.

Fundamentos teóricos

En este Capítulo se abordan las bases de los elementos estadísticos y computacionales utilizados para el desarrollo de la metodología VAR-PLS que se plantea en el siguiente Capítulo. El Capítulo se compone de tres secciones:

En la Sección 2.1 se revisan brevemente los conceptos fundamentales concernientes a las series de tiempo univariadas: definición probabilística, los supuestos que permiten su estudio estadístico, los modelos clásicos AR y su relación con las raíces unitarias.

La Sección 2.2 inspecciona el modelo VAR partiendo de su reciente historia y uso dentro del modelado econométrico hasta su formulación estadística, continuando con el concepto de cointegración, las dinámicas a corto y largo plazo y su relación con el Modelo de Corrección de Errores (VECM).

Posteriormente, en la Sección 2.3 se describen algunos métodos de reducción de dimensionalidad y de regresión supervisada.

Es importante señalar que las tres secciones que componen este Capítulo son relevantes para introducir al lector al modelo VAR-PLS, dado que dicho modelo se basa en estos conceptos.

2.1 Series de tiempo

Llamamos serie de tiempo a un conjunto de observaciones x_t , donde se observa y registra cada una de ellas en un instante de tiempo t y se asume que las observaciones son igualmente espaciadas.

Las series de tiempo pueden ser discretas en el tiempo, por ejemplo la cantidad de personas que usan diariamente el sistema de transporte colectivo; o bien continuas, como la temperatura que registra un sensor dentro de un horno a lo largo del día. En

este trabajo se trabaja con series de tiempo discretas en su dominio.

Si bien los datos son el registro de lo que se puede observar de una serie de tiempo, se requiere de un modelo matemático para representar el proceso generador de los mismos, para poder realizar inferencia al respecto y para poder desarrollar un modelo satisfactorio con el que se pueda realizar predicciones de los valores futuros de la serie.

Formalmente cada observación x_t es la realización de una variable aleatoria X_t , por lo que el conjunto de datos que forman la serie de tiempo es una realización de la familia de variables aleatorias $\{X_t, t \in T_0\}$, a su vez esta familia de variables aleatorias forman parte de un proceso estocástico $\{X_t, t \in T\}$, donde $T_0 \subset T$, definido sobre un espacio de probabilidad (Ω, \mathcal{F}, P) . En el contexto de series de tiempo llamamos índice a t y suele ser un subconjunto de \mathfrak{R} .

Recordando la definición de variable aleatoria se debe notar que para un instante de tiempo fijo $t \in T$, X_t es una función $X_t(\cdot)$ en el conjunto de eventos Ω y por otro lado al fijar un evento $w \in \Omega$, el caso de interés en este trabajo, $X_t(w)$ es una función de T .

La idea a destacar de los conceptos anteriores es que una serie de tiempo se refiere tanto a los datos observados como a la realización de un proceso estocástico, de manera general, el teorema de Kolmogorov asegura que bajo ciertas condiciones todo proceso estocástico posee una función de distribución, [Brockwell and Davis \(1986\)](#).

2.1.1 Estacionariedad y estacionariedad estricta

Si una serie de tiempo $\{X_t, t \in T\}$ tiene varianza finita para todo $t \in T$, entonces su función de autocovarianza $\gamma_x(\cdot, \cdot)$ se define como:

$$\gamma_x(r, s) = Cov(X_r, X_s) = E[(X_r - E(X_r))(X_s - E(X_s))], \quad \text{con } r, s \in T.$$

Se dice que una serie de tiempo es estacionaria si cumple las siguientes tres propiedades:

1. $E(|X_t|^2) < \infty$, para todo $t \in \mathbb{Z}$
2. $E(X_t) = m$, para todo $t \in \mathbb{Z}$
3. $\gamma_x(r, s) = \gamma_x(r + t, s + t)$, para todo $r, s, t \in \mathbb{Z}$.

Las primeras dos condiciones establecen que la varianza y el valor esperado de la serie de tiempo son iguales en cualquier instante de tiempo t y la tercera condición define la estacionariedad débil, que establece que la autocovarianza de una serie de tiempo en dos momentos diferentes es función sólo de la longitud del intervalo entre ellas y no depende de los instantes en que se observa.

Como consecuencia se tiene que si una serie de tiempo es estacionaria entonces $\gamma_x(r, s) = \gamma_x(r - s, 0)$, $\forall t, s \in \mathbb{Z}$; por lo cual su función de autocovarianza se simplifica a la forma:

$$\gamma_x(h) = \gamma_x(h, 0) = \text{Cov}(X_{t+h}, X_t), \quad \forall t, h \in \mathbb{Z},$$

y su función de autocorrelación con rezago h se define como:

$$\rho_x(h) = \gamma_x(h)/\gamma_x(0) = \text{Corr}(X_{t+h}, X_t), \quad \forall t, h \in \mathbb{Z}.$$

Por otro lado la estacionariedad estricta se tiene cuando las distribuciones conjuntas de $(X_{t_1}, \dots, X_{t_k})'$ y $(X_{t_1+h}, \dots, X_{t_k+h})'$ son iguales, esto puede interpretarse como que las realizaciones de una serie de tiempo en dos intervalos de tiempo de la misma longitud poseen características estadísticas similares. Como puede consultarse en [Brockwell and Davis \(1986\)](#) la estacionariedad estricta implica a la débil y el inverso sólo es válido en el caso de la distribución normal multivariada. Para fines prácticos la estacionariedad débil es más fácil de establecer que su versión estricta, por lo que se trabaja con ella. En lo siguiente estacionariedad se refiere a estacionariedad débil.

Es común descomponer a una serie de tiempo X_t de la siguiente manera:

$$X_t = m_t + s_t + \epsilon_t,$$

donde m_t es la componente de tendencia; s_t es la componente estacional, que puede contener ciclos de diversas longitudes, y ϵ_t es una componente aleatoria y estacionaria.

Esta representación puede estar en términos multiplicativos, lo que indicaría que es recomendable aplicar $\ln(\cdot)$ para linealizarla. De esta forma, para remover el comportamiento estacional podemos recurrir, por ejemplo, al algoritmo X-13ARIMA-SEATS, ver [Sax and Eddelbuettel \(2018\)](#), centrándonos únicamente en la tendencia y en el término de error. Por otra parte, si m_t tiene una tendencia estocástica, se suele recurrir a diferenciar la variable, lo anterior debe realizarse con precaución dado que sobrediferenciar una serie de tiempo puede incrementar la varianza de la componente aleatoria, ver por ejemplo [Chan \(2010\)](#).

Las estimaciones muestrales de la función de autocovarianza y de autocorrelación se definen, respectivamente, de la siguiente forma:

$$\hat{\gamma}(h) = n^{-1} \sum_{j=1}^{n-h} (x_{j+h} - \bar{x})(x_j - \bar{x}), \quad 0 \leq h \leq n$$

$$\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0), \quad |h| < n,$$

donde \bar{x} es la media muestral, $\bar{x} = n^{-1} \sum_{j=1}^n x_j$.

2.1.1.1 Ergodicidad

Un concepto de especial interés en el análisis de series de tiempo es el de ergodicidad. Se dice que un proceso estocástico es ergódico si la media muestral de una realización del proceso converge al subyacente parámetro del proceso.

Para procesos ergódicos no se requiere observar varias realizaciones independientes del proceso para determinar su media y momentos de orden superior, pues la convergencia de la media muestral garantiza la existencia del primer y segundo momento que es compartido por todas las realizaciones del proceso estocástico, Chan (2010), del que además se puede estimar la autocovarianza.

2.1.2 Modelos AR

Una categoría de modelos, ampliamente usados en el análisis de series de tiempo, son los modelos autorregresivos AR. Estos modelos tienen una interpretación sencilla y se parecen a los modelos de regresión lineal tradicionales cuando se reemplaza al predictor por su valor anterior rezagado.

Sea B al operador de rezago de tal forma que si $\{x_1, \dots, x_t\}$ es una serie de tiempo, se tiene que $BX_t = X_{t-1}$ y de manera análoga al iterar n veces el operador B obtenemos que $B^n X_t = X_{t-n}$.

Formalmente un modelo $AR(p)$, autorregresivo de orden p , se puede escribir como $\phi(B)X_t = \epsilon_t$, donde $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$, a fin de que:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t,$$

donde $\{X_t\}$ es estacionario. Análogamente se dice que una serie de tiempo $\{X_t\}$ es un proceso $AR(p)$ con media μ si $\{X_t - \mu\}$ es un proceso $AR(p)$.

Una característica importante que algunos modelos AR poseen es la causalidad. Se dice que un proceso AR es causal si su valor en el tiempo t depende sólo de los valores observados anteriores, $t^* < t$, y no de valores futuros, es decir que existen constantes $\{\psi_i\}$ con $\sum_{i=0}^{\infty} |\psi_i| < \infty$ tales que $X_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$. De manera equivalente un proceso $AR(p)$ es causal si las raíces del polinomio característico $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ se encuentran fuera del círculo unitario $\{z : |z| > 1\}$.

Los parámetros $\{\phi_i\}$ de un modelo $AR(p)$ pueden estimarse por mínimos cuadrados o bien por máxima verosimilitud, lo cual requiere de suponer una distribución normal a la componente aleatoria ϵ_t . Un aspecto fundamental en esta estimación es la determinación del parámetro p que indica el orden del rezago del modelo AR. Para la determinación de este parámetro se pueden seguir dos métodos: el Error de Predicción

Final (*FPE*) y el criterio de información de Akaike (*AIC*).

Para ilustrar el *FPE*, considérese una realización $X = (X_1, \dots, X_n)$ de un modelo $AR(p)$ con $p < n$. Si $\hat{\phi}_1, \dots, \hat{\phi}_p$ son las estimaciones de máxima verosimilitud a partir de X entonces se puede estimar la varianza $\hat{\sigma}^2$, de los errores $\hat{\epsilon}_t$, lo cual permite definir:

$$FPE = \hat{\sigma}^2 \left(\frac{n+p}{n-p} \right),$$

en donde se busca el valor de p que minimice el *FPE*.

2.1.3 Raíces unitarias

Hasta este punto se ha hablado de series de tiempo estacionarias, pero en la práctica se encuentran diferentes tipos de no-estacionariedad. La no-estacionariedad puede deberse a cambios sistemáticos en la media o en la varianza. Cuando la no-estacionariedad se debe a cambios en la varianza ésta puede controlarse por medio de aplicar una transformación de los datos como la función $\ln(\cdot)$, o bien cuando se observa heteroscedasticidad, bajo ciertos supuestos, el proceso puede modelarse como un ARCH,¹ ver por ejemplo, [Hamilton \(1994\)](#).

Cuando la no-estacionariedad se debe a cambios en la media se presentan las raíces unitarias. Consideremos la prueba para el coeficiente del $AR(1)$, $H : \alpha = 1$ para el modelo:

$$X_t = \beta_1 + X_{t-1} + \epsilon_t,$$

para ilustrar la idea consideremos $\beta_1 = 0$ en la ecuación anterior, entonces bajo $H, \{X_t\}$ es una caminata aleatoria y las pruebas estadísticas para este tipo de modelo son conocidas en la literatura de series de tiempo y econometría como pruebas de raíces unitarias. Si se supone una distribución normal e independencia en la componente estocástica ϵ_t la estimación por mínimos cuadrados $\hat{\alpha}$ para α es:

$$\hat{\alpha} = \frac{\sum_{t=1}^n Y_t Y_{t-1}}{\sum_{t=1}^n Y_{t-1}^2},$$

en particular

$$n(\hat{\alpha} - 1) = \frac{(1/n) \sum_{t=1}^n Y_{t-1} \epsilon_t}{(1/n^2) \sum_{t=1}^n Y_{t-1}^2}.$$

Puede demostrarse ([Chan, 2010](#)) que el lado izquierdo de la ecuación anterior tiene una convergencia casi segura a una expresión en términos del movimiento Browniano. El resultado anterior provee la prueba de Dickey-Fuller para raíces unitarias donde la restricción de $\beta = 0$ puede eliminarse y se conoce como la prueba de Dickey-Fuller

¹Del inglés Autoregressive Conditional Heteroskedasticity

aumentada.

Continuemos con el concepto de orden de integración. Consideremos el proceso $X_t = \sum_{i=1}^n \epsilon_i$, donde ϵ_t es una secuencia de variables aleatorias con distribución normal descorrelacionadas con media cero y varianza σ^2 , usualmente llamado ruido blanco. Entonces X_t tiene varianza $t\sigma^2$ así que X_t es no-estacionaria y su orden de integración es 1.

Decimos que el proceso $X_t = \sum_{i=1}^{\infty} \phi_i \epsilon_i$ es integrado de orden cero, denotado como $I(0)$, si ϵ_t es ruido blanco y $\sum_{i=1}^{\infty} \phi_i \neq 0$. De manera análoga el proceso $\{X_t\}$ es integrado de orden 1, si $\Delta X_t = (1 - B)X_t = X_t - X_{t-1}$ es $I(0)$. De manera general un proceso X_t con d raíces unitarias se dice que tiene orden de integración d , $I(d)$.

2.2 Modelos VAR

Se dice que una serie de tiempo k -variada es un proceso estocástico que contiene vectores de dimensión k , $(X_{t1}, X_{t2}, \dots, X_{tk})'$ para $t = 0, 1, 2, \dots$. Las componentes de la serie $\{X_{ti}\}$ pueden ser estudiadas independientemente como una serie de tiempo univariada, cada una caracterizada por su media y función de autocovarianza. Tal aproximación falla al considerar posible dependencia entre componentes, y tal dependencia cruzada puede ser de gran importancia para predecir valores futuros de cada componente.

Considérese la serie de vectores aleatorios $\mathbf{X}_t = (X_{t1}, \dots, X_{tk})'$, se define su vector de medias como:

$$\mu_t = E(\mathbf{X}_t) = (E(X_{t1}), \dots, E(X_{tk}))'$$

y matriz de covarianza:

$$Cov(X_{t+h,i}, X_{t,i}) = \begin{pmatrix} \gamma_{11}(t+h, t) & \dots & \gamma_{1k}(t+h, t) \\ \vdots & & \vdots \\ \gamma_{k1}(t+h, t) & \dots & \gamma_{kk}(t+h, t) \end{pmatrix},$$

donde

$$\gamma_{ij}(t+h, t) = Cov(X_{t+h,i}, X_{t,j}),$$

en notación matricial:

$$\Gamma(t+h, t) = E((X_{t+h} - \mu_{t+h})(X_t - \mu_t)').$$

Se dice que la serie \mathbf{X}_t es estacionaria si los momentos μ_t y $\Gamma(t+h, t)$ son ambos independientes de t , en ese caso se usa la notación

$$\mu = E(\mathbf{X}_t),$$

y

$$\Gamma(h) = Cov(X_{t+h}, X_t),$$

los elementos en la diagonal de la matriz anterior son las autocovarianzas de las series univariadas $\{X_{ti}\}$, mientras que los elementos fuera de la diagonal son las covarianzas entre $X_{t,i}$ y $X_{t,j}$, con $i \neq j$. De manera correspondiente la matriz de correlación se define como:

$$R(h) = \begin{pmatrix} \rho_{11}(h) & \dots & \rho_{1k}(h) \\ \vdots & & \vdots \\ \rho_{k1}(h) & \dots & \rho_{kk}(h) \end{pmatrix},$$

donde

$$\rho_{ij} = \gamma_{ij}(h)(\gamma_{ii}(0)\gamma_{jj}(0))^{-1/2}.$$

Denotamos como $\{\epsilon_t\} \sim i.i.d. (0, \Sigma)$ si $\{\epsilon_t\}$ son independientes e idénticamente distribuidas con media $\mu=0$ y matriz de covarianza Σ .

\mathbf{X}_t es un proceso lineal si puede ser expresado como:

$$X_t = \sum_{j=-\infty}^{\infty} \Phi_j \epsilon_{t-j},$$

donde ϵ_t es ruido blanco y $\{\Phi_j\}$ es una secuencia de matrices de tamaño $k \times k$ cuyas entradas son absolutamente sumables, es decir:

$$\sum_{j=-\infty}^{\infty} |\Phi_j(i, l)| < \infty, \text{ con } i, l = 1, 2, \dots, k.$$

Con lo anterior puede definirse el modelo de series de tiempo, vector autorregresivo de orden p , $VAR(p)$, como:

$$X_t = \mu + \Phi_1 X_{t-1} + \dots + \Phi_p X_{t-p} + \epsilon_t, \quad (2.1)$$

con ϵ_t ruido blanco. Nótese que $\mu = (\mu_1, \dots, \mu_k)'$ es un vector fijo de términos de intercepto, permitiendo la posibilidad de media diferente de cero.

La mayor parte de la teoría de series de tiempo univariadas puede extenderse al caso multivariado de manera natural, sin embargo surgen problemas. Por ejemplo el problema de la dimensionalidad. Cuando el número de componentes en \mathbf{X}_t se incrementa, el número de parámetros también crece. Por ejemplo para una serie de tiempo con 10 componentes, incluso en el sencillo caso de un modelo $AR(1)$ se tienen 100 parámetros libres.

Considérese un modelo $VAR(1)$:

$$X_t = \mu + \Phi_1 X_{t-1} + \epsilon_t,$$

si este modelo generador comienza en el tiempo $t = 1$ se tiene:

$$\begin{aligned} X_1 &= \mu + \Phi_1 X_0 + \epsilon_1 \\ X_2 &= \mu + \Phi_1 X_1 + \epsilon_2 = \mu + \Phi_1(\mu + \Phi_1 X_0 + \epsilon_1) + \epsilon_2 \\ &= (I + \Phi_1)\mu + \Phi_1^2 X_0 + \Phi_1 \epsilon_1 + \epsilon_2 \\ &\vdots \\ X_t &= (I + \Phi_1 + \dots + \Phi_1^{t-1})\mu + \Phi_1^t X_0 + \sum_{i=0}^{t-1} \Phi_1^i \epsilon_{t-i}, \end{aligned}$$

continuando de esta manera podemos escribir el proceso $VAR(1)$ como:

$$X_t = \mu + \sum_{i=0}^{\infty} \Phi_1^i \epsilon_{t-i},$$

donde el vector de medias de \mathbf{X}_t , $E(X_t) = \mu + \Phi_1 \mu + \Phi_1^2 \mu + \dots$ si Φ_1 tiene todos sus valores propios menor a 1 en módulo.

De esta manera se dice que un proceso $VAR(1)$ es estable si:

$$\det(I - z\Phi_1) \neq 0, \text{ para } |z| \leq 1.$$

Para un proceso general $VAR(p)$, la discusión anterior se puede extender reescribiendo cualquier $VAR(p)$ en forma de $VAR(1)$, específicamente podemos escribir:

$$\tilde{\mathbf{X}}_t = \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{pmatrix}, \tilde{\mu} = \begin{pmatrix} \mu \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tilde{\Phi} = \begin{pmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_{p-1} & \Phi_p \\ I_k & 0 & \dots & 0 & 0 \\ 0 & I_k & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_k & 0 \end{pmatrix}, \tilde{\epsilon}_t = \begin{pmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

donde $\tilde{\mathbf{X}}_t$, $\tilde{\mu}$ y $\tilde{\epsilon}_t$ son de dimensión $kp \times 1$ y $\tilde{\Phi}$ es $kp \times kp$. De esta manera se tiene que $\tilde{\mathbf{X}}_t$ es estable si:

$$\det(I_{kp} - z\tilde{\Phi}) \neq 0, \text{ para } |z| \leq 1.$$

En la forma del modelo VAR (2.1) no existe sólo un orden p correcto para el proceso. De hecho, si (2.1) es una forma correcta de escribir el proceso \mathbf{X}_t , lo mismo es cierto para:

$$X_t = \mu + \Phi_1 X_{t-1} + \dots + \Phi_p X_{t-p} + \Phi_{p+1} X_{t-p-1} + \epsilon_t,$$

con $\Phi_{p+1} = \mathbf{0}$.

En otras palabras si \mathbf{X}_t es un proceso $VAR(p)$, en este sentido también es un $VAR(p+1)$. Es útil tener un único número que determine el orden del proceso $VAR(p)$. En lo siguiente se llamará a \mathbf{X}_t un proceso $VAR(p)$ si $\Phi_p \neq \mathbf{0}$ y $\Phi_i = \mathbf{0}$ para $i > p$, de esta manera p es el orden posible más pequeño y define el orden del proceso $VAR(p)$. Esto es necesario, pues como se mencionó anteriormente en esta Sección, el ajuste de modelos VAR con innecesarios órdenes de rezago incrementa el número de parámetros y puede disminuir la calidad de las estimaciones y por ende de los pronósticos, [Lütkepohl \(2006\)](#).

De manera análoga a los modelos AR, la determinación del orden de un modelo $VAR(p)$ es un aspecto muy importante a considerar. En general, la determinación del orden p de un VAR se realiza por medio de una secuencia de pruebas donde las hipótesis son de la forma $H_0 : \Phi_m = 0$, si la hipótesis anterior se rechaza se prosigue con la prueba $H_0 : \Phi_{m+1} = 0$ hasta que no sea rechazada la hipótesis nula.

Suponiendo normalidad para la componente estocástica de la serie multivariada \mathbf{X}_t , se puede recurrir al estadístico de Wald de la prueba de razón de verosimilitud para determinar el orden p , ver por ejemplo, [Lütkepohl \(2006\)](#). Éste enfoque se centra en determinar el proceso que genera los datos, sin embargo si nuestro objetivo es pronosticar existen otras maneras de determinar el orden de un modelo $VAR(p)$ como la elección de p por medio del valor que minimice el error cuadrático medio, el error porcentual absoluto medio (MAPE), el FPE o bien el criterio de Akaike, en sus versiones multivariadas, ([Lütkepohl, 2006](#)).

El enfoque anterior en la determinación del orden p requiere de dividir la muestra en dos partes: digamos que la primer parte consiste en las primeras l observaciones del vector \mathbf{X}_t y el resto a las últimas $t - l$ observaciones, la primera será usada como conjunto de entrenamiento donde podemos utilizar la prueba de razón de verosimilitud o el FPE para determinar un orden adecuado y con este orden predecir los siguientes $t - l$ valores del vector \mathbf{X}_t , de esta manera podemos utilizar los valores predichos y los reales para la elección del orden que mejor desempeño tenga al disminuir el criterio que hayamos elegido y fijado.

En la caracterización anterior de los modelos VAR, las matrices Φ_i son constantes, aunque en la práctica es común que las series de tiempo univariadas presenten cambios estructurales que invaliden lo anterior. Por otro lado, después de que se ha determinado el orden p de un modelo VAR es importante validar los supuestos del modelo: la componente estocástica es ruido blanco y que los residuos están descorrelacionados (para diversos rezagos h). Para comprobar la normalidad de los residuos existen diversas pruebas como la prueba de Jarque-Bera, ([Sheskin, 2000](#)), en cuanto a la descorrelación entre los residuos se puede emplear la prueba de Portmanteau o la prueba de multiplicadores de Lagrange, ([Lütkepohl, 2006](#)).

2.2.1 Contexto del uso de modelos VAR en econometría

Seguindo a Juselius (2007), los economistas frecuentemente formulan modelos bien especificados *económicamente* y aplican métodos estadísticos para estimar sus parámetros. En contraste, los estadísticos formulan modelos bien especificados *estadísticamente* para los datos y analizan el modelo estadístico para resolver las cuestiones económicas de interés. En el primer caso, la estadística es usada pasivamente como una herramienta para obtener algunas estimaciones, y en el segundo caso el modelo estadístico es usado activamente como medio de análisis del subyacente proceso generador del fenómeno en cuestión.

El principio general de analizar modelos estadísticos en macroeconomía fue introducido por Haavelmo en los 40's, (Haavelmo, 1944). El enfoque probabilístico de Haavelmo a la econometría requiere de una formulación probabilística del proceso completo que genera los datos. Los aspectos computacionales involucrados en la especificación de los modelos estadísticos eran prohibitivos en los tiempos del trabajo de Haavelmo (1944), cuando incluso la estimación de una regresión múltiple era una tarea no trivial. El desarrollo del cómputo actual, hace factible las líneas adoptadas en Haavelmo (1944) en la econometría empírica. Es aquí donde los modelos VAR ofrecen ventajas como un marco de referencia general para dirigir las preguntas empíricas en economía.

Un aspecto importante mencionado en la pág. 5 en Haavelmo (1944) es el cómo las *verdaderas* variables -como funciones del tiempo- representan un ideal de mediciones precisas de la realidad, mientras que las variables definidas en teoría son las mediciones que se tienen si la realidad está de acuerdo con el modelo teórico. Es decir, que aún en el supuesto de que el modelo estadístico refleje adecuadamente la realidad, los datos pueden no medir directamente la variable en cuestión. Las mediciones disponibles de parte de las estadísticas oficiales, como es el caso de las variables que utilizaremos en los capítulos posteriores, pueden distar de las definiciones de las variables verdaderas. Cuestiones alrededor de las variables macroeconómicas que suelen reportarse como agregados¹ en unidades de tiempo como meses, trimestres, semestres, etc., se relacionan con algunos principios generales para la modelación VAR con series no-estacionarias.

2.2.2 Cointegración

Hasta este punto se ha hablado de modelos VAR estables, donde cada una de sus componentes es estacionaria. Pasemos ahora al caso general en donde el modelo VAR puede incluir series no-estacionarias.

De manera general un proceso k -dimensional \mathbf{X}_t es llamado cointegrado de orden (d, b) , escrito de manera breve $\mathbf{X}_t \sim CI(d, b)$, si todas las componentes de \mathbf{X}_t

¹Inclusive nuevas componentes pueden entrar al agregado o bien los cambios en tecnología y en la definición de los intervalos en que se realiza la medición pueden alterar el valor reportado.

comparten el mismo orden de integración, $I(d)$, y existe una combinación lineal de las componentes de \mathbf{X}_t tal que $\epsilon_t = \beta' X_t$ con $\beta = (\beta_1, \dots, \beta_k)' \neq 0$ donde ϵ_t es $I(d - b)$.

Un proceso que contiene variables cointegradas es llamado cointegrado, así los procesos VAR con variables cointegradas son llamados modelos VAR cointegrados, estudiados a profundidad en los últimos años, (Juselius, 2007). Los procesos cointegrados fueron introducidos por Granger en los 80's, (Granger, 1981).

Por ejemplo si todas las componente de \mathbf{X}_t son $I(1)$ y $\beta' X_t$ es estacionario, $I(0)$, entonces $\mathbf{X}_t \sim CI(1, 1)$. El vector β es llamado vector de cointegración.

El vector de cointegración anterior no es único. Pueden existir vectores de cointegración linealmente independientes que se pueden presentar, por ejemplo, en el caso de tener cuatro variables en un sistema donde las dos primeras están conectadas por una relación a largo plazo, al igual que las otras dos restantes. Entonces un vector de cointegración con ceros en sus dos últimas componentes refleja la primera relación, y un vector con ceros en sus dos primeras componentes refleja la segunda relación, además puede existir un vector de cointegración involucrando las cuatro variables.

La cointegración implica que ciertas combinaciones lineales de las componentes de la serie multivariada poseen un orden de integración menor al del proceso por sí mismo. Las variables cointegradas son accionadas por los mismos choques persistentes, o cambios en el tiempo, a los cuales las variables responden. Entonces podemos pensar que la no-estacionariedad de una variable corresponde, o se relaciona, con la no-estacionariedad de otra variable, por lo que existe una combinación lineal de ellas que es estacionaria. Otra manera de expresar lo anterior es que cuando dos o más variables tienen una tendencia estocástica, ellas muestran una tendencia a moverse juntas en el largo plazo.

Una ventaja de los modelos VAR, es que el principio de cointegración es invariante a extensiones del conjunto de datos, es decir que se pueden incorporar nuevas series, a diferencia del análisis de regresión donde una nueva variable puede alterar las estimaciones existentes dramáticamente, además de que los modelos VAR no requieren de especificar la diferencia entre variables endógenas y exógenas.

2.2.3 Modelos VEC

Antes de que se introdujera el concepto de proceso cointegrado en la literatura econométrica existía un concepto relacionado, el concepto de modelo de corrección de errores o modelo de corrección de equilibrio, (Juselius, 2007). En un modelo de corrección de errores, denotado en este trabajo por modelos VEC, los cambios en una variable dependen de las desviaciones a partir de una relación de equilibrio. Para ejemplificar considerese que X_{t1} representa el precio de un bien en un mercado en particular y X_{t2}

es el correspondiente precio del mismo bien en otro mercado. Asíumase que la relación de equilibrio de las dos variables está dada por $X_{t1} = \beta_1 X_{t2}$ y que los cambios de X_{t1} dependen de las desviaciones de este equilibrio en el periodo $t - 1$, es decir:

$$\Delta X_{t1} = \alpha_1(X_{t-1,1} - \beta_1 X_{t-1,2}) + \epsilon_{t1},$$

una relación similar puede ser válida para X_{t2}

$$\Delta X_{t2} = \alpha_2(X_{t-1,1} - \beta_1 X_{t-1,2}) + \epsilon_{t2},$$

en un modelo de corrección de errores más general, los ΔX_{ti} deben de depender además de cambios previos en ambas variables, como por ejemplo el siguiente modelo:

$$\begin{aligned} \Delta X_{t1} &= \alpha_1(X_{t-1,1} - \beta_1 X_{t-1,2}) + \gamma_{11,1} \Delta X_{t-1,1} + \gamma_{12,1} \Delta X_{t-1,2} + \epsilon_{t1} \\ \Delta X_{t2} &= \alpha_2(X_{t-1,1} - \beta_1 X_{t-1,2}) + \gamma_{21,1} \Delta X_{t-1,1} + \gamma_{22,1} \Delta X_{t-1,2} + \epsilon_{t2}, \end{aligned}$$

para notar la relación entre el modelo de corrección de errores y el concepto de cointegración, supongamos que X_{t1} y X_{t2} son ambas $I(1)$. En este caso los términos en la expresión anterior que incluyen ΔX_{ti} son estables, más aún, ϵ_{t1} y ϵ_{t2} son ruido blanco por lo cual también son estables, entonces la siguiente expresión es estable:

$$\alpha_i(X_{t-1,1} - \beta_1 X_{t-1,2}) = \Delta X_{ti} - \gamma_{i1,1} \Delta X_{t-1,1} - \gamma_{i2,1} \Delta X_{t-1,2} - \epsilon_{ti}, \quad (2.2)$$

si $\alpha_i \neq 0$, $X_{t1} - \beta_1 X_{t2}$ es estable y representa una relación de cointegración. En notación matricial el modelo anterior puede ser escrito como:

$$\Delta \mathbf{X}_t = \alpha \beta' \mathbf{X}_{t-1} + \Gamma_1 \Delta \mathbf{X}_{t-1} + \epsilon_t,$$

donde $\mathbf{X}_t = (X_{t1}, X_{t2})'$, $\epsilon_t = (\epsilon_{t1}, \epsilon_{t2})'$,

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \beta' = (1, -\beta_1), \text{ y } \Gamma_1 = \begin{pmatrix} \gamma_{11,1} & \gamma_{12,1} \\ \gamma_{21,1} & \gamma_{22,1} \end{pmatrix},$$

reordenando 2.2 se tiene la representación $VAR(2)$:

$$\mathbf{X}_t = (I_k + \Gamma_1 + \alpha \beta') \mathbf{X}_{t-1} - \Gamma_1 \mathbf{X}_{t-2} + \epsilon_t$$

donde de nuevo las variables cointegradas son generadas por un proceso VAR. Para el caso general del proceso $VAR(p)$, k -dimensional donde sus componentes son $I(1)$ o $I(0)$:

$$\mathbf{X}_t = \Phi_1 \mathbf{X}_{t-1} + \dots + \Phi_p \mathbf{X}_{t-p} + \epsilon_t,$$

es llamado cointegrado de rango r si :

$$\Pi = -(I_k - \Phi_1 - \dots - \Phi_p),$$

tiene rango r , entonces Π puede factorizarse de la forma $\alpha\beta'$ con α y β siendo de dimensión $k \times r$ y de rango r . La matriz β es llamada matriz de cointegración y α es llamada la matriz de cargas. Si $r = 0$ entonces ΔX_t es estable y tiene una representación $VAR(p - 1)$ y si $r = k$ entonces el proceso VAR no tiene raíces unitarias y es un VAR estable.

2.3 Reducción de dimensionalidad y componentes principales supervisados

La regresión por Mínimos Cuadrados Parciales (PLS por sus siglas en inglés), que puede ser vista como un método de componentes principales supervisados, es particularmente útil cuando el número de variables es mayor al número de observaciones. La regresión PLS será de utilidad para realizar la estimación del pronóstico conjunto en la metodología VAR-PLS.

También la regresión PLS puede ser considerada como un método de reducción de dimensión pues crea variables que son combinaciones lineales de las variables originales -llamadas componentes- presentes en un conjunto de datos y la elección de un menor número de estas implica una reducción en la dimensión del conjunto de datos con el cual se trabaja. El trabajar con un menor número de variables manteniendo fijo el tamaño de la muestra tiende a mejorar la calidad de la inferencia estadística, al igual que puede mejorar el desempeño de la regresión y en el contexto de este trabajo los pronósticos conjuntos.

La reducción de dimensionalidad no debe confundirse con los métodos de selección de variables. Los primeros producen nuevas variables que son combinación de las demás ya sean lineales como en el caso del Análisis de Componentes Principales (PCA por sus siglas en inglés) o no lineales como ISOMAP, (Hastie et al., 2009); mientras que los segundos realizan una búsqueda en el conjunto potencia generado por el conjunto de variables de un conjunto de datos, por ejemplo la regresión LASSO o la búsqueda exhaustiva de subconjuntos, (Hastie et al., 2009).

2.3.1 PCA

Las componentes principales son un conjunto de datos en \mathbb{R}^p que proporcionan una secuencia de la mejor aproximación lineal de un conjunto de datos, todas de rango $q \leq p$. Consideremos las observaciones x_1, x_2, \dots, x_N y un modelo lineal de rango q para representarlas:

$$f(\lambda) = \mu + V_q \lambda$$

donde μ es un vector de localización y \mathbf{V}_q es una matriz de dimensiones $p \times q$ con q vectores unitarios como columnas y λ es un vector de q parámetros. Notemos que esta es la representación paramétrica de un plano afín de rango q . Ajustar tal modelo a un conjunto de datos por medio de mínimos cuadrados implica minimizar el *error de reconstrucción*:

$$\min_{\mu, \{\lambda_i\}, \mathbf{V}_q} \sum_{i=1}^N \|x_i - \mu - \mathbf{V}_q \lambda_i\|^2$$

cuya solución está dada por:

$$\begin{aligned} \hat{\mu} &= \bar{x} \\ \hat{\lambda}_i &= \mathbf{V}_q^t (x_i - \bar{x}) \end{aligned}$$

lo anterior se reduce a encontrar la matriz con columnas ortogonales \mathbf{V}_q :

$$\min_{\mathbf{V}_q} \sum_{i=1}^N \|(x_i - \bar{x}) - \mathbf{V}_q \mathbf{V}_q^t (x_i - \bar{x})\|^2$$

se puede asumir que la media muestral es cero, lo cual se logra centrando las observaciones, la matriz $\mathbf{H} = \mathbf{V}_q \mathbf{V}_q^t$ es una matriz de proyección y mapea a cada punto x_i en su reconstrucción de rango reducido q , $\mathbf{H}_q x_i$, que no es más que la proyección de x_i en el subespacio generado por las columnas de \mathbf{V}_q .

Una manera de calcular la solución anterior es por medio de la descomposición en valores singulares de la matriz de datos¹ \mathbf{X} , [Golub and Van-Loan \(1996\)](#):

$$\mathbf{X} = \mathbf{U} \mathbf{D} \mathbf{V}^t,$$

donde los vectores columna u_i , son los vectores singulares por la izquierda y son ortogonales entre sí, de manera análoga la matriz \mathbf{V} es ortogonal de dimensiones $p \times p$ cuyas columnas son los vectores singulares por la derecha y finalmente \mathbf{D} es una matriz diagonal de dimensión $p \times p$ cuyos elementos no cero son los valores singulares de la matriz \mathbf{X} tales que $d_1 \geq d_2 \geq \dots \geq d_p \geq 0$. Para cada rango q los vectores que forman \mathbf{V}_q consiste en las primeras q columnas de \mathbf{V} . Las columnas de $\mathbf{U} \mathbf{V}$ forman las componentes principales de \mathbf{X} .

La elección óptima de un número de componentes principales es un problema abierto, existen alternativas para su elección como: el criterio del codo, el criterio de Kaiser

¹A pesar de que existen diferentes algoritmos para efectuar tal descomposición de manera general la complejidad en tiempo de tal factorización es de $O(m^2n + n^3)$ para una matriz densa de tamaño $m \times n$.

Kaiser (1959), el *parallel analysis* propuesto en Horn (1965) y otros métodos basados en simulación Monte Carlo y resultados de matrices aleatorias, Braeken and Assen (2017).

Si bien los componentes principales tienen una práctica interpretación geométrica, poseen otras agradables propiedades como que la combinación lineal $\mathbf{X}v_1$ tiene la varianza más grande entre todas las combinaciones lineales de las variables originales, $\mathbf{X}v_2$ posee la mayor varianza con respecto a todas las combinaciones lineales que son ortogonales a la anterior, etc.

Existen otras maneras de estimar las componentes principales; por ejemplo suponiendo que poseen una distribución normal, en este contexto se pueden estimar por medio de máxima verosimilitud y usando el método EM, (Hastie et al., 2009). Con el supuesto anterior existen métodos iterativos para estimar las componentes en matrices no densas (Roweis, 1998).

2.3.2 PLS

El método de PLS puede ser usado para regresión multivariada, así como univariado. Puede haber varias variables dependientes Y_1, \dots, Y_l y para formar la relación entre las variables \mathbf{Y} y las variables explicativas X_1, \dots, X_m , PLS construye nuevas variables, también llamadas variables latentes o componentes donde cada una de ellas es una combinación lineal de X_1, \dots, X_m .

Este método tiene similitud con PCA donde las componentes principales forman las variables independientes en una regresión. La mayor diferencia entre ambos métodos radica en que las componentes de PCA son determinadas únicamente por los valores de las variables \mathbf{X} mientras que en PLS los valores de ambos conjuntos de variables \mathbf{X} y \mathbf{Y} influyen en la construcción de las componentes.

La intención de PLS es formar componentes que capturen la mayor parte de la información en las variables \mathbf{X} que sea útil para predecir Y_1, \dots, Y_l , mientras se reduce la dimensionalidad del problema de regresión usando un número menor de componentes que el número total de variables en \mathbf{X} , Garthwaite (1994).

En lo siguiente, primero se detalla la manera de estimar PLS para el caso univariado y posteriormente se da el algoritmo para el caso multivariado concluyendo con las propiedades de las componentes y algunas consideraciones importantes.

Las componentes derivadas por PLS pueden ser vistas como promedios ponderados de los predictores, (Garthwaite, 1994). Para el caso univariado supóngase que se tiene una muestra de tamaño n para las cuales se requiere estimar la relación lineal entre Y y \mathbf{X} . Para $i = 1, \dots, n$ el i -ésimo reglón de la muestra es denotado por $(x_{i1}, \dots, x_{im}, y_i)$. Y los valores observados de Y y X_j son denotados como y y x_j , con lo que se tiene

$y = (y_1, \dots, y_n)^t$ y para $j = 1, \dots, m$, $x_j = (x_{j1}, \dots, x_{jn})^t$. Considérese la ecuación de regresión dada por:

$$\hat{Y} = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \dots + \beta_p T_p, \quad (2.3)$$

donde cada componente T_k es una combinación lineal del conjunto de X_j y la correlación muestral para cada par de componentes es 0.

Una ecuación con una gran cantidad de parámetros es típicamente más flexible que una que contiene pocos parámetros, sin embargo, se tiene la desventaja de que los parámetros estimados pueden ser influenciados más fácilmente por el error aleatorio (o error de medición) de los datos. Es en este contexto es donde la regresión de PLS tiene ventaja al reducir el número de componentes T_i en 2.3 empleando menos variables que las que posee \mathbf{X} . Para lograr esto PLS adopta el principio de considerar la relación entre Y y alguna variable específica de \mathbf{X} , las otras variables de \mathbf{X} no son admitidas en tener influencia en la estimación de la relación directamente, pero se permite su influencia a través de las componentes T_k .

Para simplificar notación supongamos que en la primera iteración del algoritmo Y y \mathbf{X} están centradas, es decir que sus columnas tienen media cero y denotémoslas como $Y^{(1)}$ y $X^{(1)}$. En la estimación de PLS las componentes, T_k , son determinadas secuencialmente. La primer componente, T_1 , pretende ser útil para predecir Y y es construida como combinación lineal de las columnas de $\mathbf{X}^{(1)}$, durante su construcción las correlaciones muestrales entre las columnas de $\mathbf{X}^{(1)}$ son ignoradas. Para obtener T_1 , se realiza la regresión¹ de $Y^{(1)}$ contra la primer columna $X_1^{(1)}$, luego contra $X_2^{(1)}$ y así sucesivamente para cada $X_j^{(1)}$. Como $Y^{(1)}$ y $\mathbf{X}^{(1)}$ están centrados, tienen medias muestrales iguales a cero por lo que las regresiones de mínimos cuadrados resultantes son:

$$\hat{Y}_j^{(1)} = b_j^{(1)} X_j^{(1)}, \quad (2.4)$$

donde $b_j^{(1)} = x_j^{(1)t} y^{(1)} / (x_j^{(1)t} x_j^{(1)})$. Dados los valores de $X_j^{(1)}$ para una iteración, las demás m ecuaciones de 2.4 proporcionan una manera de estimar $Y^{(1)}$. Para reconciliar estas estimaciones mientras se ignora las relaciones entre las columnas $V_j^{(1)}$, es posible tomar un simple promedio $\sum_{j=1}^m b_j^{(1)} V_j^{(1)} / m$ o de manera más general un promedio ponderado. Fijemos la componente T_1 igual al promedio ponderado siguiente:

$$T_1 = \sum_{j=1}^m w_j^{(1)} b_j^{(1)} V_j^{(1)}, \quad (2.5)$$

con la restricción de que $w_j^{(1)}$ tenga norma unitaria, aunque esto no es esencial. Como veremos posteriormente, multiplicar T_1 por una constante no afectara a los valores

¹En este caso el superíndice indica el número de iteración en el que se encuentra el algoritmo.

de las subsecuentes componentes ni las predicciones de Y . La ecuación 2.5 permite un abanico de posibilidades para la construcción de T_1 , dependiendo de los pesos que se usen, consideraremos posteriormente dos formas particulares de estos pesos.

Como T_1 es un promedio ponderado de los predictores de $Y^{(1)}$, será un predictor útil para Y , pero como las variables de \mathbf{X} potencialmente contienen información útil para predecir Y , la información en X_j que no está contenida en T_1 puede ser estimada por medio de regresar los residuos de X_j en T_1 , los cuales son idénticos a los residuos de la regresión de $X_j^{(1)}$ contra T_1 . De manera análoga la variabilidad de Y que no es explicada por T_1 puede ser estimada por medio de los residuos de la regresión de $Y^{(1)}$ en T_1 . Estos residuos son denotados por $X_j^{(2)}$ para $X_j^{(1)}$ y por $Y^{(2)}$ para $Y^{(1)}$. La siguiente componente, T_2 , es una combinación lineal de los $X_j^{(2)}$ que serán de utilidad para predecir $Y^{(2)}$.

El procedimiento se extiende iterativamente para obtener los componentes T_3, \dots, T_p donde cada componente es determinada a partir de los residuos de las regresiones de las componentes precedentes relacionando la variabilidad residual en Y con la información residual en las columnas de \mathbf{X} . específicamente suponga que $T_i, (i \geq 1)$ simplemente ha sido construido con las variables $Y^{(i)}$ y $X_j^{(i)}$ ($j = 1, \dots, m$) y sean los valores muestrales de $T_i, Y^{(i)}$ y las $X_j^{(i)}$ denotados por $t_i, y^{(i)}$ y $x_j^{(i)}$.

Para obtener T_{i+1} , primero se determinan $X_j^{(i+1)}$ y $Y^{(i+1)}$. Para $j = 1, \dots, m$ se realiza la regresión de las columnas $X_j^{(i)}$ contra las T_i , teniendo $t_i^t x_j^{(i)} / (t_i^t t_i)$ como los coeficientes de la regresión, y $X_j^{(i+1)}$ se define como:

$$X_j^{(i+1)} = V_j^{(i)} - \{t_i^t v_j^{(i)} / (t_i^t t_i)\} T_i, \quad (2.6)$$

con valores muestrales $x_j^{(i+1)}$. De manera similar $Y^{(i+1)}$ es definida como $Y^{(i+1)} = Y^{(i)} - \{t_i^t y^{(i)} / (t_i^t t_i)\} T_i$, y sus valores muestrales, $y^{(i+1)}$ son los residuales de la regresión de $Y^{(i)}$ contra T_i .

La variabilidad residual en Y es $Y^{(i+1)}$ y la información restante en X_j está en $X_j^{(i+1)}$, así que el siguiente paso es realizar la regresión de $Y^{(i+1)}$ contra cada $X_j^{(i+1)}$. La j -ésima regresión produce a $b_j^{(i+1)} X_j^{(i+1)}$ como un predictor de $Y^{(i+1)}$ donde:

$$b_j^{(i+1)} = x_j^{(i+1)t} y^{(i+1)} / (x_j^{(i+1)t} x_j^{(i+1)}), \quad (2.7)$$

y formando finalmente la nueva combinación lineal de los predictores:

$$T_{i+1} = \sum_j w_j^{(i+1)} b_j^{(i+1)} V_j^{(i+1)}.$$

Una característica de PLS es que la correlación muestral entre cualquier par de componentes es cero, esto es consecuencia de que los residuos de una regresión es-

tan descorrelacionadas con el regresor, así por ejemplo $X_j^{(i+1)}$ están descorrelacionadas con T_i , además como las componentes T_{i+1}, \dots, T_p son combinación lineal de las $X_j^{(i+1)}$ entonces están descorrelacionadas con T_i . Para completar el algoritmo el requisito de $\sum_{j=i}^k w_j^{(i)} = 1$ puede relajarse e igualar $w_j^{(i)}$ a $v_j^{(i)t} v_j^{(i)}$ para todos los pares i, j ; de esta manera $w_j^{(i)} \propto \text{Var}(X_j^{(i)})$ y las componentes $T_i = \sum_j (x_j^{(i)t} y^{(i)}) X_j^{(i)} \propto \sum_j \hat{Cov}(X_j^{(i)}, Y^{(i)}) X_j^{(i)}$.

Esta es la manera usual de estimar vía PLS, la forma anterior de elegir los pesos $w_j^{(i)}$, es que sean inversamente proporcionales a las varianzas de los coeficientes $b_j^{(i)}$, también si $\text{Var}(X_j^{(i)})$ es pequeña con relación a la varianza muestral de X_j , entonces X_j es aproximadamente colineal con respecto a los componentes T_1, \dots, T_k , entonces la contribución a T_i debe ser pequeña si $w_j^{(i)}$ es pequeño. La otra manera de elegir los pesos, es que sean iguales lo que permite que cada variable X_j tenga el mismo peso para realizar predicciones. En este punto es importante destacar que la primer forma de considerar los pesos tiene como consecuencia que las componentes sean invariantes bajo rotaciones ortogonales de los datos y la segunda forma las hace invariantes bajo escala de los datos \mathbf{X} y Y , [Garthwaite \(1994\)](#)

Decidir el número de componentes a incluir, en un modelo de regresión es un problema difícil, usualmente se puede efectuar por validación cruzada, ([Hastie et al., 2009](#)), o bien en el caso multivariado por medio de la prueba de razón de verosimilitudes, suponiendo una distribución normal de los componentes, ([Höskuldsson, 1988](#)).

Pasando al caso multivariado existen dos maneras de realizar la estimación, realizar la regresión PLS para cada Y_i o bien hacerlo de manera múltiple como reseñamos a continuación.

El algoritmo básico para la regresión de PLS multivariado fue desarrollado en [S. Wold C. Albano W. Dunn U. Edlund K. Esbensen P. Geladi S. Hellberg and Sjöström \(1984\)](#). Se parte de dos matrices escaladas y centradas \mathbf{X} de dimensiones $N \times M$ y \mathbf{Y} con dimensiones de $N \times K$. El escalamiento corresponde a trabajar con matrices de correlación. El algoritmo es como sigue:

Resultado: \mathbf{X} -loadings, \mathbf{Y} -loadings

1. Iguale u a la primer columna de \mathbf{Y}
2. $w = \mathbf{X}^t u / (u^t u)$
3. Escale w para que tenga norma unitaria
4. $t = \mathbf{X} w$
5. $c = \mathbf{Y}^t t / (t^t t)$
6. Escale c para que tenga norma unitaria
7. $u = \mathbf{Y}^t c / (c^t c)$;
8. Si se ha logrado la convergencia pase a 9, en otro caso valla a 2
9. \mathbf{X} -loadings: $p = \mathbf{X}^t t / (t^t t)$
10. \mathbf{Y} -loadings: $q = \mathbf{Y}^t u / (u^t u)$
11. Realice la regresión de (u contra t): $b = u^t t / (t^t t)$
12. Estime las matrices residuales:
 - $\mathbf{X} \leftarrow \mathbf{X} - t p^t$
 - $\mathbf{Y} \leftarrow \mathbf{Y} - b t c^t$

Algoritmo 1: PLS iterativo

En el algoritmo anterior la siguiente iteración comienza con las nuevas matrices \mathbf{X} y \mathbf{Y} como las matrices residuales de la previa. Las iteraciones continúan hasta que un criterio de paro se logra o bien \mathbf{X} llega a ser la matriz cero.

Puede demostrarse, Höskuldsson (1988), que en cada nueva iteración los vectores $u^{(n)}$, $c^{(n)}$, $t^{(n)}$ y $w^{(n)}$ son proporcionales a multiplicar su iteración anterior por las matrices $\mathbf{Y}\mathbf{Y}^t\mathbf{X}\mathbf{X}^t$, $\mathbf{Y}^t\mathbf{X}\mathbf{X}^t\mathbf{Y}$, $\mathbf{X}\mathbf{X}^t\mathbf{Y}\mathbf{Y}^t$ y $\mathbf{X}^t\mathbf{Y}\mathbf{Y}^t\mathbf{X}$ respectivamente. Lo anterior muestra que el algoritmo actúa de manera similar al *power method*, (Golub and Van-Loan, 1996), determinando el valor propio de cada matriz respectivamente. Cuando se logra la convergencia se tiene:

$$\begin{aligned} \mathbf{Y}\mathbf{Y}^t\mathbf{X}\mathbf{X}^t u &= a u \\ \mathbf{Y}^t\mathbf{X}\mathbf{X}^t\mathbf{Y} c &= a c \\ \mathbf{X}\mathbf{X}^t\mathbf{Y}\mathbf{Y}^t t &= a t \\ \mathbf{X}^t\mathbf{Y}\mathbf{Y}^t\mathbf{X} w &= a w, \end{aligned}$$

así los vectores u, c, t y w son los vectores propios correspondientes al máximo valor propio. Lo anterior tienen entre sus consecuencias las siguientes: los vectores w son ortogonales entre sí, es decir $w_i^t w_j = 0$ para $i \neq j$; los vectores t_i también lo son, $t_i^t t_j = 0$ para $i \neq j$; los vectores t_i forman una base ortogonal del espacio generado por las columnas de \mathbf{X} y los vectores w_i son ortogonales a los vectores p_j con $i < j$, (Höskuldsson, 1988).

Otra interpretación y derivación de las componentes de PLS es la de componentes con máxima covarianza. Considere dos componentes f y g en el espacio de \mathbf{X} y \mathbf{Y} respectivamente que cumplan lo siguiente:

$$\begin{aligned} f &= Xd, \quad |d| = 1 \\ g &= Ye, \quad |e| = 1, \end{aligned}$$

la covarianza muestral entre estas dos componentes está dada por $Cov(f, g) = f^t g / N$. Así con la notación del algoritmo 1 los vectores w y c satisfacen la siguiente maximización:

$$[Cov(t, u)]^2 = [Cov(Xw, Yc)]^2 = \max[Cov(f, g)]^2, \quad |d| = |e| = 1, \quad (2.8)$$

para probar lo anterior se recurre a la descomposición SVD, (Golub and Van-Loan, 1996), de la matriz $\mathbf{X}^t \mathbf{Y}$:

$$\mathbf{X}^t \mathbf{Y} = \sum_i a_i f_i g_i^t,$$

el valor singular a_1 tiene la interpretación:

$$(a_1)^2 = \max(d^t \mathbf{X}^t \mathbf{Y} e)^2, \quad |d| = |e| = 1,$$

donde el máximo se obtiene cuando $d = f_1$ y $e = g_1$, en el algoritmo 1 se tiene que $w = f_1$ y $c = g_1$. Por la igualdad 2.8 se tiene que la covarianza anterior es igual a la covarianza entre las componentes t y u por lo que se puede interpretar que ellas son las componentes del espacio de \mathbf{X} y \mathbf{Y} que tienen máxima covarianza sobre todas las componentes en este espacio.

Para concluir esta Sección sólo se menciona la existencia de otra derivación de las componentes de PLS la cual se aborda desde el punto de vista de la regresión por mínimos cuadrados pesados, es decir se tiene la matriz $\mathbf{X}^t \mathbf{V} \mathbf{X}$, donde la matriz de pesos \mathbf{V} se sustituye por la matriz $\mathbf{Y} \mathbf{Y}^t$, (Höskuldsson, 1988).

Con los elementos reseñados en este Capítulo tenemos los elementos necesarios para exponer la metodología VAR-PLS en el siguiente.

Metodología VAR-PLS

Como se mencionó en la introducción, la metodología para pronósticos conjuntos de series de tiempo VAR-PLS es una extensión del trabajo de [Frances \(2006\)](#) en el sentido de que se generaliza el número de componentes de la serie al igual que se dan intervalos de confianza para tales pronósticos, utilizando una adaptación del método Bootstrap.

En este capítulo detallamos primero la construcción de las estimaciones puntuales de los pronósticos conjuntos para diferentes horizontes $1, \dots, h$ y posteriormente tratamos la construcción de los intervalos de confianza.

El caso de interés de esta metodología es predecir una de las componentes de una serie de tiempo multivariada, como las reseñadas en la sección 2.2, de manera tan precisa como sea posible, aunque por su naturaleza multivariada se obtienen pronósticos para las demás variables que forman la serie. Las relaciones multivariadas existentes en un vector autorregresivo surgen al considerar la relación de una de las componente con el resto.

Como paso previo al comienzo de la metodología se sugiere verificar que el conjunto de variables a utilizar no presenten cambios estructurales, para esto podemos hacer uso de la prueba de Bussetti-Harvey, [Bussetti and Harvey \(2001\)](#), y descartar las series que presenten dichos cambios.

El primer paso en la metodología consiste en aplicar una transformación a la serie de tiempo para disminuir la varianza componente a componente.

Consideremos el vector k -variado $Z^*_{t+h,j}$ donde cada componente es de longitud $t+h$ y h es el horizonte a pronosticar. Para simplificar la notación consideremos dividir el vector anterior de tal forma que denotamos por Y^* a la matriz que contienen los elementos $Z^*_{t+h-i,j}$ con $i = 0, 1, 2, \dots, h-1$ y $j = 1, 2, \dots, k$, esta será la matriz que utilizaremos en la regresión PLS para pronosticar, así Y^* es una matriz con las t observaciones más recientes y con k variables; y como X^* a las observaciones $Z^*_{t+h-i,j}$

con $i = h, \dots, t + h - 1$ y $j = 1, 2, \dots, k$, por lo que la matriz X^* de dimensiones (t, k) contiene las primeras t observaciones de la serie de tiempo.

Después de dividir la muestra tal como se menciona en el párrafo anterior, se determina el orden del modelo $VAR(p)$ utilizando el criterio de FPE, reseñado en la sección 2.2; luego de determinado el orden p se realiza una prueba para sustentar la hipótesis nula de que el modelo VAR es cointegrado.

Varias pruebas de cointegración se han desarrollado desde que el concepto nació. Entre estas se encuentran la de dos pasos de Engle y Granger, [Engle and Granger \(1987\)](#), la prueba de Phillips–Ouliaris, [Phillips and Ouliaris \(1990\)](#), y la famosa -y más usada- se debe al mismo Johansen, [Johansen \(1991\)](#). Estas pruebas pueden emplearse en la metodología para verificar la cointegración del proceso VAR, sin embargo tienen la desventaja técnica -más no conceptual- de solo poder ser usadas con menos de 12 componentes. Tal desventaja técnica se debe a que en los trabajos de Johansen los valores críticos de los respectivos estadísticos solo fueron calculados hasta 11 series y para tres niveles de significancia.

Sin embargo, si el número de componentes del proceso VAR que se analiza es mayor a 11, como es el caso del ejercicio empírico que se realiza en el siguiente capítulo, puede emplearse el procedimiento descrito en [Carlomagno \(2017\)](#) el cual no tiene limitante en el número de variables pero su complejidad computacional incrementa exponencialmente. Este procedimiento posee diversas propiedades muy útiles para la metodología VAR-PLS:

- Cuando el tamaño de muestra disminuye o el número de componentes se incrementa, este procedimiento provee mejores resultados que la prueba de Johansen. Lo cual justifica a la metodología tratar con conjuntos de datos que tienen más variables que observaciones, High Dimension Low Sample Size (HDLSS), [Jung and Marron \(2009\)](#).
- Debido a que la potencia de la prueba de Johansen disminuye al aumentar el número de componentes, [Lütkepohl \(2006\)](#), este procedimiento tiene mayor potencia pues emplea pruebas de cointegración de solo dos componentes.
- El procedimiento no se ve sustancialmente afectado por la asimetría o curtosis de los residuos.

El procedimiento planteado en [Carlomagno \(2017\)](#), parte de considerar a cada componente del vector X_i como un nodo de una gráfica, al considerar los posibles $\binom{k}{2}$ pares de nodos estos se unen por medio de una arista si la prueba de cointegración de Johansen para este par de variables establece que hay cointegración.

En caso de que los $\binom{k}{2}$ pares de componentes sean cointegrados, utilizando la prueba de Johansen, [Carlomagno \(2017\)](#) demuestra que el vector con las k componentes

también es cointegrado.

En caso de que la gráfica inducida por los nodos y vértices resultado de la acción anterior no sea completa se pueden encontrar conjuntos de componentes que son cointegradas, por medio de un algoritmo para encontrar clanes maximales en una gráfica, lo que en nuestra metodología se emplearía como un criterio para determinar variables cointegradas en conjunto.

Una vez que se conoce el orden p del modelo VAR y que se realiza la prueba de cointegración se construyen las siguientes matrices, con las columnas de Y^* y X^* :

$$X = (X_{t-1,1}^* \quad \cdots \quad X_{t-1,k}^* \quad X_{t-2,1}^* \quad \cdots \quad X_{t-2,k}^* \quad \cdots \quad X_{t-p,1}^* \quad \cdots \quad X_{t-p,k}^*)$$

$$Y = (Y_{t+h,1}^* \quad \cdots \quad Y_{t+h,k}^* \quad Y_{t+h-1,1}^* \quad \cdots \quad Y_{t+h-1,k}^* \quad \cdots \quad Y_{t,1}^* \quad \cdots \quad Y_{t,k}^*).$$

Nótese que tanto las nuevas matrices X y Y tienen ambas t observaciones sin embargo las dimensiones de X son $t \times ((p+1) \times k)$ y las de Y son $t \times (h \times k)$. Es decir que X se construye con las columnas de X^* concadenando las mismas columnas, pero primero con un rezago, luego con dos rezagos hasta p rezagos y de manera análoga Y se construye con las columnas de Y^* uniendo a la derecha los h primeros rezagos.

3.1 Estimación de los h pronósticos conjuntos

Tomando como punto de partida las matrices X y Y mencionadas en la sección anterior procedemos a realizar la estimación del pronóstico conjunto para los h horizontes de la serie de tiempo multivariada.

Como en el contexto de PLS, consideremos que tenemos l variables latentes por lo que utilizando la regresión por PLS podemos factorizar las matrices X y Y como sigue:

$$\begin{aligned} X &= KW \\ Y &= KBC, \end{aligned}$$

donde la matriz K , de dimensiones $t \times l$, es la matriz de *scores* y W , con dimensiones $l \times ((p+1) \times k)$, es la matriz de *loadings*. Análogamente C es una matriz de *loadings*, con dimensiones $l \times (h \times k)$, y B es una matriz de pesos, de dimensiones $l \times l$. En [Frances \(2006\)](#) no se especifica como calcular B , sin embargo, se puede inferir que es el resultado de la multiplicación por la izquierda de la inversa generalizada K^\dagger con L , donde L es la matriz de *scores* de la descomposición $Y = LC$.

En la descomposición anterior por medio de PLS puede emplearse cualquier algoritmo, ya sea el que es invariante a escalas de los datos o bien el que produce vectores

de cargas ortogonales. En particular es recomendable centrar y escalar los datos para incrementar la precisión numérica así que tendremos matrices de cargas ortogonales.

Finalmente para pronosticar h horizontes con un número de componentes fija l se efectúa el siguiente producto para obtener los pronósticos \hat{Y} , donde tenemos estimaciones para las k componentes en los siguientes h pasos:

$$\hat{Y} = X\hat{B}_{PLS},$$

donde \hat{B}_{PLS} es el producto $W^\dagger BC$ con dimensiones $((p+1) \times k) \times l$. Nótese que los h pronósticos de las k componentes se encuentran en el primer renglón de \hat{Y} .

De manera análoga a los modelos $AR_j(p)$ de los que se habló en la introducción se realiza el pronóstico de la técnica anterior un paso a la vez, y estimando las matrices de parámetros correspondientes e incorporando el nuevo valor predicho a la muestra y descartando el más viejo, de esta forma aprovechamos la correlación entre las observaciones.

El procedimiento anterior requiere de fijar el número de componentes o variables latentes l . En la práctica, podemos buscar en el dominio de las componentes $l = 1, 2, \dots, ((p+1) \times k)$ el valor que minimice una medida de error de los pronósticos. En una implementación o codificación puede surgir el detalle de que el extremo superior de l no sea precisamente $(p+1) \times k$ pues la última columna de la derecha de X puede contener valores nulos que afecten a la estimación por PLS acotando a $l < ((p+1) \times k)$.

3.2 Construcción de los intervalos de confianza

Una vez que se determina el número de componentes, l , de la regresión PLS, que minimice algún criterio de error de los pronósticos, se procede a construir intervalos de confianza para tales pronósticos.

En [Pascual et al. \(2011\)](#) se propone una manera de construir intervalos de confianza para pronósticos de modelos VAR utilizando una adaptación del método Bootstrap. Tal procedimiento se puede adecuar a la metodología VAR-PLS para obtener intervalos de confianza. Tal adaptación sigue los siguientes pasos:

1. Estimar los coeficientes de la matriz \hat{B}_{PLS} y con ellos obtener los residuos, \hat{a}_t , de $Y - \hat{Y}$, donde $\hat{Y} = X\hat{B}_{PLS}$. Denotemos por \hat{F}_a a la función de distribución empírica de los residuos \hat{a}_t centrados y multiplicados por el factor $[(t-p)/(t-2p)]^{(1/2)}$, recomendado en [Stine \(1987\)](#).
2. De un conjunto inicial de p observaciones de la matriz X , $X_t^{(*)} = X_{t,j}$ con $j = 1, \dots, (k \times (p+1))$, construir una serie de Bootstrap $\{Y_1^{(*)}, Y_2^{(*)}, \dots, Y_t^{(*)}\}$ como

sigue:

$$Y_i^{(*)} = X_{i-1}^{(*)} \hat{B}_{PLS} + \hat{a}_{i-1}^*, \quad i = 1, \dots, t,$$

donde \hat{a}_t^* es una muestra independiente de \hat{F}_a . Una vez que se tienen las t observaciones estimamos nuevamente \hat{B}_{PLS}^* , una réplica de Bootstrap del ajuste del paso anterior.

3. Pronosticar utilizando PLS con los nuevos parámetros \hat{B}_{PLS}^* .
4. Repetir los pasos 1 a 3, R veces.

De esta manera obtenemos R réplicas de las observaciones contenidas en la matriz Y , con las cuales podemos obtener intervalos de confianza para la n -ésima componente por medio de los percentiles:

$$C.I._{t+h} = \{y_{t+k,n} | y_{t+k,n} \in [q_B^*(\tau), q_B^*(1-\tau)]\},$$

donde $q_B^*(\tau)$ es el percentil τ de la función de distribución empírica generada por las R réplicas Bootstrap de la n -ésima componente, $G_{B,n}^*$.

En este punto se ha concluido de exponer la metodología VAR-PLS. En el capítulo siguiente realizamos un ejercicio empírico a manera de ejemplo donde se busca pronosticar la inflación mexicana utilizando la metodología expuesta.

Antes de finalizar este capítulo destacaremos que los pasos de la metodología como el procedimiento descrito en [Carlomagno \(2017\)](#), al igual que la determinación del número de componentes y la construcción de los intervalos de confianza son procedimientos que se pueden paralelizar en diferentes arquitecturas de cómputo como la multihilo o la de memoria compartida.

La metodología se implementó en el ambiente de cómputo estadístico R, [R Core Team \(2018\)](#), y una codificación que hace uso de paralelismo en una arquitectura multihilo puede ser consultada en <https://github.com/fou-foo/Tesis.git>

Resultados

Con la metodología detallada en el capítulo anterior realizaremos en este un ejercicio empírico para medir la calidad de los pronósticos de la inflación mexicana¹ en conjunto con otras 14 variables macroeconómicas.

4.1 Los datos

Como se mencionó en la sección 2.2.1 existen los problemas entre las variables *verdaderas* -desde el punto de vista económico- y las variables observables. Además Juselius (2007) menciona, que los datos de las variables macroeconómicas tienen una fuerte dependencia temporal que sugiere una formulación basada en procesos estocásticos. Es útil distinguir entre:

- Variables estacionarias con una dependencia temporal a corto plazo.
- Variables no-estacionarias con dependencia temporal a largo plazo.

El orden de integración de una variable no es en general una propiedad económica de la variable sino una conveniente aproximación estadística para distinguir entre variaciones a corto, mediano o largo plazo en los datos.

En el caso de la inflación, de acuerdo con Juselius (2007), existen argumentos a favor de considerarla una raíz unitaria con tendencia estocástica.

La presencia y determinación de raíces unitarias en series de tiempo de variables macroeconómicas es un tema ampliamente estudiado. En el segundo capítulo de Juselius (2007) se puede encontrar un ejemplo que involucra a la inflación anual de Dinamarca observada en los diferentes intervalos: 1901-1992, 1945-1992 y 1975-1992, donde para

¹Nótese que la inflación es función del Índice Nacional de Precios al Consumidor, variable pronosticada en este trabajo.

los dos primeros se tiene una serie estacionaria y en el tercero -subconjunto del segundo- la estacionariedad se pierde.

Las variables macroeconómicas mensuales que se utilizan en el siguiente apartado son las siguientes donde después de su descripción se añaden las opciones para su descarga (cuando es necesario) de la url que se encuentra en la bibliografía:

1. Índice Nacional de Precios al Consumidor (INPC nacional): Mide la variación de los precios de la canasta de bienes y servicios representativa de los hogares mexicanos (INEGI, 2019). Un aumento de esta variable conlleva costos sociales enormes para un país, ya que afecta: el consumo, la inversión y el empleo, (Huertas, 2019).

Indicadores económicos de coyuntura>Índices de Precios>

Índice nacional de precios al consumidor

2. Salarios mínimos: Índice correlacionado con la cantidad mínima de pesos por día que un empleado recibe por jornada laboral, (Banxico, 2019a).
3. Índice Nacional de Precios al Consumidor de los Estados Unidos: Variable análoga a la mexicana pero para una canasta diferente. Sin ajuste estacional, (United States Department Of Labor, 2019).
4. Tipo de cambio Pesos por dólar E.U.A.: Para solventar obligaciones denominadas en moneda extranjera, Fecha de liquidación Cotizaciones promedio, (Banxico, 2019b).
5. Índice de producción industrial de los Estados Unidos: Datos sobre productos físicos que se obtienen de asociaciones comerciales privadas y de agencias gubernamentales. Los datos sobre las horas trabajadas por los trabajadores de producción se recopilan en la encuesta mensual de establecimientos realizada por la Oficina de Estadísticas Laborales, (Organización para la Cooperación Económica y el Desarrollo, 2019)

Industry and Services>Industry and Service Statistics(MEI)>

Production and Sales (MEI)>Production

6. IGAE de actividades económicas primarias: Segmento del Índice Global de la Actividad Económica correspondiente a actividades primarias, (INEGI, 2019).

Indicadores económicos de coyuntura>

Indicador global de la actividad económica, base 2013>Series originales>

Índice de volumen físico>Actividades primarias (Periodicidad: Mensual)

-
7. IGAE de actividades económicas secundarias: Segmento del Índice Global de la Actividad Económica correspondiente a actividades secundarias, (INEGI, 2019).

Indicadores económicos de coyuntura>

Indicador global de la actividad económica, base 2013>Series originales>

Índice de volumen físico>Actividades secundarias>

Total actividades secundarias (Periodicidad: Mensual)

8. IGAE de actividades económicas terciarias: Segmento del Índice Global de la Actividad Económica correspondiente a actividades terciarias, (INEGI, 2019).

Indicadores económicos de coyuntura>

Indicador global de la actividad económica, base 2013>Series originales>

Índice de volumen físico>Actividades terciarias>

Total actividades terciarias (Periodicidad: Mensual)

9. Minería: Componente del indicador de actividades secundarias correspondiente a las actividades derivadas de la minería, (INEGI, 2019).

Indicadores económicos de coyuntura>

Indicador global de la actividad económica, base 2013>Series originales>

Índice de volumen físico>Actividades secundarias>

21 Minería (Periodicidad: Mensual)

10. Manufactura: Componente del indicador de actividades secundarias correspondiente a las actividades derivadas de la industria manufacturera, (INEGI, 2019).

Indicadores económicos de coyuntura>

Indicador global de la actividad económica, base 2013>Series originales>

Índice de volumen físico>Actividades secundarias>

31-33 Industrias manufactureras (Periodicidad: Mensual)

11. Construcción: Componente del indicador de actividades secundarias correspondiente a las actividades derivadas de la industria de la construcción, (INEGI, 2019).

Indicadores económicos de coyuntura>

Indicador global de la actividad económica, base 2013>Series originales>

Índice de volumen físico>Actividades secundarias>

23 Construcción (Periodicidad: Mensual)

12. Energía: Componente del indicador de actividades secundarias correspondiente a las actividades derivadas de la generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final, (INEGI, 2019).

Indicadores económicos de coyuntura>

Indicador global de la actividad económica, base 2013>Series originales>

Índice de volumen físico>Actividades secundarias>

Generación, transmisión y distribución de energía eléctrica,
suministro de agua y de gas por ductos al consumidor final

13. Tasa de Desempleo Abierta: Porcentaje de la población mexicana mayor a 15 años que no tiene empleo, (INEGI, 2019).

Ocupación, empleo y remuneraciones> Tasas de ocupación, desocupación y subocupación
(resultados mensuales de la ENOE, 15 años y más)> Nacional>

Población total>Población desocupada

14. Ingreso: Indicador Global de la Actividad Económica. Permite conocer y dar seguimiento a la evolución del sector real de la economía en el corto plazo, (INEGI, 2019).

Indicadores económicos de coyuntura>Indicador global de la actividad económica, base 2013>Series originales>Índice de volumen físico>

Total (Periodicidad: Mensual)

15. Billetes y monedas: Fuentes y usos de la base monetaria, Usos, Billetes y monedas en circulación, medidos en miles de pesos, (Banxico, 2019c).
16. TIIE 28: Tasa de Interés Interbancaria a 28 días, medida en porcentaje, (Banxico, 2019d).
17. TIIE 91: Tasa de Interés Interbancaria a 91 días, medida en porcentaje, (Banxico, 2019d).

4.2 Aplicación empírica

Aplicaremos la metodología VAR-PLS con las 17 variables mensuales de la sección anterior para realizar pronósticos sobre la inflación mexicana sobre los últimos 12 meses con los que se dispone, la variable INPC. Para ello utilizamos las series mensuales a partir de enero del 2005, pues la variable relativa al desempleo inicia en este punto, hasta marzo de 2019 que es la última fecha de actualización de las variables relativas a las actividades económicas (primarias, secundarias y terciarias). Con lo que contamos con 171 observaciones de las cuales utilizaremos las primeras 159 para pronosticar las últimas 12.

En la figura 4.1 se pueden apreciar las series de tiempo con las que se trabaja. Las series presentan diferentes escalas y comportamientos, por ejemplo, las series sobre la construcción y la energía aparentan tener estacionalidad. La variable de inflación exhibe una tendencia a la alza al igual que la mayoría de las series incluyendo las dos tasas interbancarias que presentan una alza reciente después de un periodo decreciente, aunado a lo anterior al aplicar la prueba de Bussetti-Harvey para el mes de Agosto del 2016 (nuestra observación número 140) se tiene que las dos series de tiempo de tasas interbancarias presentan cambios estructurales por lo que las descartamos del conjunto de variables.

Al inicio aplicamos la función logaritmo a cada uno de las componentes, después de utilizar el procedimiento descrito en [Carlomagno \(2017\)](#) no se rechaza la hipótesis nula de cointegración y al estimar el orden p del modelo VAR se obtiene un rezago de 5 unidades utilizando el criterio FPE.

Con lo anterior las matrices X y Y descritas en el capítulo anterior tienen $(p + 1) \times k = 6 * 15 = 90$ y $k \times h = 15 * 12 = 180$ columnas respectivamente, por lo cual existen 180 posibles elecciones de número de componentes para la regresión PLS. Se utiliza el error porcentual absoluto medio, mean absolute percentage error (MAPE), como criterio para determinar el número de componentes. En la figura 4.2 se puede apreciar el MAPE para cada una de las 15 variables, en particular como es de interés pronosticar la inflación se elige el número de componentes que minimicen el MAPE para esta componente, el cual es 4.

Una vez que se ha determinado el número de componentes que minimiza el MAPE, se procede a realizar 10,000 réplicas Bootstrap para construir los intervalos de confianza.

En la figura 4.3 se presentan los resultados de los pronósticos puntuales junto al intervalo de confianza (al 95%) construido con los percentiles de Bootstrap. En este punto vale la pena resaltar que los pronósticos puntuales sobreestiman la inflación en los primeros 4 meses próximos (mayo, junio, julio y agosto de 2018) mientras que la subestima para los últimos 7 meses, además los intervalos de confianza contienen a los

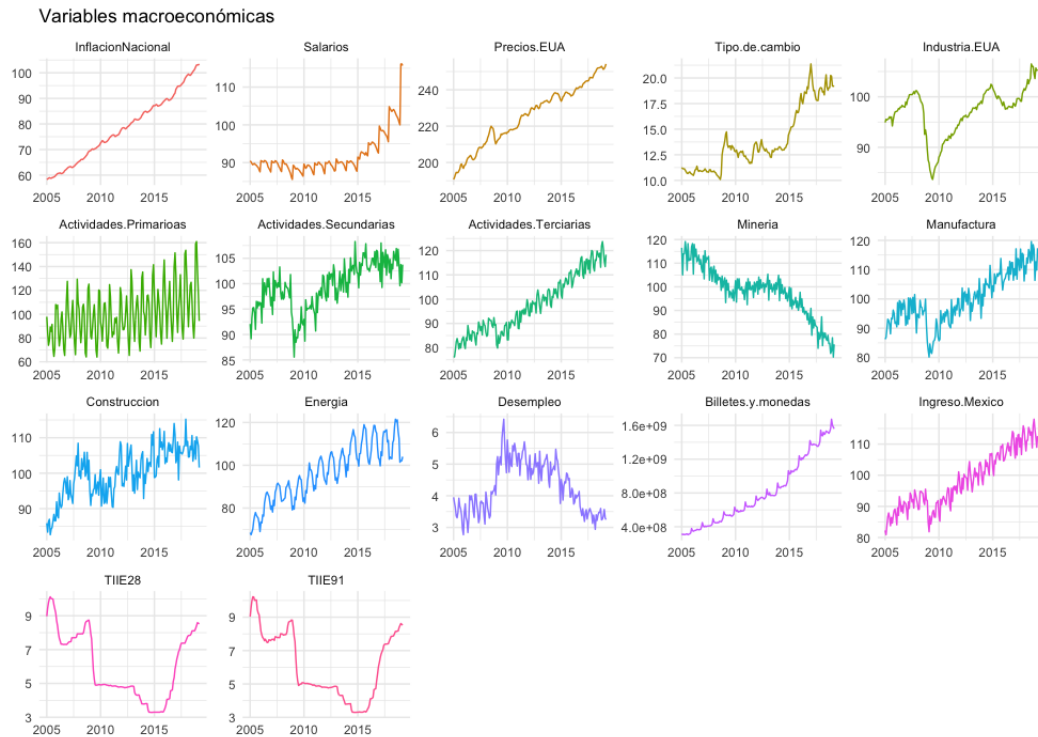


Figura 4.1: Las 17 series con las que se ilustra la metodología.

valores verdaderos de la serie pero la longitud de los mismo no se incrementa de igual forma que lo hacen los típicos intervalos de confianza para pronósticos $VAR(p)$.

Para evaluar el desempeño de la metodología, estimamos un solo modelo $VAR(5)$ con la muestra de las primeras 159 observaciones y predecimos los siguientes 12 meses. Los resultados de ambos métodos se muestran en la tabla 4.1, donde se observa que la metodología propuesta VAR-PLS se desempeña mejor que el modelo VAR, obteniendo errores relativos menores.

Finalmente en la tabla 4.2 mostramos la estimación de los intervalos de confianza para la metodología VAR-PLS.

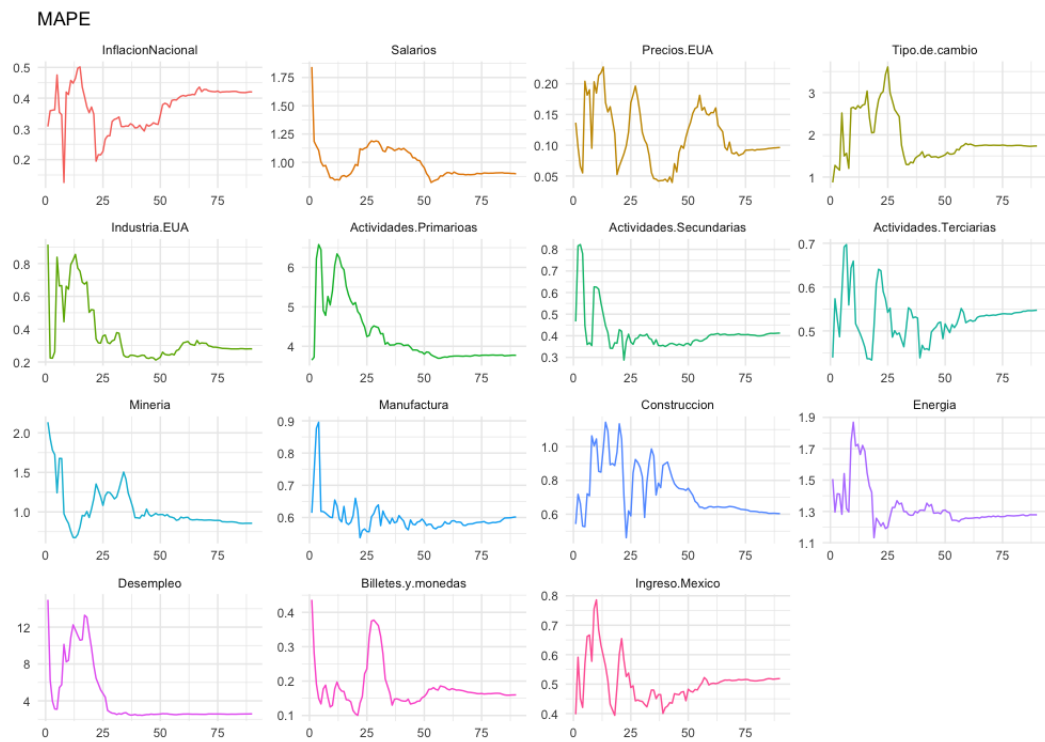


Figura 4.2: MAPE para cada una de las series con las que se trabaja, en el eje vertical se encuentra la medida de error y en el eje horizontal el número de componentes consideradas en el pronóstico VAR-PLS.

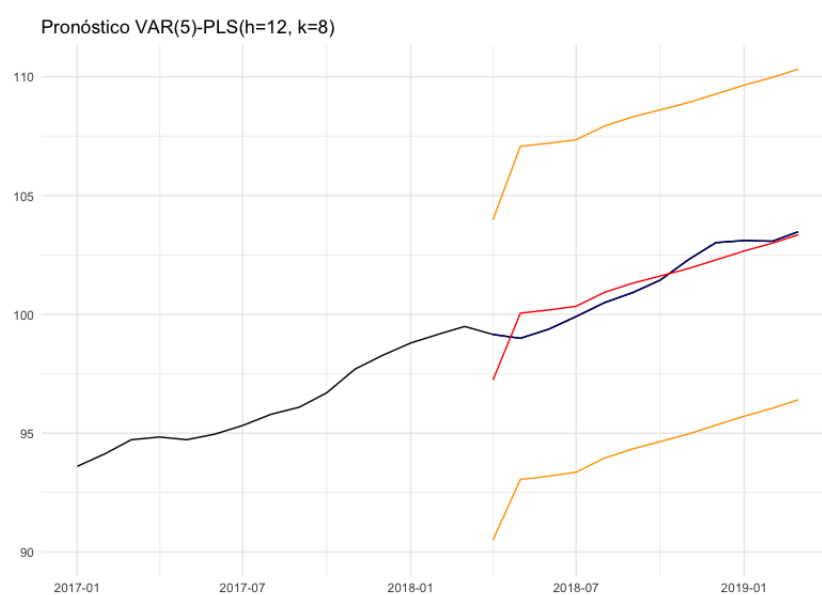


Figura 4.3: Pronósticos de la metodología VAR-PLS para para el Índice Nacional de Precios al Consumidor en los últimos 12 meses de la muestra (abr. 2018 a mar. 2019). La línea roja corresponde al pronóstico, mientras que la negra es el valor real, por su parte las líneas naranjas forman el intervalo de confianza de los pronósticos obtenidos por Bootstrap con una confianza de 95 %.

	Valor Real	VAR-PLS	VAR	Precisión VAR-PLS	Precisión VAR
1	99.15	97.24	102.82	98.07	96.31
2	98.99	100.05	101.86	98.93	97.11
3	99.38	100.19	101.55	99.19	97.81
4	99.91	100.34	101.68	99.57	98.22
5	100.49	100.93	102.02	99.57	98.48
6	100.92	101.31	102.24	99.61	98.69
7	101.44	101.61	102.63	99.83	98.83
8	102.30	101.93	103.69	99.64	98.65
9	103.02	102.29	104.44	99.29	98.62
10	103.11	102.67	104.77	99.57	98.38
11	103.08	103.00	104.74	99.92	98.39
12	103.48	103.35	105.14	99.88	98.39

Tabla 4.1: Estimaciones y precisión de los pronósticos de la metodología VAR-PLS y VAR para cada uno de los doce horizontes pronosticados.

h	VAR-PLS	2.5 %	97.5 %
V1	97.24	90.50	103.98
V2	100.05	93.05	107.07
V3	100.19	93.19	107.20
V4	100.34	93.36	107.35
V5	100.93	93.95	107.93
V6	101.31	94.34	108.31
V7	101.61	94.65	108.61
V8	101.93	94.97	108.92
V9	102.29	95.34	109.27
V10	102.67	95.71	109.65
V11	103.00	96.05	109.98
V12	103.35	96.40	110.32

Tabla 4.2: Intervalos de confianza al 95 % para cada uno de los doce horizontes pronosticados, obtenidos por Bootstrap, de la metodología VAR-PLS.

Conclusiones

A lo largo de este trabajo revisamos los conceptos necesarios para la exposición detallada de una nueva metodología para realizar pronósticos conjuntos de una serie de tiempo multivariada para diferentes horizontes.

Entre los retos que surgieron y que la metodología resuelve se encuentran: el manejo de series multivariadas y cointegradas con dimensión mayor a 11. Los resultados de la metodología son prometedores en un contexto macroeconómico como es ilustrado con un ejercicio empírico donde se pronosticó la inflación mexicana en 12 meses recientes con precisión mayor al 99%.

La mayor parte de los pasos de la metodología como la determinación de cointegración, la elección del número de componentes en la regresión PLS y la construcción de los intervalos de confianza son procedimientos que se pueden implementar en una arquitectura multihilo.

Como trabajo futuro se plantean extensiones conceptuales de la metodología como: la extensión al caso en que las matrices X y Y , como se definen en el tercer capítulo, no sean densas. Durante la implementación surgieron detalles que sugieren extensiones como la imputación de datos cuando surgen valores nulos en la construcción de las matrices X y Y .

Aunque la metodología es conceptualmente útil en casos en que existan más variables que observaciones falta verificarlo empíricamente. Se mantiene la sugerencia planteada en [Frances \(2006\)](#), sobre el desarrollo de una metodología PLS-VAR, que pronostique un VAR utilizando PLS pero sin la estimación directa del modelo VAR además de utilizar métodos Monte Carlo para verificar las bondades de la metodología VAR-PLS para diferentes procesos generadores de datos.

Lista de Figuras

4.1. Las 17 series con las que se ilustra la metodología.	33
4.2. MAPE para cada una de las series con las que se trabaja, en el eje vertical se encuentra la medida de error y en el eje horizontal el número de componentes consideradas en el pronóstico VAR-PLS.	34
4.3. Pronósticos de la metodología VAR-PLS para para el Índice Nacional de Precios al Consumidor en los últimos 12 meses de la muestra (abr. 2018 a mar. 2019). La línea roja corresponde al pronóstico, mientras que la negra es el valor real, por su parte las líneas naranjas forman el intervalo de confianza de los pronósticos obtenidos por Bootstrap con una confianza de 95 %.	35

Lista de Tablas

4.1. Estimaciones y precisión de los pronósticos de la metodología VAR-PLS y VAR para cada uno de los doce horizontes pronosticados.	36
4.2. Intervalos de confianza al 95 % para cada uno de los doce horizontes pronosticados, obtenidos por Bootstrap, de la metodología VAR-PLS. . .	37

Bibliografía

- Banxico (2019a). *Sistema de Información Económica*. <http://www.banxico.org.mx/SieInternet/consultarDirectorioInternetAction.do?sector=10&accion=consultarCuadroAnalitico&idCuadro=CA601&locale=es> Consultado: 2019-06-16. 29
- Banxico (2019b). *Sistema de Información Económica*. <http://www.banxico.org.mx/SieInternet/consultarDirectorioInternetAction.do?sector=6&accion=consultarCuadro&idCuadro=CF86&locale=es> Consultado: 2019-06-16. 29
- Banxico (2019c). *Sistema de Información Económica*. <http://www.banxico.org.mx/SieInternet/consultarDirectorioInternetAction.do?sector=11&accion=consultarCuadro&idCuadro=CF315&locale=es> Consultado: 2019-06-16. 31
- Banxico (2019d). *Sistema de Información Económica*. <http://www.banxico.org.mx/SieInternet/consultarDirectorioInternetAction.do?sector=18&accion=consultarCuadro&idCuadro=CF113&locale=es> Consultado: 2019-06-16. 31
- Braeken, J. and Assen, V. (2017). An empirical kaiser criterion. *Psychological Methods*, 22(3):450–466. 17
- Brockwell, P. J. and Davis, R. A. (1986). *Time Series: Theory and Methods*. Springer-Verlag, Berlin, Heidelberg. 4, 5
- Buseti, F. and Harvey, A. (2001). Testing for the presence of a random walk in series with structural breaks. *Journal of Time Series Analysis*, 22(2):127–150. 23
- Carlomagno, G. (2017). *Discovering common trends in a large set of disaggregates: statistical procedures and their properties*. PhD thesis, Department of Statistics and Econometrics, University Carlos III of Madrid. 24, 27, 32
- Chan, N. H. (2010). *Time Series: Applications to Finance with R and S-Plus(R)*. Wiley Series in probability and Statistics. Wiley, 2nd edition. 5, 6, 7
- Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276. 24

- Frances, P. H. (2006). Forecasting 1 to h steps ahead using partial least squares. *Econometric Institute Report*, 47. [1](#), [2](#), [23](#), [25](#), [38](#)
- Garthwaite, P. H. (1994). An interpretation of partial least squares. *Journal of the American Statistical Association*, 89(425):122–127. [17](#), [20](#)
- Golub, G. H. and Van-Loan, C. F. (1996). *Matrix computations*. Johns Hopkins studies in the mathematical sciences. Johns Hopkins University Press, 3rd ed edition. [16](#), [21](#), [22](#)
- Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16:121–10. [13](#)
- Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica*, 12(Supplement):1–118. [12](#)
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, 1 edition. [7](#)
- Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction, 2nd Edition*. Springer series in statistics. Springer. [15](#), [17](#), [20](#)
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(179-85). [17](#)
- Huertas, C. (2019). *Metodologías Competitivas para la predicción de la Inflación en México (Tesis de licenciatura inédita)*. Universidad Nacional Autónoma de México; CDMX, México. [29](#)
- Höskuldsson, A. (1988). Pls regression method. *Journal of Chemometrics*, 2:211 – 228. [20](#), [21](#), [22](#)
- INEGI (2019). *Banco de Información Económica*. Instituto Nacional de Estadística y Geografía. <https://www.inegi.org.mx/sistemas/bie/> Consultado: 2019-06-16. [29](#), [30](#), [31](#)
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580. [24](#)
- Jung, S. and Marron, J. S. (2009). Pca consistency in high dimension, low sample size context. *The Annals of Statistics*, 37(6B):4104–4130. [24](#)
- Juselius, K. (2007). *The Cointegrated VAR Model: Methodology and Applications*. Advanced Texts in Econometrics. Oxford University Press, USA, 2 edition. [2](#), [12](#), [13](#), [28](#)
- Kaiser, H. (1959). The application of electronic computers to factor analysis. *Meeting of Amer. Psychol. Ass.*, (Supplement):1–118. [17](#)

- Lütkepohl, H. (2006). *New Introduction To Multiple Time Series Analysis*. Springer. 11, 24
- Organización para la Cooperación Económica y el Desarrollo (2019). *OECD Stats*. <https://stats.oecd.org/> Consultado: 2019-06-16. 29
- Pascual, L., Ruiz, E., and Fresoli, D. (2011). Bootstrap forecast of multivariate var models without using the backward representation. *Statistics and Econometrics Series*, Working Paper 11-34. 26
- Phillips, P. C. B. and Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica*, 58(1):165–193. 24
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 27
- Roweis, S. (1998). Em algorithms for pca y spca. *Proceeding NIPS '97 Proceedings of the 1997 conference on Advances in neural information processing systems*, 10:626–632. 17
- S. Wold C. Albano W. Dunn U. Edlund K. Esbensen P. Geladi S. Hellberg, E. J. W. L. and Sjostrorn, M. (1984). *Multivariate Data Analysis in Chemistry, in Chemometrics*. Mathematics and Statistics in Chemistry. B. R. Kowalski, Reidel Publishing Company. 20
- Sax, C. and Eddelbuettel, D. (2018). Seasonal adjustment by X-13ARIMA-SEATS in R. *Journal of Statistical Software*, 87(11):1–17. 5
- Sheskin, D. J. (2000). *Handbook of parametric and nonparametric statistical procedures*. Chapman Hall/CRC, 2nd ed edition. 11
- Stine, J. (1987). Estimating properties of autoregressive forecasts. *Journal of Economic Perspectives*, 15(4):101–115. 26
- United States Department Of Labor (2019). *CPI-All Urban Consumers (Current Series)*. <https://beta.bls.gov/dataViewer/view/timeseries/CUUR0000SA0> Consultado: 2019-06-16. 29