

Proyecto de tesis: Impacto de las variaciones de los precios
internacionales de los principales bienes primarios sobre los
distintas clases de la inflación uruguaya

Laura Marta Montaldo Iglesias
C.I.: 3.512.962-7
Maestría en ingeniería matemática

9 de octubre de 2023

Cantidad de hojas:

Índice

Descripción del problema

En este estudio se busca analizar en qué medida las perturbaciones en los precios internacionales de ciertos bienes transables denominados *commodities* afectan a la inflación uruguaya a través de sus distintos componentes.

La economía uruguaya presenta ciertas características particulares. Es una economía pequeña y abierta, tomadora de precios en los mercados internacionales y exportadora de bienes primarios. Según fuentes oficiales ¹, en el año 2020, mientras que los tres principales bienes de exportación se concentraban alrededor de productos cárnicos, celulosa y habas, los tres principales productos de importación en orden descendente fueron los combustibles, maquina y equipos de transporte. Por estos motivos, sería interesante evaluar el grado de vulnerabilidad de la misma frente a los cambios en los precios de algunos de los principales bienes primarios transables a nivel global.

El presente estudio se enmarca en la literatura relacionada a los modelos dinámicos de factores. En un primer momento, se va a emplear un modelo de vectores autorregresivos aumentado por factores, conocido como FAVAR². Éstos modelos son una extensión de los modelos VAR ³ y son útiles cuando se cuenta con un gran número de variables, porque permiten hacer frente al problema de la *maldición de dimensionalidad* presente en los modelos VAR. Posteriormente se van a comparar los resultados de éste enfoque clásico con los de un modelo más reciente de factores dinámicos profundo (D^2FM)⁴. Este tipo de modelos se ubica en el marco de redes neuronales profundas permitiendo no linealidades entre los factores y las variables observables. Se va a analizar en qué medida éste modelo de última generación es mejor en desempeño en relación a uno clásico.

Para la estrategia empírica se emplean datos del índice de precios domésticos y precios internacionales de determinados bienes primarios. En cuanto a los variables domésticas se utilizan las ochenta y ocho clases del IPC de Uruguay, extraídas del Instituto Nacional de Estadística (INE). Con respecto a los precios internacionales de los bienes primarios, se utilizan precios de tres bienes primarios (carne, soja y petróleo) y el promedio simple del precio de los alimentos. Los mismos son tomados del Fondo Monetario Internacional. Para el precio de la carne se toma la serie correspondiente a la carne de Australia y Nueva Zelanda en centésimos de dólar por tonelada, para el petróleo se toma el promedio de tres precios spot expresados en dólares por barril, para la soja se toma el precio futuro de la soja en Chicago expresado en dólares por tonelada métrica y para los alimentos se realiza un promedio simple de los precios de commodities alimenticios disponibles.

¹Ficha País, Uruguay XXI

²Factor-Augmented VAR model

³Vector autoregression model

⁴Deep Dynamic Factor Model

Marco Conceptual

El presente análisis se enmarca teóricamente en lo que W. Enders (2014) denomina Modelos multivariados, que permiten analizar las interrelaciones entre múltiples series temporales. El enfoque dominante en la literatura para analizar el efecto de perturbaciones sobre determinadas variables endógenas es el de modelos autorregresivos VAR. Sin embargo, los modelos VAR padecen el problema de la *maldición de la dimensionalidad* cuando se cuenta con una gran cantidad de series. Éste es un concepto introducido por Bellman (1961) que indica que la cantidad de muestras necesarias para estimar una función arbitraria dado el grado de precisión, crece de manera exponencial con respecto a la cantidad de variables de entrada de la función. La maldición de la dimensionalidad significa que el número de objetos del conjunto de datos a los que hay que acceder crece exponencialmente con la dimensionalidad subyacente. Por lo tanto, cuando se dispone de un número elevado de variables explicativas, como es el caso para este estudio, los modelos VAR ya no son útiles porque sería necesario dejar fuera del análisis variables relevantes al mismo para poder preservar los grados de libertad del modelo.

Una alternativa es emplear otro tipos de modelos como puede ser el de vectores autorregresivos aumentado por factores (FAVAR) introducido por Bernanke (2005) que constituye un caso particular de los modelos dinámicos de factores (MDF). En este marco, la premisa de los MDF consiste en que la dinámica común de un número grande de variables de series temporales se desprende de un conjunto relativamente menor de factores inobservables o latentes, que a su vez evolucionan en el tiempo. Este tipo de modelos son un caso particular de los modelos de espacio de estado o modelos de markov ocultos, para los cuales variables observables se expresan en términos de variables latentes inobservables que evolucionan de acuerdo a una dinámica retardada con dependencia finita.

Stock y Watson (2016) establecen que los MDF se destacan en las aplicaciones macroeconómicas debido a que los complejos movimientos de un número potencialmente grande de series observables se pueden resumir en un número menor de factores que conducen las fluctuaciones comunes de todas las series.

Los MDF se pueden expresar tanto en términos dinámicos como estáticos, lo que conlleva a dos métodos de estimación distintos. En ambos casos, se asume que tanto las variables observables como latentes son estacionarias de segundo orden e integradas de orden cero. Sin embargo, para el caso dinámico se va a representar la dependencia de las series temporales X_t en los rezagos de los factores de manera explícita, para el caso estático, se va a representar tal dinámica de manera implícita.

Los MDF en su forma dinámica, se definen como un vector X_t de series temporales observables de tamaño $N \times 1$ que dependen de un número menor de factores latentes e inobservables f_t de tamaño q y de un componente idiosincrático e_t con media cero. En general, los factores latentes y el término idiosincrático van a estar serialmente correlacionados.

El MDF se representa como

$$X_t = \lambda(L)f_t + e_t, \quad (1)$$

$$f_t = \Psi(L)f_{t-1} + \eta_t \quad (2)$$

donde la matrices de rezagos polinomiales $\lambda(L)$ y $\Psi(L)$ son de tamaño $N \times q$ y $q \times q$, respectivamente, mientras que η_t es el vector de innovaciones, de tamaño $q \times 1$, con media cero, serialmente incorrelacionadas a los factores.

Se asume que las perturbaciones idiosincráticas están incorrelacionadas con el factor de innovaciones en todos sus rezagos, esto es, $Ee_t\eta_{t-k} = 0$ para todo k . En general, e_t puede estar serialmente correlacionado.

Cuando las perturbaciones idiosincrásicas e_t en (1) estén serialmente correlacionadas, los modelos (1) y (2) van a estar completamente especificados. Suponiendo que la i -ésima perturbación idiosincrática sigue una autorregresión univariada, se tiene que

$$e_{it} = \delta_i(L)e_{it-1} + \nu_{it}, \quad (3)$$

donde ν_{it} está serialmente incorrelacionado.

En el MDF estático o apilado la forma dinámica (1) y (2) depende de $r \geq q$ factores estáticos F_t en vez de q factores dinámicos f_t con r . En este esquema, la estimación de los factores se puede realizar mediante componentes principales y se puede reescribir al modelo de la siguiente forma.

Sea p los grados de la matriz de rezagos polinomial $\lambda(L)$ y sea $F_t = (f'_t, f'_{t-1}, \dots, f'_{t-p})'$ el vector de de factores estáticos de tamaño $r \times 1$. Sea, también, $\Lambda = (\lambda_0, \lambda_1, \dots, \lambda_p)$, donde λ_h es la matriz de coeficientes, de tamaño $N \times q$, sobre el h -ésimo rezago de $\lambda(L)$. De forma similar, sea $\Phi(L)$ una matriz de unos, ceros y los elementos de $\Psi(L)$ tales que el vector de autorregresión en (2) se reescriba en términos de F_t . Con esta notación el MDF (1) y (2) se puede reescribir como,

$$X_t = \Lambda F_t + e_t, \quad (4)$$

$$F_t = \Phi F_{t-1} + G\eta_t \quad (5)$$

donde $G = [I_q \quad 0_{q \times (r-q)}]'$.

Cuando, además, se asume que las innovaciones son una combinación lineal de unas perturbaciones estructurales inobservables tales que

$$\eta_t = H\varepsilon_t, \quad (6)$$

Y las perturbaciones estructurales estén incorrelacionadas

$$E\varepsilon_t\varepsilon'_t = \Sigma_\varepsilon = \begin{bmatrix} \sigma_{\varepsilon_1}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{\varepsilon_n}^2 \end{bmatrix}$$

se está en el marco de un modelo dinámico de factores estructural (MDFS), que se expresa como

$$X_t^{n \times 1} = \Lambda^{n \times r} F_t^{r \times 1} + e_t^{n \times 1}, \quad (7)$$

$$\Phi(L)^{r \times r} F_t^{r \times 1} = G^{r \times q} \eta_t^{q \times 1} \quad \text{donde } \Phi(L) = I - \Phi_1 L - \dots - \Phi_p L^p \quad (8)$$

$$\eta_t^{q \times 1} = H^{q \times q} \varepsilon_t^{q \times 1} \quad (9)$$

En este sistema, las q perturbaciones estructurales ε_t impactan a los factores comunes pero no así al término idiosincrático. Bajo este esquema, todos los factores son inobservables, cuando se consideran algunos observables el MFDS se transforma en un modelo FAVAR. En este sentido, mientras que algunos factores van a ser observados otros se mantienen inobservados. Los FAVAR imponen restricciones en los MDF dado que uno o más factores van a estar medidos sin error por una o más variables observables. Una representación típica de éstos modelos es la presentada por Bernanke (2005), donde Y_t es el vector de tamaño $M \times 1$ variables observables que conducen la dinámica de la economía y los factores inobservados F_t son de tamaño $K \times 1$ donde K es pequeño. De esta manera, la dinámica conjunta de (F'_t, Y'_t) se puede representar por la *ecuación de transición*.

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + v_t, \quad (10)$$

donde $\Phi(L)$ es el polinomio de rezagos conformable finito de orden d . El término de error v_t tiene media cero con matriz de covarianzas Q . La ecuación (10) es un VAR en (F'_t, Y'_t) . Cuando los términos de $\Phi(L)$ sean iguales a cero se obtiene un VAR de Y_t en su forma estandar. La ecuación (10) no puede ser estimada directamente porque los factores F_t son inobservables.

Se supone que se cuenta con un conjunto N de series "informacionales" X_t , siendo N grande, en particular $N > T$ con T representando la cantidad de períodos y mucho mayor a la cantidad de factores y variables observables en el sistema FAVAR ($K + M \ll N$).

Las series de tiempo informacionales X_t se relacionan a los factores inobservables F_t y a las variables observables Y_t mediante una *ecuación de observación* de la forma

$$X_t = \Lambda^f F_t + \Lambda^y Y_t + e_t \quad (11)$$

donde Λ^f es la matriz de pesos de los factores de tamaño $K \times N$, Λ^y es de tamaño $N \times M$ y el vector de errores e_t es de tamaño $N \times 1$ con media cero. La ecuación (11) captura la idea de que tanto Y_t como F_t , que en general pueden estar correlacionados, representan a los factores comunes que conducen la dinámica de X_t . Cuando en esta ecuación no se tengan factores observables el modelo es un MDF.

El autor presenta dos formas para estimar a (10)-(11), la primera consiste en un enfoque semiparamétrico de componentes principales en dos pasos y la segunda, un enfoque completamente paramétrico de verosimilitud bayesiano de un sólo paso.

Para que se pueda estimar el modelo, se tienen que imponer restricciones sobre (10)-(11), es decir, hacer la identificación del sistema. Como se deja a la dinámica del VAR en (10) irrestricta, es necesario imponerlas sobre los factores y los coeficientes de la ecuación (11). Si se asume que $\hat{\Lambda}^f$ y \hat{F}_t son una solución al problema de estimación, se podría definir $\tilde{\Lambda}^f = \tilde{\Lambda}^f H$ y $\tilde{F}_t = H^{-1} \hat{F}_t$ donde H es una matriz no singular de tamaño $K \times K$ que también satisfaga ala ecuación (11). Además, se tiene que imponer una normalización. Para el caso de la estimación por componentes principales, en el primer paso, se puede emplear la normalización implícita en la misma. Por otro lado, identificar a las perturbaciones estructurales en la ecuación de transición requiere imponer más restricciones.

Dos objetivos en este tipo de análisis consisten en estimar la función impulso-respuesta (SIRF⁵) y descomposición de la varianza del error predicción estructural (FEVD⁶). Ambas representaciones se obtienen a partir de la representación MA del VAR del sistema y a partir de la varianza de los shocks estructurales. En este sentido, considerando al VAR estructural de Y_t como $A(L)Y_t = H\varepsilon_t$ y a su respectivo MA estructural como $Y_t = D(L)\varepsilon_t$ donde $D(L) = C(L)H$ y H^{-1} es invertible, se define a la función impulso respuesta como

$$SIRF_{ij} = \{D_{h,ij}\}, \quad h = 0, 1, \dots, D_h = C_h H \quad (12)$$

donde D_h es el h -ésimo rezago de la matriz de coeficientes en $D(L)$. Entonces $D_{h,ij}$ es el efecto causal de la i -ésima variable del aumento unitario en el j -ésimo shock luego de h períodos.

Por su parte la $FEVD_{h,ij}$ mide la importancia de la j -ésima perturbación para explicar la variación en Y_{it} calculando la contribución relativa del shock en la varianza frente a cambios inesperados en Y_{it} en h períodos, esto es, la varianza de la predicción del error h pasos hacia adelante,

⁵structural impulse response functions

⁶structural forecast error variance decomposition

$$FEVD_{h,ij} = \frac{\sum_{k=0}^h D_{k,ij}^2 \sigma_{\varepsilon_j}^2}{\text{var}(Y_{it+h}|Y_t, Y_{t-1}, \dots)} = \frac{\sum_{k=0}^h D_{k,ij}^2 \sigma_{\varepsilon_j}^2}{\sum_{j=1}^n \sum_{k=0}^h D_{k,ij}^2 \sigma_{\varepsilon_j}^2} \quad (12)$$

con $D(L) = A(L)^{-1}H$

Éstos modelos, desde el punto de vista del aprendizaje profundo, presentan a grandes rasgos dos debilidades. Primero, asumen una estructura lineal y segundo, podrían presentar limitaciones en cuanto a escalabilidad en términos computacionales cuando se tiene una gran cantidad de variables. Por esto, Los modelos de factores en un marco de aprendizaje profundo permitirían sobrepasar estas limitaciones, manteniendo el mismo grado de flexibilidad y de interpretabilidad que los DFM. En este sentido, en una segunda etapa de este estudio, se va a emplear un modelo de factores dinámicos profundo ("Deep Dynamic Factor Model", de ahora en más D^2FM).

El D^2FM adopta una estructura asimétrica de Autoencoders (AE). La formulación dinámica de los autoencoders puede ser vista como una generalización no lineal de los modelos dinámicos de factores.

Los AE son redes neuronales entrenadas para mapear un conjunto de variables dentro de sí mismas, primero codificando a las entradas en una representación de menor dimensión (o incompleta) y luego decodificándola de vuelta hacia sí misma.

La forma general de los modelos D^2FM se puede expresar como

$$\mathbf{f}_t = G(\mathbf{y}_t), \quad (13)$$

$$\mathbf{y}_t = F(\mathbf{f}_t) + \varepsilon_t, \quad (14)$$

$$\mathbf{f}_t = B_1 \mathbf{f}_{t-1} + \dots + B_p \mathbf{f}_{t-p} + u_t, \quad u_t \stackrel{iid}{\sim} \mathcal{N}(0, U) \quad (15)$$

$$\varepsilon_t = \Phi_1 \varepsilon_{t-1} + \dots + \Phi_d \varepsilon_{t-d} + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, Q), \quad (16)$$

se mantienen los supuestos de linealidad en las ecuaciones dinámicas mientras que se permiten no linealidades entre las variables y los factores.

El modelo D^2FM se puede implementar usando una estructura de AE simétricos con un perceptrón multicapas (MLP⁷) para la función de codificación (13) y otro MLP para la decodificación (14). Por su parte, el supuesto de innovaciones gaussianas e *iid* permite interpretar a la red estimada como un MAP de la verosimilitud del modelo.

La esencia del modelo es el AE con un codificador no lineal multicapas y una estructura del decodificador que puede ser tanto asimétrica como simétrica con una única capa lineal. Se adoptan ecuaciones lineales estocásticas autorregresivas para modelar la dinámica de los factores y los componentes idiosincráticos.

En la red de decodificación, la cantidad de capas ocultas y la cantidad de neuronas no precisan estar predefinidas y se pueden seleccionar con validación cruzada.

Para estimar D^2FM se emplea un procedimiento en dos pasos. Primero, se estiman todos los parámetros del modelo. Segundo, se une a la parte de decodificación al marco de espacios de estados para permitir actualizaciones en línea de los estados latentes dados los observables.

Como en las series temporales no puede aplicarse el método de validación cruzada $K - fold$ porque hay correlación serial en los datos. Por lo tanto se emplea un enfoque de *fuera de la muestra* (*out-of-sample*) que consiste en dividir al conjunto de observaciones hasta un cierto momento T entre un conjunto de entrenamiento $[0, T - k * h - 1]$ y un conjunto de validación $[T - k * h, T - (h - 1) * h]$ donde h determina el largo del conjunto mientras que $k = K, \dots, 1$ con $K \ll \frac{T-1}{h}$. Al hacer un promedio sobre las pérdidas

⁷Multilayer perceptron

calculadas en los K conjuntos de validación, se llega a obtener una "pérdida de validación". Esto implica que se va a estimar un modelo dado con hiperparámetros fijos K veces, y esto va a ser así para cada combinación posible de hiperparámetros.

Las ventajas de modelos se pueden resumir en que mantienen el mismo nivel de interpretabilidad que un MDF clásico, por lo que es fácil compararlos. Segundo, la estructura de autoencoder permite introducir técnicas de aprendizaje profundo para poner a prueba su potencial. Tercero, la red de decodificador lineal y su marco de espacios de estado permiten emplear el filtro de kalman estandar para actualizar los estados inobservables en tiempo real. Por último, la adopción de técnicas de filtrado lineal permiten una fácil interpretación de las previsiones del modelo.

Referencias

- [1] *Adaptive Filter Theory*, Haykin, S., Prentice Hall, 3rd Edition, 1995
- [2] *Statistical Digital Signal Processing and Modeling*, Hayes M.H., Wiley, 1st Edition, 1996.