



UNIVERSIDAD
DE LA REPUBLICA
URUGUAY

Comparación de Modelos Dinámicos de Factores

Un análisis del Impacto de Política Monetaria

Laura Marta Montaldo Iglesias

Programa de Posgrado en Ingeniería Matemática
Facultad de Ingeniería
Universidad de la República

Montevideo – Uruguay
Junio de 2024



UNIVERSIDAD
DE LA REPUBLICA
URUGUAY

Comparación de Modelos Dinámicos de Factores

Un análisis del Impacto de Política Monetaria

Laura Marta Montaldo Iglesias

Tesis de Maestría presentada al Programa de Posgrado en Ingeniería Matemática, Facultad de Ingeniería de la Universidad de la República, como parte de los requisitos necesarios para la obtención del título de Magíster en Ingeniería Matemática.

Director:

Dr. Marco Scavino

Director académico:

Dra. Paola Bermolen

Montevideo – Uruguay

Junio de 2024

Montaldo Iglesias, Laura Marta

Comparación de Modelos Dinámicos de Factores / Laura Marta Montaldo Iglesias. - Montevideo: Universidad de la República, Facultad de Ingeniería, 2024.

XII, 125 p. 29, 7cm.

Director:

Marco Scavino

Director académico:

Paola Bermolen

Tesis de Maestría – Universidad de la República, Programa en Ingeniería Matemática, 2024.

Referencias bibliográficas: p. 78 – 83.

1. Modelos dinámicos de factores, 2. Modelo de vectores autorregresivos aumentados por factores, 3. Modelo dinámico de factores profundo, 4. FAVAR profundo, 5. Política monetaria, 6. Funciones de respuesta al impulso, 7. Descomposición de la varianza del error de predicción. I. Scavino, Marco, . II. Universidad de la República, Programa de Posgrado en Ingeniería Matemática. III. Título.

INTEGRANTES DEL TRIBUNAL DE DEFENSA DE TESIS

Dra. Elizabeth Bucacos

Dr. Gonzalo Perera

Dr. Juan Kalemkerian

Montevideo – Uruguay
Junio de 2024

A mi padre ♡

Agradecimientos

Quisiera agradecer, en un primer momento, a Elizabeth Bucacos que me introdujo a los modelos FAVAR y me dió la dirección para la creación de esta tesis, a Marco Scavino por su paciencia infinita, por presentarme el mundo de los modelos dinámicos de factores y por ser mi tutor. También a Paola Bermolen por ayudarme en todo este proceso y a la Facultad de Ingeniería por darme la oportunidad de cursar la Maestría en Ingeniería Matemática.

*Una teoría económica es una
abstracción del mundo real.
Entre otras razones, la inmensa
complejidad de la realidad
económica hace imposible
comprender todas las
interrelaciones a un tiempo;
tampoco tienen todas ellas igual
importancia en lo que atañe a
los fenómenos económicos que
estudiamos. Por eso, lo más
sensato es elegir los que, a
nuestro juicio, constituyen los
factores y relaciones
primordiales que interesan a
nuestro problema, y concentrar
la atención exclusivamente en
ellos. Un esquema analítico de
esta índole, deliberadamente
simplificado, se llama modelo
económico, pues representa la
realidad económica de una
manera esquemática y
aproximada.*

Alpha Chiang

RESUMEN

Esta tesis presenta una exploración del impacto de una política monetaria contractiva a través del prisma de modelos de factores dinámicos. Profundizando en el análisis, se parte con un modelo Vectores autorregresivos aumentados por factores (FAVAR) que sirve de punto de referencia, se lo compara predictivamente con un modelo dinámico de factores profundo, basado en Autoencoders y un modelo FAVAR profundo. El núcleo innovador de este trabajo reside en la comparación de un modelo FAVAR lineal tradicional con una propuesta vanguardista, el modelo Deep Dynamic Factor Model, un paradigma emergente en el aprendizaje profundo, y en el desarrollo de un modelo FAVAR profundo. Comparando los resultados para distintas cantidades de factores, esta investigación no solo arroja luz sobre el rendimiento predictivo de estos modelos, sino que también ofrece una perspectiva única sobre las consecuencias de las políticas monetarias en un conjunto adecuado de variables económicas. Además, este estudio abre el debate sobre si incorporar modelos de aprendizaje profundo en los estudios de política monetaria, dada la discusión relacionada a emplear modelos de cajas negras y sus posibles consecuencias negativas en temas sensibles como es la política económica. Por tanto, esta investigación contribuye a la literatura existente con un análisis comparativo robusto e introduce una herramienta analítico-computacional sofisticada para la interpretación de políticas monetarias en un entorno económico complejo.

Palabras claves:

Modelos dinámicos de factores, Modelo de vectores autorregresivos aumentados por factores, Modelo dinámico de factores profundo, FAVAR profundo, Política monetaria, Funciones de respuesta al impulso, Descomposición de la varianza del error de predicción.

Lista de siglas

- ADF** Prueba de Dickey-Fuller aumentada 38, 39
- AE** Autoencoders 4, 27, 28, 29, 32
- AI** Inteligencia artificial 23, 25
- Adam** Estimación adaptativa de momentos 53, 121, 122
- CSIRF** Función de respuesta al impulso acumulada 13
- DDFM** Modelo dinámico de factores profundo 4, 6, 7, 22, 23, 26, 27, 28, 32, 33, 34, 41, 48, 49, 53, 54, 55, 76
- DFM** Modelos dinámicos de factores 3, 15, 16, 21, 22, 27, 32, 45
- DL** Aprendizaje profundo 4, 22, 23, 25, 26, 27, 59, 76, 119
- EM** Esperanza-Maximización 77
- FAVAR** Vectores autorregresivos aumentados por factores VIII, 2, 3, 4, 6, 9, 15, 16, 19, 20, 22, 41, 45, 59, 60, 73
- FED** Reserva Federal de Estados Unidos 67, 125
- FEVD** Descomposición de la varianza del error de predicción 5, 6, 9, 12, 68, 72
- FFR** Tasa de Fondos Federales 5, 13, 46, 67, 68, 72, 125
- FIR** Funciones de respuesta al impulso 5, 6, 9, 68
- FOMC** Comité Federal de Mercado Abierto 67
- FRED** Datos Económicos de la Reserva Federal 36

ICS Desplazamiento de covariación interna 53

ICr Criterios de información 44

KPSS Prueba de Kwiatkowski, Phillips, Schmidt y Shin 38, 39

MA Medias móviles 72

MAE Error absoluto medio 57

MCMC Cadena de Markov Monte Carlo 49

MCO Mínimos cuadrados ordinarios 43, 44, 55

ML Aprendizaje automático 23, 24, 25, 51, 59, 118

MLP Perceptrón multicapa 29, 30, 31, 32

MSE Error cuadrático medio 28, 50, 57, 58

PC Componentes principales 6, 7, 18, 21, 28, 29, 43, 45, 46, 47, 73

PM Política monetaria 2, 3, 4, 5, 6, 9, 46, 67, 68, 72, 73, 74, 76, 125

RMSFE Error de predicción cuadrático medio 57

RU Raíz unitaria 7, 38, 39, 40

SDFM Modelos dinámicos de factores estructurales 9, 16, 19

SFIR Funciones de respuesta al impulso estructurales 12, 19

SGD Descenso por gradiente estocástico 118, 121

SLP Perceptrón de una capa 32

SMA Medias móviles estructurales 11, 12

SVAR Vectores autorregresivos estructurales 9, 10, 11, 12, 16, 19

VAR Vectores autorregresivos 2, 3, 6, 9, 10, 13, 14, 15, 20, 22, 34, 46, 55, 72

VC Validación cruzada 52

VMA Vectores de medias móviles 11

Tabla de contenidos

Lista de siglas	x
1 Introducción	1
1.1 Motivación y objetivos	2
1.2 Organización del trabajo	6
2 Fundamentos teóricos	8
2.1 Modelos de vectores autorregresivos	9
2.2 Modelos dinámicos de factores lineales	16
2.3 Aprendizaje Automático, Profundo e Inteligencia artificial . . .	23
2.4 Modelos dinámicos de factores profundos	27
3 Estrategia empírica	35
3.1 Datos del estudio	36
3.2 Inferencia	42
4 Resultados empíricos	56
4.1 Desempeño de los modelos estimados	57
4.2 Importancia de las variables y correlación entre factores	59
4.3 Análisis de impacto de política monetaria	67
5 Consideraciones finales	75
5.1 Conclusiones y trabajos futuros	76
Referencias bibliográficas	78
Apéndices	84
Apéndice 1 Conjunto de datos	85
Apéndice 2 Resultados de las pruebas ADF y KPSS	90

Apéndice 3	Importancia de las variables sobre los factores	96
Anexos		108
Anexo 1	Estacionariedad de las series temporales	109
Anexo 2	Cantidad óptima de factores	114
Anexo 3	Modelo dinámico de factores profundo	116
Anexo 4	Cometidos de la política monetaria	124

Capítulo 1

Introducción

1.1. Motivación y objetivos

El presente estudio se ubica en lo que Bernanke y Boivin (2003) denominaron *entorno rico en datos*. El mismo se caracteriza por la disponibilidad de un amplio conjunto de variables y observaciones temporales. En este contexto, considerando un marco de Política monetaria (PM), los bancos centrales suelen seguir y monitorear una gran cantidad de datos para tomar conclusiones sobre las dinámicas del sistema económico (Stock y Watson (2016, cap.1)). De esta manera, se define un conjunto de series que representan el sistema de variables para una determinada realidad económica¹.

El conjunto de series informacionales de esta investigación surge de la base de datos de McCracken y Ng (2016). Es un conjunto de serie temporales que por su volumen y temporalidad se considera *big data* y permite brindar a la comunidad científica una base de datos común para la experimentación empírica, como son los estudios de Andreini et al. (2023a) y de Klieber (2024) que sirven de antecedentes para el presente trabajo.

Para caracterizar los determinantes y efectos de la PM sobre lo que se puede conceptualizar como la economía, la literatura tradicional se destaca por el uso de los modelos de Vectores autorregresivos (VAR)². Sin embargo, estos modelos padecen lo que se conoce en la literatura como la maldición de la dimensionalidad (Bellman (1957)), lo que limita su eficacia para manejar grandes volúmenes de datos y reflejar adecuadamente la complejidad de la economía moderna. Lo anterior implica que para representar la realidad, los VAR son capaces de incluir sólo un número limitado de variables en el sistema. Aunque los modelos VAR son apreciados por su simplicidad, surge una discrepancia entre la abundancia de series temporales que los bancos centrales supervisan y la capacidad limitada de estos modelos para capturar y medir los efectos de una perturbación en una variable considerada exógena sobre otras variables macroeconómicas (Bernanke et al. (2005)).

En este contexto, han surgido otros enfoques que permiten incorporar información adicional y uno de los más destacados es el modelo FAVAR³. El mismo fue propuesto inicialmente por Bernanke et al. (2005) como una herramienta para abordar el problema de la gran dimensionalidad de los datos económi-

¹A estas series temporales se las puede llamar series informacionales (como en Bernanke et al. (2005)).

²Por sus siglas en inglés relativas a Vector Autoregressive Models.

³Por sus siglas en inglés relativas a Factor-Augmented Vector Autoregression.

cos. Desde su introducción, el mencionado modelo se ha establecido como un enfoque esencial en la modelización macroeconómica. Ha encontrado aplicación en una variedad de estudios, abarcando desde la evaluación de políticas económicas hasta el análisis de los impactos de perturbaciones financieras en la economía. Como referencia nacional, Bucacos (2015) realiza un análisis novedoso del impacto de la PM internacional sobre variables relevantes para la economía uruguaya empleando el FAVAR. Se busca estudiar la vulnerabilidad de la economía uruguaya frente a este tipo de choque externo considerando un espacio temporal de 20 años. Este trabajo de investigación tiene como objetivo entender los canales de transmisión de perturbaciones externas sobre una economía pequeña, abierta y dolarizada.

El modelo FAVAR es una extensión del VAR y se basa en la idea de que múltiples series temporales macroeconómicas pueden ser influenciadas por factores comunes inobservables, también conocidos como factores latentes. Siguiendo a Bernanke et al. (2005), estos factores son constructos no directamente medibles como puede ser la actividad económica, condiciones crediticias, tendencias generales de la economía y no sería posible representarlos por una o pocas series temporales. Como consecuencia, la inclusión de estos factores en el análisis mejoraría significativamente la capacidad de pronóstico y comprensión de la economía, al capturar la riqueza de datos de numerosas series temporales, como señalan investigaciones de Stock y Watson (Stock y Watson (2005, 2015, 2016)).

En el presente trabajo se sigue la premisa de los Modelos dinámicos de factores (DFM). Estos plantean que un número reducido de factores estimados resume un amplio conjunto de información económica (Stock y Watson (2016)). Por consiguiente, se adopta la metodología propuesta por Stock y Watson (2016) relativa a esquemas de reducción de la dimensionalidad. La misma consiste en estimar unos factores inobservables a incorporar en modelos estructurales para evaluar cómo los shocks de PM¹ afectan a un grupo de variables macroeconómicas. De esta manera, se tiene como objetivo analizar y cuantificar el impacto de las variaciones en la tasa de interés de referencia de la Reserva Federal de Estados Unidos² sobre un conjunto particular de series temporales similar al empleado en Bernanke et al. (2005). En este sentido, las

¹En español se llaman choques de PM (Bucacos (2015)) por lo que a lo largo de la presente tesis se emplearán indistintamente los términos choque y shock de PM.

²Tasa de Fondos Federales.

investigaciones de Christiano et al. (1997, 1999) señalan que los shocks de PM son buenos candidatos para aplicar lo que se conoce en la literatura como el programa de Lucas. Esto se debe a que diversos modelos muestran respuestas variadas ante este tipo de shock. Dicho programa se compone de tres pasos:

1. Se aísla al choque de PM de la economía: se establecen los supuestos de identificación.
2. Se caracterizan las respuestas esperadas de las variables de la economía.
3. Se comparan los resultados de los modelos con la realidad.

Con lo anterior, se selecciona aquel modelo que mejor responda a la pregunta de cómo reacciona la economía frente a un shock exógeno de PM. Este programa sirve de guía para comparar resultados empíricos con lo que establece la teoría y el conocimiento. Como en Christiano et al. (1999), se adopta la sugerencia de Lucas (1980) que indica que para experimentar en política económica es necesario poner a prueba modelos que sean imitaciones útiles de la realidad, sujetándolos a choques para los cuales se tenga una idea a priori de cómo va a reaccionar la economía. En la medida que el modelo sea capaz de reponder a preguntas simples, será posible avanzar con preguntas más complejas. En este marco, se toma el modelo FAVAR como referencia y se busca comparar sus resultados con otro modelo.

En línea con el concepto anterior, desde la perspectiva del Aprendizaje profundo (DL)¹, el Teorema de No Free Lunch establece que el rendimiento medio de cualquier par de algoritmos para todos los problemas posibles es idéntico (Goodfellow et al. (2016, p. 114-116), Wolpert y Macready (1997)). Esto implica que en promedio, considerando todas las distribuciones de datos posibles, ningún algoritmo es mejor que otro para resolver un problema.

Siguiendo estas ideas, se estiman tres modelos: primero, un modelo FAVAR lineal como en Bernanke et al. (2005); luego, uno novedoso de DL, basado en Autoencoders (AE)² propuesto por Andreini et al. (2023a), llamado Modelo dinámico de factores profundo (DDFM); por último, una vez estimados los factores por este último modelo, se infiere lo que se denomina como modelo FAVAR profundo³. Posteriormente, se analiza el desempeño predictivo de cada modelo y su capacidad de generalización para nuevas entradas de datos. Se

¹Por sus siglas en inglés relativas a Deep Learning.

²En español se denomina Codificadores Automáticos.

³Como el propuesto en Klieber (2024), llamado Deep Dynamic FAVAR.

busca evaluar si, a pesar de su complejidad, los resultados obtenidos con los modelos de aprendizaje profundo son comparables a los de un enfoque lineal manteniendo, a la vez, niveles similares de flexibilidad e interpretabilidad. Una vez realizado lo anterior, se analiza el impacto de un shock de PM sobre un conjunto de indicadores macroeconómicos empleando los modelos FAVAR lineal y profundo. Para este propósito, se calculan las Funciones de respuesta al impulso (FIR) y la Descomposición de la varianza del error de predicción (FEVD) ante un aumento de la Tasa de Fondos Federales (FFR) para un conjunto de series de tiempo.

En cuanto a los hallazgos del estudio, los tres modelos presentan un desempeño predictivo similar. El análisis de las funciones de respuesta al impulso y la descomposición de la varianza del error de predicción arrojan resultados comparables para los tres modelos y en relación con la literatura para varias de las series macroeconómicas consideradas. Sin embargo, respecto a la importancia de las variables informacionales en relación con los factores estimados por los modelos FAVAR lineal y profundo, se encontraron diferencias significativas en los resultados. Mientras que en el caso lineal los factores están vinculados a grupos más homogéneos de variables, en el caso profundo los resultados son más heterogéneos.

1.2. Organización del trabajo

En el capítulo 2 se describen los modelos y conceptos teóricos que fundamentan la estrategia empírica del capítulo 3 y el análisis de los resultados del capítulo 4. Para esto se propone un esquema de 3 etapas como se explica a continuación:

1. **Preprocesamiento de las series temporales:** Esta etapa consiste en una serie de pasos con el fin de obtener un panel de datos de series temporales estacionarias, sin datos faltantes ni datos atípicos. Los detalles relativos a la implementación del preprocesamiento de las series univariadas y construcción del conjunto de datos que sirve de entrada para la modelización se encuentran en [este enlace](#) (archivo README.md).
2. **Estimación de los modelos FAVAR lineal, DDFM y FAVAR profundo:** Primero, se estima el modelo FAVAR lineal adoptando la metodología propuesta por Bernanke et al. (2005), relativa a un análisis por Componentes principales (PC). Segundo, se estima el modelo DDFM siguiendo a Andreini et al. (2023a). Por último, como en Klieber (2024), se incluyen los factores no lineales estimados en un esquema VAR para inferir un modelo FAVAR profundo.
3. **Análisis de los resultados de las estimaciones:**
 - a) **Error de generalización:** Tras la estimación de los modelos se comparan sus desempeños predictivos con distintas medidas de error de generalización.
 - b) **Importancia de las variables:** Se obtiene la correlación entre factores y se analiza la importancia de las variables del conjunto de datos sobre los factores estimados siguiendo los criterios empleados en Klieber (2024).
 - c) **Impacto de perturbaciones de PM:** Como en Bernanke et al. (2005) y en Klieber (2024), se calculan las FIR para un conjunto de variables consideradas de interés con el objetivo de evaluar si se comportan de acuerdo a lo que indica la literatura (Christiano et al. (1999), Bernanke et al. (2005)). Por último, se computan las FEVD para cada modelo FAVAR como en Bernanke et al. (2005).

En el capítulo 5 se concluye el estudio y se proponen posibles trabajos futuros.

Los cálculos computacionales fueron desarrollados tanto en lenguaje de programación R como en Python (R Core Team (2023), Python Software Foundation (2023)). De manera general, los datos y el código del estudio están disponibles en el siguiente [repositorio público](#). En particular, se contribuye en la generación de dos clases en Python. Las mismas sirven para el análisis de pruebas de Raíz unitaria (RU) y de estacionariedad del conjunto de datos, como se describe en el capítulo 3, sección 3.1. Para el caso del modelo FAVAR lineal, se extiende el código de Duarte (2020) creando una función para el análisis por PC de acuerdo a la metodología que se explica en el capítulo 3, sección 3.2. Para el caso del modelo DDFM, se amplía el código desarrollado por Andreini et al. (2023b); se incluye una función para estimar las predicciones que se comparan con el conjunto de prueba, como se detalla en el capítulo 4, sección 4.1.

Capítulo 2

Fundamentos teóricos

2.1. Modelos de vectores autorregresivos

El análisis comienza explicando los conceptos de perturbaciones estructurales, modelos de Vectores autorregresivos estructurales (SVAR), de Vectores autorregresivos en su forma reducida (VAR) y abordando los problemas de identificación e invertibilidad que surgen en este contexto. Esto se debe a que, en primer lugar, según Stock y Watson (2016, p.443), los esquemas de identificación utilizados en los SVAR pueden adaptarse directamente a los Modelos dinámicos de factores estructurales (SDFM)¹. Estos, al imponer ciertas restricciones, permiten la construcción de Vectores autorregresivos aumentados por factores (FAVAR), siendo el modelo de referencia en el presente trabajo de investigación. En segundo lugar, el FAVAR es en sí mismo un VAR en sus argumentos. Por último, de la representación VAR en medias móviles, se desprenden las Funciones de respuesta al impulso (FIR) y la Descomposición de la varianza del error de predicción (FEVD) que sirven para examinar el impacto de un shock de PM sobre la economía. Complementando lo anterior, en Christiano et al. (1999) se sugiere utilizar modelos estructurales para llevar a cabo experimentos destinados a responder preguntas relacionadas con los efectos de una determinada política monetaria (PM). Esta recomendación surge en respuesta a las limitaciones que podrían presentar las alternativas de emplear técnicas puramente estadísticas² o de realizar pruebas directamente sobre la realidad.

En este escenario, uno de los objetivos de la macroeconomía empírica consiste en estimar el efecto que tienen las perturbaciones estructurales no anticipadas (shocks)³ sobre la economía, representada por un sistema de ecuaciones simultáneas⁴. Según Bernanke (1986), los choques son fuerzas exógenas primitivas que no están correlacionadas entre sí y tienen un significado económico. Ejemplos de shocks, por nombrar algunos, incluyen un aumento inesperado de la tasa de interés de referencia del Banco Central (shock de PM), un incremento sorpresivo en los precios del petróleo (shock de oferta de petróleo), una mejora imprevista en la productividad (shock de productividad), entre otros. Como consecuencia de un shock inesperado, se pueden producir marcadas variaciones

¹Por sus siglas en inglés relativas a Structural dynamic factor models.

²Porque cada realidad económica constituye un caso particular.

³Dado que en español se traduce como choque, a lo largo del presente estudio se emplean los conceptos de perturbaciones estructurales, shocks y choques de manera indistinta.

⁴Como en Ramey (2016), en este estudio adopto el concepto de shock de Blanchard y Watson (1984), Bernanke (1986) y Stock y Watson (2016).

en otras variables del sistema económico (Stock y Watson (2016)).

Según Ramey (2016), los shocks presentan las siguientes características: primero, son exógenos respecto a las otras variables endógenas del modelo; segundo, no están correlacionados con otros choques exógenos; tercero, son representaciones de movimientos no anticipados en variables exógenas o noticias sobre movimientos futuros en variables exógenas. En este marco, se asume que las perturbaciones estructurales van a ser tanto autónomas como no anticipadas, lo que implica que, por un lado, están incorrelacionadas con otros shocks y, por el otro, están serialmente incorrelacionadas. Esto significa que no es posible observar a las perturbaciones sin errores de medida, por lo que en la literatura han surgido enfoques para poder identificarlas.

Siguiendo a Stock y Watson (2016), el esquema de identificación dominante es el de Vectores autorregresivos propuesto por Sims (1980), el cual se fundamenta en la Descomposición de Cholesky¹. Estos modelos permitieron establecer un vínculo entre las innovaciones en un sistema lineal y los choques estructurales macroeconómicos. La premisa fundamental de los modelos SVAR consiste en suponer que el espacio de innovaciones relativo a una serie temporal Z_t está generado por el espacio de perturbaciones estructurales. Entonces, para identificar a los shocks de interés alcanza con encontrar una combinación lineal de innovaciones relativas a esas perturbaciones estructurales². Originalmente, para realizar la identificación del sistema, se tomaron restricciones de corto plazo, pero luego fueron emergiendo otros enfoques para identificar shocks con restricciones de más largo plazo (Stock y Watson (2016), Ramey (2016)).

Para llegar a la expresión del SVAR, en Stock y Watson (2016) se parte de la representación de un modelo VAR de orden p , en su forma reducida. En esta dirección, para un vector de series temporales estacionarias Z_t de tamaño $n \times 1$ con media cero, se tiene que

$$VAR(p) : \quad Z_t = A_1 Z_{t-1} + \dots + A_p Z_{t-p} + \eta_t \quad \text{ó} \quad A(L)Z_t = \eta_t \quad (2.1)$$

donde $A(L) = I - A_1 L - A_2 L^2 - \dots - A_p L^p$ y L es el operador de rezagos $L^p Z_t = Z_{t-p}$. El vector Z_t es una función lineal de sus p valores rezagados a lo que se le suma una innovación serialmente incorrelacionada η_t , siendo

¹También conocida como triangulación.

²En Ramey (2016) se indica que los primeros en identificar los shocks de la tasa de fondos federales como choques de política monetaria fueron Bernanke y Blinder (1992).

una secuencia martingala en diferencias con matriz de covarianzas Σ_η . Como (2.1) es la regresión poblacional de Z_t en sus rezagos entonces $A(L)$ y Σ_η están identificados. También se puede describir a Z_t en términos de los valores pasados y presentes de sus innovaciones a través de una representación Vectores de medias móviles (VMA)³ como

$$Z_t = C(L)\eta_t \quad \text{donde} \quad C(L) = I + C_1L + C_2L^2 + \dots = A(L)^{-1}. \quad (2.2)$$

Por su parte, el SVAR representa a Z_t en términos de sus perturbaciones estructurales ε_t , las que implican cambios exógenos e inesperados sobre las relaciones económicas del sistema. Se asume que las innovaciones η_t son una combinación lineal de los shocks inobservables bajo la forma

$$\eta_t = H\varepsilon_t. \quad (2.3)$$

Si se supone además que las perturbaciones estructurales están incorrelacionadas

$$E\varepsilon_t\varepsilon_t' = \Sigma_\varepsilon = \begin{pmatrix} \sigma_{\varepsilon_1}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{\varepsilon_n}^2 \end{pmatrix} \quad (2.4)$$

sustituyendo (2.3) en (2.1) y (2.2) se llega a la formulación del SVAR tal que

$$A(L)Z_t = H\varepsilon_t \quad \text{ó} \quad B(L)Z_t = \varepsilon_t \quad (2.5)$$

donde $B(L) = H^{-1}A(L)$ si H^{-1} existe.

Es posible obtener la representación de Medias móviles estructurales (SMA)¹ tal que

$$Z_t = D(L)\varepsilon_t \quad \text{donde} \quad D(L) = C(L)H. \quad (2.6)$$

Bajo este esquema, los parámetros de (2.5) y (2.6) van a estar identificados si H y Σ_ε están identificados. Es más, como $D(L)$ es la proyección de Z_t sobre los valores pasados y corrientes de ε_t , se dice que se está en un problema de

³Por sus siglas en inglés relativas a Vector moving averages.

¹Por sus siglas en inglés relativas a Structural moving averages.

identificación de las perturbaciones estructurales. El problema de identificación implica identificar completamente a H . Para esto, se asume que el espacio generado por las innovaciones es el de las perturbaciones estructurales. Entonces, se supone que H es una matriz invertible,

$$\text{Si } H^{-1} \text{ existe entonces } \varepsilon_t = H^{-1}\eta_t. \quad (2.7)$$

Cuando se cumpla (2.7) se dice que el SVAR existe. Además, las ecuaciones (2.3) y (2.4) implican que

$$\Sigma_\eta = H\Sigma_\varepsilon H' \quad (2.8)$$

de donde se tienen n^2 parámetros en H y n en Σ_ε . Con esto, se tienen en total $n(n+1)$ parámetros libres pero como la matriz de covarianzas es simétrica, la cantidad de ecuaciones en (2.8) es $n(n+1)/2$. Entonces identificar a H y Σ_ε requiere $n(n+1)/2$ supuestos adicionales. Con una normalización de la escala de las perturbaciones $H_{jj} = 1$, se obtienen n restricciones y se dejan $n(n-1)/2$ restricciones adicionales para identificar a H . Bajo el supuesto de que los shocks son gaussianos, las restricciones en (2.8) son las únicas que se tienen que identificar. En caso contrario, se van a precisar supuestos adicionales (Stock y Watson (2016, p448)).

Del SVAR se desprenden las Funciones de respuesta al impulso estructurales (SFIR) de Z_t en sus shocks. Agregando el supuesto (2.4) del SMA también se puede calcular la FEVD. La SFIR es la trayectoria temporal del efecto causal dinámico sobre Z_{it} , la i -ésima componente de Z_t , que tiene un aumento en una unidad en el momento 0 de ε_{jt} , la j -ésima componente de ε_t . Sea D_h la h -ésima matriz de rezagos de los coeficientes en $D(L)$. Entonces $D_{h,ij}$ es el efecto causal sobre la i -ésima variable que tiene un aumento unitario en la j -ésima perturbación luego de h períodos. Por consiguiente, $D_{h,ij}$ es el efecto en Z_{it+h} de un aumento unitario en ε_{jt} . Por lo tanto, la función de respuesta al impulso estructural ($SIRF_{ij}$) es la secuencia de los coeficientes del SMA, que se expresa como

$$SIRF_{ij} = \{D_{h,ij}\}, \quad h = 0, 1, \dots, \quad \text{donde } D_h = C_h H \quad (2.9)$$

con $C(L) = A(L)^{-1}$ de (2.2) y $D(L) = C(L)H$ de (2.6). En el momento

$t = 0$ se tiene $D_0 = H$, llamado efecto contemporáneo o efecto de impacto¹.

Considerando h períodos tras el aumento inicial, se puede obtener la Función de respuesta al impulso acumulada (CSIRF)¹ como

$$CSIRF_{ij} = \sum_{k=0}^h D_{k,ij}.$$

La Descomposición de la Varianza del Error de Predicción ($FEVD_{h,ij}$) mide la importancia del j -ésimo shock para explicar la variación en Z_{it} . Se calcula la contribución relativa del shock a la varianza de los cambios inesperados en Z_{it} , h períodos hacia adelante, es decir, la varianza del error de predicción h pasos hacia adelante:

$$FEVD_{h,ij} = \frac{\sum_{k=0}^h D_{k,ij}^2 \sigma_{\varepsilon_j}^2}{var(Z_{i,t+h}|Z_t, Z_{t-1}, \dots)} = \frac{\sum_{k=0}^h D_{k,ij}^2 \sigma_{\varepsilon_j}^2}{\sum_{j=1}^n \sum_{k=0}^h D_{k,ij}^2 \sigma_{\varepsilon_j}^2} \quad (2.10)$$

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

Con respecto a la escala y el signo de las perturbaciones estructurales, como son inobservables es necesario normalizarlas. En este marco, un tipo de normalización es la de efecto unitario, que se expresa como

$$H_{jj} = 1. \quad (2.11)$$

Esta normalización implica fijar el signo y la escala del j -ésimo shock de forma que un aumento unitario en ε_{jt} induzca a un aumento unitario contemporáneo en una determinada variable observada como puede ser Z_{jt} . O sea, un aumento unitario en ε_t implica un aumento del mismo tamaño y signo en η_t ; lo que, a su vez, implica un aumento unitario en Z_{jt} . Para este estudio, como la FFR² está medida en puntos porcentuales, entonces un shock de política monetaria induce un aumento en una unidad porcentual en la mencionada tasa

¹En Ramey (2016), se expresa $D_h = \{D_{h,ij}\}$ a partir de $D_h = C_h H$. Con esto se representa la respuesta al impulso de la variable Z_i en un horizonte $t + h$ dado un shock ε_{jt} como

$$\frac{\partial Z_{i,t+h}}{\partial \varepsilon_{jt}} = D_{h,ij}$$

donde los parámetros $D_{h,ij}$ son funciones no lineales de los parámetros del VAR en su forma reducida.

¹Por sus siglas en inglés relativas a Cumulative structural impulse response function

²Por sus siglas en inglés relativas a Federal Funds Rate.

de interés de referencia.

Un esquema para identificar al sistema es el que propone Sims (1980), que se basa en ordenar a las variables de forma tal que la primera innovación responde sólo al primer shock en el período; la segunda innovación, por su parte, responde al primer y segundo shock en el período y así sucesivamente. Es un enfoque recursivo que refiere a cadenas causales de Wold (1954). En este contexto, se asume que H es triangular inferior, lo que implica que $H\Sigma_\varepsilon^{1/2} = Chol(\Sigma_\eta)$ donde $Chol$ es la factorización de Cholesky¹. Además, cuando se emplea la normalización de efecto unitario, se tiene que $H = Chol(\Sigma_\eta)\Sigma_\varepsilon^{-1/2}$ donde $\Sigma_\varepsilon = diag(\{[Chol(\Sigma_\eta)_{jj}]^2, j = 1, \dots, n\})$. Este tipo de identificación brinda $n(n-1)/2$ restricciones contemporáneas para la identificación del sistema. Las restricciones contemporáneas² se basan en criterios temporales y asumen el efecto de un determinado shock sobre cierta variable en el período³.

Aunque los modelos VAR son simples y útiles, han sido objeto de varias críticas. Primero, la estrategia de identificación dominante indica una única manera de identificar a los shocks, sin la necesidad de identificar al resto de las variables macroeconómicas del sistema (Bernanke et al. (2005)). Por lo anterior, en la literatura, han surgido estrategias alternativas para la identificación de los modelos (Christiano et al. (1999)). Segundo, este tipo de modelos examina sólo el efecto de cambios inesperados en la política monetaria sin considerar los posibles anuncios institucionales del cambio con su consiguiente efecto sobre las expectativas de los agentes, o la elección de la regla de política monetaria, entre otros (Stock y Watson (2016, p.460-461)). Tercero, la cantidad de variables que se pueden incluir en el modelo es limitada. Esto puede llevar a que las instituciones obtengan resultados contaminados porque la cantidad de parámetros aumenta rápidamente con la cantidad de variables que se agregan. Esto plantea un problema en los grados de libertad del modelo. Además, el VAR solo permite observar las respuestas a impulsos con un conjunto reducido de variables, lo que impide capturar completamente la dimensión de los shocks estructurales (Bernanke et al. (2005)). Las anteriores limitaciones podrían explicar algunos resultados desconcertantes⁴ en la

¹Se supone que los errores están mutuamente incorrelacionados.

²También llamadas de corto plazo.

³También se pueden identificar a los shocks sobre la base de restricciones de largo plazo, es decir, sobre el efecto de largo plazo de una determinada perturbación sobre cierta variable.

⁴Uno de ellos es el llamado *the price puzzle*. En el caso de una política monetaria contractiva, el nivel general de precios debería disminuir, sin embargo los resultados muestran

investigación de los VAR, donde los resultados empíricos no coinciden con lo que espera la teoría económica. Por último, otro motivo para querer entender los efectos de los shocks de política monetaria sobre un conjunto amplio de variables macroeconómicas se asocia al hecho de que ninguna serie temporal se asocia sólo a un único constructo teórico (Bernanke et al. (2005)).

Para hacer frente a las debilidades del VAR, Bernanke et al. (2005) proponen el modelo FAVAR. Este modelo permite condicionar el análisis VAR a conjuntos de datos más ricos en información, sin dejar de lado las ventajas estadísticas de restringir el análisis a un número reducido de series temporales. De esta forma, se combinan los análisis VAR estándar con uno de factores. Siguiendo la premisa de los DFM, que plantea que unos pocos factores estimados resumen una gran cantidad de información económica, una solución al problema de los grados de libertad en los VAR, consiste en aumentarlos por dichos factores. Como indica Klieber (2024), dado que los FAVAR se basan en el supuesto de que la dinámica de la economía puede ser capturada por una representación de baja dimensión, los factores latentes pasan a jugar un rol fundamental.

lo contrario (Bernanke et al. (2005)).

2.2. Modelos dinámicos de factores lineales

Stock y Watson (2016) proponen una generalización de los modelos SVAR al definir los modelos dinámicos de factores estructurales (SDFM)¹. Los mismos se enmarcan, a su vez, en el contexto de los DFM. El supuesto clave en estos modelos radica en que la dinámica común de un amplio conjunto de series económicas está vinculada a un número significativamente menor de factores latentes e inobservables.

Estos modelos representan un caso particular de los modelos de espacio de estado o modelos de Markov ocultos, para los cuales ciertas variables observables se expresan en términos de variables latentes inobservables que evolucionan de acuerdo a una dinámica rezagada con dependencia finita. De esta manera, se pueden resumir los movimientos complejos de un número potencialmente grande de series observables en un número menor de factores que conducen las fluctuaciones comunes de un determinado sistema económico.

Los modelos basados en factores se pueden expresar tanto en términos dinámicos como estáticos, lo que conlleva a dos métodos de estimación distintos. 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, mientras que para el caso estático, dicha representación es implícita. En ambos casos, se asume que tanto las variables observables como las latentes son estacionarias de segundo orden e integradas de orden cero.

A continuación, se presentan los DFM dinámico (2.12), estático (2.13), estructural (2.15) y el modelo FAVAR lineal (2.17)-(2.18).

¹Por sus siglas en inglés, Structural Dynamic Factor Models.

DFM dinámicos

En su forma dinámica, estos modelos se describen como un vector X_t de series temporales observables de tamaño $N \times 1$ que dependen de un número reducido de factores latentes e inobservables f_t de tamaño q (con $q \ll N$) y de un componente idiosincrático e_t con media cero¹.

El DFM dinámico (2.12), en Stock y Watson (2016), se representa como

$$X_t = \lambda(L)f_t + e_t, \quad (2.12a)$$

$$f_t = \Psi(L)f_{t-1} + \eta_t \quad (2.12b)$$

donde L es el operador de rezagos tal que $L^i f_t = f_{t-i}$ con $i = 0, 1, 2, \dots$ (Hamilton (1994, p. 25–42)). Entonces

$$\lambda(L) = \sum_{i=0}^{\infty} \lambda_i L^i$$

y

$$\Psi(L) = \sum_{i=0}^{\infty} \Psi_i L^i$$

son las matrices de polinomios de rezagos de tamaño $N \times q$ y $q \times q$, respectivamente. Por consiguiente, se tiene que

$$\lambda(L)f_t = \sum_{i=0}^{\infty} \lambda_i f_{t-i} \quad y \quad \Psi(L)f_{t-1} = \sum_{i=0}^{\infty} \Psi_i f_{t-1-i}.$$

La i –ésima fila de $\lambda(L)$, o sea a $\lambda_i(L)$, es la matriz de pesos de los factores dinámicos de la serie X_{it} . Entonces, el término $\lambda_i(L)f_t$ de (2.12a) es el componente común de esa i –ésima serie².

Por otro lado, η_t en (2.12b) representa el vector de innovaciones de tamaño $q \times 1$, con media cero, serialmente incorrelacionado con los factores f_t . Además, e_t en (2.12a) son las perturbaciones idiosincráticas, que surgen del error de medición y de características especiales específicas de las series individuales (Stock y Watson, 2011, p. 3). Las perturbaciones e_t se suponen incorrelacionadas con

¹En general, los factores latentes f_t y el término idiosincrático e_t van a estar serialmente correlacionados.

²Se va a suponer que $\lambda(L)$ es unidireccional, por lo que el componente común de cada serie va a estar determinado por los valores pasados y presentes de f_t .

el vector de innovaciones η_t en todos sus rezagos, esto es, $Ee_t\eta'_{t-k} = 0$ para todo k . Cuando las perturbaciones idiosincrásicas e_t en (2.12a) estén serialmente incorrelacionadas, el modelo (2.12) va a estar completamente especificado. En este caso, se dice que el DFM es exacto. En caso contrario, el DFM será aproximado.

DFM estáticos

El DFM estático¹ (2.13), en Stock y Watson (2016), se expresa como

$$X_t = \Lambda F_t + e_t, \quad (2.13a)$$

$$F_t = \Phi(L)F_{t-1} + G\eta_t. \quad (2.13b)$$

Este modelo reescribe el DFM dinámico (2.12) de manera que

- $F_t = (f'_t, f'_{t-1}, \dots, f'_{t-p})'$ es el vector de factores estáticos de tamaño $r \times 1$. Ahora, en vez de tener q factores f_t dinámicos, se tienen $r \geq q$ factores F_t estáticos.
- $\Lambda = (\lambda_0, \lambda_1, \dots, \lambda_p)$, siendo p el grado de la matriz de rezagos polinomiales $\lambda(L)$. Ahora, en el h -ésimo rezago de $\lambda(L)$, λ_h es una matriz de $N \times q$ coeficientes.
- El componente común para la i -ésima variable X_{it} es $\Lambda_i F_t$.
- $\Phi(L)$ es una matriz compuesta por unos, ceros y los elementos de $\Psi(L)$, lo que permite reexpresar el VAR de (2.12b) en términos de F_t .
- $G = [I_q \quad 0_{q \times (r-q)}]'$ si $r \geq q$. Cuando $r = q$ se tiene que $G = I$ entonces η_t son las innovaciones de los factores F_t . A lo largo de este estudio se supone que $r = q$, por lo que $G = I$.

Bajo esta forma estática, es posible realizar análisis por PC.

¹Se denomina también DFM apilado.

DFM estructurales

Partiendo de un DFM en su forma estática (2.13), cuando se supone que las innovaciones son una combinación lineal de unas perturbaciones estructurales inobservables tal que

$$\eta_t = H\varepsilon_t, \quad (2.14)$$

y que las perturbaciones estructurales están 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 SDFM (2.15), que en Stock y Watson (2016) se especifica como

$$X_t = \Lambda F_t + e_t, \quad (2.15a)$$

$$\Phi(L)F_t = G\eta_t \quad \text{donde } \Phi(L) = I - \Phi_1L - \dots - \Phi_pL^p \quad (2.15b)$$

$$\eta_t = H\varepsilon_t \quad (2.15c)$$

donde se tiene r factores estáticos y q factores dinámicos con $q \leq r$. En este sistema, las q perturbaciones estructurales ε_t impactan en los factores comunes pero no en el término idiosincrático e_t . Si se asume que H es invertible y que los shocks están mutuamente incorrelacionados (Σ_ε es diagonal) entonces se obtiene

$$X_t = \Lambda\Phi(L)^{-1}GH\varepsilon_t + e_t. \quad (2.16)$$

El efecto causal dinámico de las n variables al aumentar ε_t en una unidad es la SFIR que se expresa como $\Lambda\Phi(L)^{-1}GH$. Este tipo de modelos generaliza a los SVAR porque permite incorporar más variables que las perturbaciones estructurales y que cada variable tenga una dinámica idiosincrática que puede tener error de medida¹. Bajo el modelo (2.15), todos los factores son inobservables. Sin embargo, cuando alguno de los factores se considera observable, el SDFM se transforma en un FAVAR. Entonces, algunos factores van a ser

¹En este modelo, se requiere la normalización de Λ , G y H .

observados mientras que el resto sigue siendo latente. Como consecuencia, se imponen restricciones adicionales dado que al menos un factor va a estar medido sin error por al menos una variable observable. Como resultado se obtiene un tipo de modelo que es una combinación entre modelos de factores y los modelos VAR al mismo tiempo (Bernanke et al. (2005)).

FAVAR

Sea Y_t un vector de variables observables con efectos generalizados sobre la economía, de tamaño $M \times 1$. Si además se dispone de información adicional en unos factores inobservados¹ F_t de tamaño $r \times 1$ donde r es pequeño, entonces la dinámica conjunta de (F_t', Y_t') se puede representar por la siguiente *ecuación de transición*:

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + \eta_t, \quad (2.17)$$

donde $\Phi(L)$ es la matriz de rezagos polinomiales de orden p . El término de error η_t tiene media cero con matriz de covarianzas Q . La ecuación (2.17) es un VAR en (F_t', Y_t') . Se dice que (2.17) es un FAVAR (Bernanke et al. (2005)).

Bajo esta configuración, se supone además que se cuenta con un conjunto de series informacionales X_t , de tamaño N , siendo N grande² y mucho mayor a la cantidad de factores y variables observables en el sistema ($r + M \ll N$).

Se asume que las series de tiempo informacionales X_t se relacionan con F_t y Y_t mediante una *ecuación de observación* bajo la forma

$$X_t = \Lambda^f F_t + \Lambda^y Y_t + e_t \quad (2.18)$$

donde Λ^f es la matriz de pesos de los factores de tamaño $r \times N$, 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 (2.18) 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 la economía X_t .

En este estudio se va a identificar a una única innovación de política monetaria. Entonces, el factor observable Y_t va a ser la Tasa de Fondos Federales

¹Los factores inobservados se pueden considerar como conceptos difusos como la actividad económica o las condiciones crediticias.

²En particular, $N > T$ con T siendo la cantidad de períodos.

(FFR) y se la va a tratar como el instrumento de política monetaria partiendo de la idea de que tiene un efecto generalizado sobre X_t . Como consecuencia, se asume que Y_t no presenta errores de medición. Además, se va a asumir que los factores F_t capturan la actividad económica y los movimientos del nivel generalizado de precios. Con esto, para realizar la identificación del sistema, se adopta un esquema recursivo como en Bernanke et al. (2005), el cual establece un ordenamiento entre variables lentas primero y luego las rápidas (ver Apéndice 1). Con este ordenamiento de las variables se impone el siguiente supuesto de identificación: los factores latentes no responden a las perturbaciones de política monetaria en el período.

Las variables de movimiento lento, están en gran medida predeterminadas en el período actual, entonces, no se ven afectadas por los cambios de política monetaria en el período. Las de movimiento rápido son altamente sensibles a cambios contemporáneos relativos a perturbaciones o novedades en la economía. De esta manera, considerando esta categorización de variables, se plantea el siguiente esquema recursivo (2.19)

$$\begin{pmatrix} X_t^s \\ X_t^f \end{pmatrix} = \begin{pmatrix} \Lambda_{ss} & 0 & 0 \\ \Lambda_{fs} & \Lambda_{ff} & \Lambda_{fr} \end{pmatrix} \begin{pmatrix} F_t^s \\ F_t^f \\ r_t \end{pmatrix} + e_t, \quad (2.19a)$$

$$\begin{pmatrix} F_t^s \\ F_t^f \\ r_t \end{pmatrix} = \Phi(L) \begin{pmatrix} F_{t-1}^s \\ F_{t-1}^f \\ r_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_t^s \\ \eta_t^f \\ \eta_t^r \end{pmatrix}, \quad (2.19b)$$

$$\begin{pmatrix} \eta_t^s \\ \eta_t^f \\ \eta_t^r \end{pmatrix} = \begin{pmatrix} H_{ss} & 0 & 0 \\ H_{fs} & H_{ff} & 0 \\ H_{rs} & H_{rf} & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_t^s \\ \varepsilon_t^f \\ \varepsilon_t^r \end{pmatrix}. \quad (2.19c)$$

Además de la identificación del sistema, como los factores son inobservables, para que se pueda inferir el modelo es necesario imponer restricciones adicionales. Para esto, se aplica una normalización sobre la ecuación de observación (2.18), como se detalla en el capítulo 3, sección 3.2.

Como se está en el contexto de un DFM estático (2.13) exacto¹, la estimación del FAVAR se puede hacer mediante PC. Este tipo de estimación no

¹Bajo un DFM exacto se supone que las varianzas idiosincrásicas son homogéneas y los factores son tratados como parámetros a estimar (Stock y Watson (2016)).

paramétrica se basa en la idea de que cuando la cantidad de variables X_t es lo suficientemente grande, sólo con la variación que se da entre las mismas ya es posible estimar el espacio generado por los factores (Stock y Watson (2016, p. 429)).

Si bien los FAVAR se han establecido en la literatura como uno de los modelos de referencia para evaluar los impactos de cambios de política monetaria sobre la economía¹, recientemente han surgido otros enfoques desde la perspectiva del DL² que generalizan a los DFM lineales y al FAVAR (Andreini et al. (2023a) y Klieber (2024)). En este sentido, en una segunda etapa de este estudio, se va a estimar el Deep dynamic factor model (DDFM) propuesto por Andreini et al. (2023a). Posteriormente, así como en Klieber (2024), se van a incluir los factores estimados con el DDFM en un esquema VAR para inferir un modelo FAVAR profundo con el fin de compararlo con los resultados del FAVAR lineal presentado en esta sección.

¹En procedimientos de reducción dimensional.

²Por sus siglas en inglés relativas a Deep Learning.

2.3. Aprendizaje Automático, Profundo e Inteligencia artificial

Antes de pasar a describir el DDFM de la próxima sección 2.4, se introducen los conceptos de Inteligencia artificial (AI)¹, Aprendizaje automático (ML)² y Aprendizaje profundo (DL)³ presentados en Goodfellow et al. (2016). La AI busca automatizar, computacionalmente, trabajo rutinario. Pretende resolver tareas que son fáciles de resolver para las personas pero difíciles de describir de manera formal⁴. Para esto plantea como solución que las computadoras aprendan de la experiencia y entiendan el mundo en términos de una jerarquía de conceptos. Esto permite que la computadora aprenda abstracciones complejas a partir de otras más simples. Se dice que los conceptos se construyen unos encima de otros, en forma de capas, por lo que se define al presente enfoque como *AI deep learning* (Goodfellow et al. (2016, cap.1)). Los sistemas de AI utilizan técnicas de ML, fundamentadas en la estadística aplicada para extraer patrones de los datos. El desempeño de un modelo de ML va a depender de la representación que se realice de los mismos y del proceso de aprender dichas representaciones expresadas bajo la forma de abstracciones. El DL es un caso particular del ML y se distingue por el empleo de redes neuronales profundas. Esto implica, que se basa en conceptos complejos a partir de otros más simples⁵.

En este contexto, se dice que un algoritmo de ML es capaz de aprender de los datos, cuando a través de ciertas tareas, con la experiencia, mejora una cierta medida de desempeño. Las tareas indican cómo se debería procesar un ejemplo, que se puede representar como $\mathbf{x} \in \mathbb{R}^n$. Un ejemplo es una colección de características que han sido medidas cuantitativamente a partir de algún objeto o evento. Relacionado a la experiencia, los algoritmos de ML se pueden categorizar en supervisado y no supervisado. Por un lado, los modelos de tipo supervisado asocian una entrada \mathbf{x} a una determinada salida \mathbf{y} de datos. Experimentan con un conjunto de datos que contiene características y cada

¹Por sus siglas en inglés relativas a Artificial Intelligence.

²Por sus siglas en inglés relativas a Machine Learning.

³Por sus siglas en inglés relativas a Deep Learning.

⁴Como reconocer caras humanas en un conjunto de imágenes.

⁵Un sistema de DL puede representar el concepto de una imagen de una persona combinando ideas más simples, como esquinas y contornos, que a su vez se definen en términos de bordes.

ejemplo está asociado con un objetivo o etiqueta. Por otro lado, los de tipo no supervisado basan su experiencia en las características y luego aprenden propiedades útiles de la estructura del conjunto de datos. En este estudio, se realizan tareas de predicción de un conjunto de variables y la experiencia se relaciona a un proceso de aprendizaje supervisado. Lo anterior implica que se cuenta con un conjunto de variables objetivo a predecir. Como resultado, se busca que sus predicciones sean precisas. Las medidas de precisión⁶ permiten evaluar el desempeño de un determinado modelo. La precisión mide en qué medida una proporción de datos es estimada correctamente. Alternativamente, como medida de desempeño se puede emplear la tasa de error⁷ que mide la proporción de datos para los cuales el modelo obtiene un resultado incorrecto.

El principal desafío de los modelos de ML se basa en la capacidad que tiene el algoritmo de performar correctamente sobre datos nuevos que no hayan sido entradas previas del modelo, es decir, se busca poder generalizarlo. Para esto se puede dividir al conjunto de datos en entrenamiento y conjunto de prueba. Se dice que sobre el primero se entrena el modelo, mientras que en los datos de prueba se evalúa su desempeño. Del resultado de entrenar al modelo sobre el conjunto de datos de entrenamiento, se obtiene un error de entrenamiento¹. Sin embargo, el objetivo del ML es obtener un error de generalización lo más pequeño posible. En este sentido, el error de generalización² se define como el valor esperado del error en un conjunto de datos de prueba, distinto al conjunto de entrenamiento. Se asume que los conjuntos de datos de entrenamiento y prueba son *i.i.d.*, esto es, que son independientes e idénticamente distribuidos. Lo anterior implica que los datos de entrenamiento y prueba comparten la misma distribución de probabilidad subyacente p_{data} .

Considerando lo anterior, se puede afirmar que un modelo de ML presenta un buen desempeño cuando el error de entrenamiento es bajo y la diferencia entre este y el error de prueba es mínima. Relacionados a estas dos condiciones, surgen los conceptos de subajuste (underfitting) y sobreajuste (overfitting). El subajuste ocurre cuando el modelo no es capaz de tener un error lo suficientemente bajo en el conjunto de entrenamiento. El sobreajuste se da cuando la diferencia entre el error de entrenamiento y el de prueba es demasiado grande. Se puede controlar en qué medida un modelo sobreajusta o subajusta median-

⁶Accuracy en inglés.

⁷Error Rate en inglés.

¹Training error en inglés.

²También denominado error de prueba.

te la capacidad del mismo. La capacidad de un modelo va a estar dada por su habilidad en ajustarse a una gran variedad de funciones. Esto implica, que el desempeño de un algoritmo va a estar en gran medida determinado por la representación que se realice a partir conjunto de datos dado. Un algoritmo se va a desempeñar mejor cuando su capacidad sea la apropiada en relación a la verdadera complejidad de la tarea a realizar y a la disponibilidad de datos para el entrenamiento. Modelos con baja capacidad van a ser incapaces de resolver tareas complejas. Sin embargo, cuando su capacidad sea mayor a la necesaria es probable que sobreajuste (Goodfellow et al. (2016, cap.5)).

A grandes rasgos, el ML descansa sobre algunos principios, teoremas y conceptos clave. Uno de ellos es el concepto de parsimonia, conocido como La Navaja de Occam. Establece que si varios modelos son buenos para explicar las observaciones, entonces se debería elegir al más sencillo. El teorema No Free Lunch establece que cualquier algoritmo de ML es capaz de generalizar bien a partir de un conjunto finito de entrenamiento (Wolpert y Macready (1997)). Esto implica que ningún algoritmo de ML es universalmente mejor que otro. Como consecuencia se debería buscar el que mejor performe sobre una tarea específica. De este modo, la elección del modelo a emplear va a depender también de un conjunto de preferencias personales que deberían estar alineadas con el problema a resolver. Las preferencias se conocen como regularizaciones y son todas aquellas modificaciones que se le realizan al algoritmo de ML con el propósito de reducir el error de generalización sin afectar al de entrenamiento¹ (Goodfellow et al. (2016, cap.5)).

Si bien la mayoría de los algoritmos de ML funcionan correctamente sobre una gran variedad de problemas, pueden presentar ciertas limitaciones para algunas tareas de la AI². Como consecuencia, el desarrollo del DL se ve impulsado en parte por la necesidad de abordar los desafíos que enfrenta el ML (Goodfellow et al. (2016, p.152-161)). En este sentido, un algoritmo de DL está compuesto, en general, por datos, una función de costo (o pérdida) que puede incluir regularizaciones, un procedimiento de optimización y un modelo. Para escenarios no lineales, no es posible emplear funciones de costo para soluciones cerradas por lo que se emplean, típicamente, métodos de optimización basados en gradientes (para más detalles ver Anexo 3).

¹Por el teorema de *No Free Lunch*, así como no hay un algoritmo que sea mejor que otro, tampoco existe una regularización universal que supere al resto.

²Cuando aumenta la cantidad de datos o falla la consistencia local a priori, entre otros.

En este marco, las redes profundas feedforward³ constituyen la piedra angular del DL. Se basan en la aproximación de funciones. Esto implica, que se define un mapeo $\mathbf{y} = f(\mathbf{x}; \boldsymbol{\theta})$ y se aprende el valor del parámetro $\boldsymbol{\theta}$ que resulta de la mejor función aproximada. Se habla de redes porque típicamente se representan mediante la composición de varias funciones diferentes⁴. El modelo se asocia a un grafo acíclico directo y describe cómo las funciones se componen conjuntamente. Se dice que las funciones están compuestas en cadena y se denominan redes neuronales. Las mismas se representan por capas de funciones y la cantidad de capas brinda la profundidad de la red. El nombre *deep learning* se desprende de estos conceptos (Goodfellow et al. (2016, cap.6)). Este tipo de modelos se llama feedforward⁵ porque la información fluye a través de la función que se evalúa en \mathbf{x} mediante computaciones intermedias que se emplean para definir f , para finalmente calcular la salida \mathbf{y} . Esto implica que durante el entrenamiento, la propagación hacia adelante puede continuar hasta que produce un escalar $J(\boldsymbol{\theta})$. A su vez, para actualizar los pesos de la red neuronal durante la fase de entrenamiento, se emplea un algoritmo de optimización llamado Back-propagation que calcula el gradiente de la función de costo con respecto a los parámetros del modelo $\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta})$, con el objetivo de minimizar la función de pérdida. El mismo está relacionado a los conceptos de grafos computacionales, operaciones y a la Regla de la cadena (Goodfellow et al. (2016, cap.6)). En la próxima sección 2.4 se describe el el DDFM y en el capítulo 3 se desarrollan los métodos de estimación y optimización empleados.

³También llamadas redes neuronales feedforward o Perceptrones multicapa.

⁴ Siguiendo el ejemplo de Goodfellow et al. (2016, p.164), se puede pensar el caso de tres funciones $f^{(1)}$, $f^{(2)}$, $f^{(2)}$ conectadas por una cadena que forman $f(x) = f^{(3)}(f^{(2)}(f^{(1)}(x)))$ donde $f^{(1)}$ es la primer capa, $f^{(2)}$ es la segunda y $f^{(3)}$ la tercera.

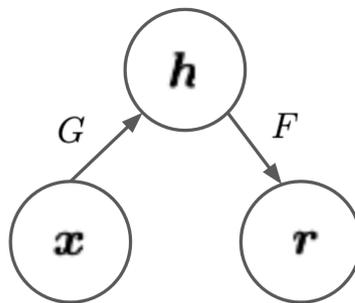
⁵ En este contexto no hay conexiones de retroalimentación para las cuales las salidas se retroalimenten así mismas, cuando se permiten conexiones de retroalimentación se está en el marco de redes neuronales recurrentes (Goodfellow et al. (2016, cap.10)).

2.4. Modelos dinámicos de factores profundos

DDFM

En Andreini et al. (2023a), se plantea que desde la perspectiva del Aprendizaje profundo (DL), los DFM podrían presentar ciertas debilidades al asumir una estructura lineal y al enfrentar problemas de escalabilidad computacional cuando se incrementa la cantidad de variables. Los autores proponen comparar los resultados de los DFM lineales con aquellos basados en un modelo de DL para evaluar si son capaces de superar dichas dificultades. En este sentido, se busca mantener al mismo tiempo el mismo grado de flexibilidad e interpretabilidad que un modelo lineal pero mejorando el desempeño de los resultados. De esta manera, en la presente sección se describe el Modelo dinámico de factores profundo (DDFM) basado en Autoencoders (AE).

Siguiendo a Goodfellow et al. (2016, cap.14), los AE pueden considerarse como un caso particular de redes neuronales feedforward y pertenecen al campo del aprendizaje no supervisado. Son redes neuronales entrenadas con el propósito de copiar, de manera aproximada, las entradas en salidas. La red se compone de dos partes: una función de encoder¹ que se define como $\mathbf{h} = G(\mathbf{x})$ y una función de decoder² que produce una reconstrucción tal que $\mathbf{r} = F(\mathbf{h})$. Internamente, poseen capas ocultas \mathbf{h} que describen un código para representar a las entradas. Las redes neuronales profundas se basan en el concepto de grafos. En este sentido, la estructura general de un AE se puede representar como en la siguiente Figura 2.1.



Fuente: Goodfellow et al. (2016)

Figura 2.1: Estructura general de un AE.

¹En español se denomina codificador.

²En español se denomina decodificador.

Los AE tradicionalmente se han empleado en esquemas de reducción de la dimensionalidad. En este marco, la representación en baja dimensión fuerza al AE a capturar las características más importantes de los datos. Es por esto que pueden considerarse como una generalización no lineal de componentes principales (PC). Una forma de obtener las características más sobresalientes de \mathbf{x} a partir de las entradas es restringiendo \mathbf{h} de manera que tenga un tamaño menor que \mathbf{x} . Como resultado, la dimensión del código es menor que la dimensión de la entrada y se dice que el AE es subcompleto³.

El proceso de aprendizaje se describe, a grandes rasgos, como la minimización de una función de pérdida \mathcal{L} , que puede ser el Error cuadrático medio (MSE)¹, tal que $\mathcal{L}(\mathbf{x}, F(G(\mathbf{x})))$ penaliza a $F(G(\mathbf{x}))$ por no ser similar a \mathbf{x} . Cuando el decoder es lineal y \mathcal{L} es el MSE, un AE subcompleto aprende a generar el mismo subespacio que PC. En por esto que AE con estructuras no lineales en el codificador y en el decodificador se entienden como generalizaciones no lineales de PC.

Considerando lo anterior, para describir el DDFM propuesto por Andreini et al. (2023a), se parte de una formulación general de modelos de factores latentes con componentes idiosincrásticos. Se define $\mathbf{x}_t = (x_{t,1}, \dots, x_{t,N})$ como el vector que recoge las N variables de interés en el momento t . Con esto, un modelo de factores latentes general se puede expresar como

$$\mathbf{x}_t = F(\mathbf{f}_t) + \mathbf{e}_t = \tilde{\mathbf{x}}_t + \mathbf{e}_t, \quad (2.20)$$

donde \mathbf{f}_t es un vector de componentes comunes estocásticas, de tamaño $r \times 1$, con $r = \dim(\mathbf{f}) \ll \dim(\mathbf{x})$. Por su parte, \mathbf{e}_t es el término de perturbaciones estocásticas y $F(\cdot)$ es una función genérica que mapea a los factores inobservados en las variables observables. Se asume que \mathbf{f}_t y \mathbf{e}_t son independientes, con media cero y varianza finita. Por su parte, $\tilde{\mathbf{x}}_t$ es el componente de \mathbf{x}_t que se relaciona con los factores comunes². Cuando $F(\cdot)$ no sea lineal, se puede expresar al componente de factores como

$$\tilde{\mathbf{x}}_t = F(G(\mathbf{x}_t)) = (F \circ G)(\mathbf{x}_t) = (F \circ G)(\tilde{\mathbf{x}}_t + \mathbf{e}_t) \quad (2.21)$$

donde

³En inglés el término es undercomplete.

¹Por sus siglas en inglés relativas a Mean Square Error.

²Cuando se asuma que $F(\cdot)$ es una función lineal, el modelo se reduce a un modelo de factores lineal estándar.

- $G(\cdot)$ es la función de encoding: mapea a las observables en el código \mathbf{f}_t .
- $F(\cdot)$ es la función de decoding: mapea de vuelta a los factores en \mathbf{x}_t .

Los AE resuelven el problema paramétrico de encontrar un mapeo¹ de la forma $\tilde{\mathbf{x}}_t = F(G(\mathbf{x}_t))$ bajo la restricción de minimizar una función de pérdida tal que

$$\mathcal{L}(\mathbf{x}_t, \tilde{\mathbf{x}}_t; \boldsymbol{\theta}) = \mathcal{L}(\mathbf{x}_t, F(G(\mathbf{x}_t))), \quad (2.22)$$

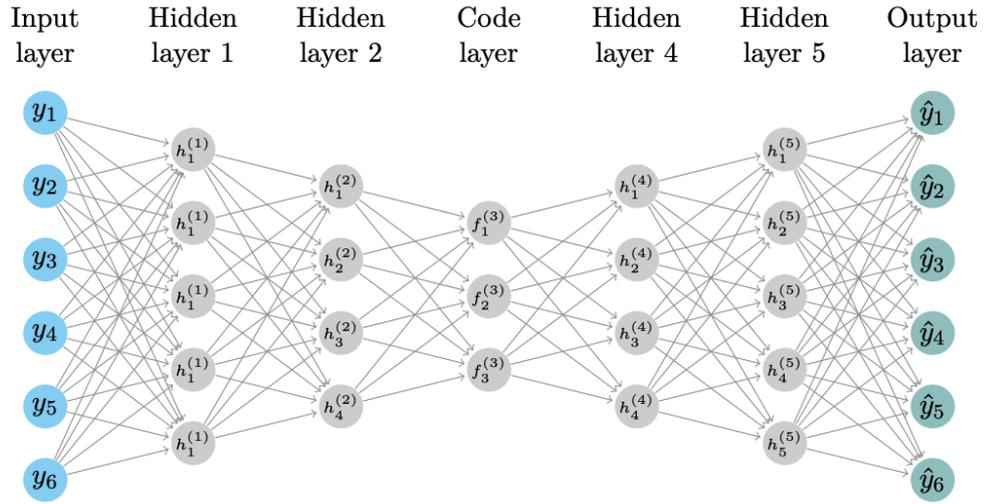
donde $\mathcal{L}(\cdot)$ es la función de pérdida y $\boldsymbol{\theta}$ es el vector que contiene todos los parámetros en $G(\cdot)$ y $F(\cdot)$. En este marco, el análisis de PC se puede interpretar como un AE que minimiza una función de pérdida tal que

$$\mathcal{L}(\mathbf{x}_t, \tilde{\mathbf{x}}_t; \boldsymbol{\theta}) = \|\mathbf{x}_t - \tilde{\mathbf{x}}_t\|^2. \quad (2.23)$$

Como en principio $G(\cdot)$ y $F(\cdot)$ pueden ser cualquier función no lineal, encontrar una forma funcional que recoja el proceso generador de los datos podría ser un problema en sí mismo. Los AE brindan una implementación práctica para este problema al expresar la composición de dos funciones como una cadena de Perceptrón multicapa (MLP). En la primer cadena, opera el encoding mientras que en la segunda, se produce la salida de decoding.

En la siguiente Figura 2.2, se tiene un AE simétrico, con 6 observables y 3 neuronas en la capa de código. Las dos primeras capas ocultas operan como el encoding mientras que en las dos últimas se decodifican los datos para obtener un resultado final. La elección del tipo de capa de salida está estrechamente ligada a la decisión que se realiza del tipo de función de pérdida. En este sentido, la capa de salida juega el rol de aportar información adicional a las características del modelo con el fin de completar la tarea del algoritmo (Goodfellow et al. (2016, p.177)).

¹Lo que es equivalente a aprender una representación.



Fuente: Andreini et al. (2023a, p.11)

Figura 2.2: Ejemplo de AE simétrico.

Un MLP es un tipo de red neuronal artificial feedforward compuesta por varias capas ocultas¹. Cada una está conformada por un conjunto de neuronas². Cada capa oculta de la red suele tener un valor vectorial y la dimensión de las capas ocultas determina el ancho del modelo. Cada capa consta de un conjunto de neuronas que actúan en paralelo, cada una de las cuales representa una relación vector-escalar. Cada neurona en cada capa recibe ciertas entradas de las neuronas de la capa anterior y devuelve, para la siguiente capa, una salida de activación (h_{ml}^l). La función de activación³ $g_{ml}^l(\cdot)$ de cada neurona⁴ es una función no lineal parametrizada, que se expresa como

$$h_{ml}^l = g_{ml}^l (\mathbf{W}_{ml}^l \mathbf{h}^{l-1} + b_{ml}^l), \quad (2.24)$$

donde l es la capa para $l = 1, \dots, L$, m_l es el nodo y $\theta_{ml}^l \equiv \{\mathbf{W}_{ml}^l, b_{ml}^l\}$ son los parámetros de la función de activación a determinar bajo la forma de un conjunto de pesos y una constante (o sesgo). En la literatura se cuenta con una variedad de funciones de activación. El diseño de unidades ocultas es un campo de investigación muy activo y aún no cuenta con muchos principios

¹Las capas ocultas son las capas intermedias del modelo que se diferencian de la capa de entrada y de salida.

²También llamadas nodos o unidades.

³En inglés se llama es link function.

⁴O unidad oculta.

teóricos rectores definitivos. Sin embargo, el enfoque de aproximación universal brinda una solución al problema de la elección de las funciones de activación al relacionar a las redes neuronales feedforward con las capas ocultas (Goodfellow et al. (2016, p.187-197)).

El Teorema de Aproximación Universal establece que cualquier red neuronal feedforward que cuente con una capa de salida lineal y al menos una capa oculta con una función de activación arbitraria, será capaz de aproximar cualquier función en un espacio compacto de \mathbb{R}^n , para un error distinto de cero, siempre y cuando la red tenga un número suficiente de unidades ocultas. También, las derivadas de la red feedforward pueden aproximar correctamente a las derivadas de la función seleccionada. En suma, lo anterior implica que independientemente de la función que se trate de aprender, un MLP grande podrá representar a esa función. Sin embargo, no está garantizado que el algoritmo de entrenamiento sea capaz de aprender esa función particular (Cybenko (1989), Hornik et al. (1989), Hornik et al. (1990), Lu et al. (2017), Goodfellow et al. (2016), Andreini et al. (2023a)).

A nivel de cada neurona en las capas de la red, se aplica la función de activación elegida sobre una transformación afín tal que $\mathbf{z} = g(\mathbf{W}'\mathbf{x} + \mathbf{b})$, siendo \mathbf{x} el vector de entradas de las unidades ocultas. Por lo tanto, cada neurona recibe un vector de entradas \mathbf{x} y calcula una transformación afín para luego, aplicar una función de activación $g(\mathbf{z})$, elemento a elemento. En este estudio considero dos funciones de activación: la tangente hiperbólica (*tanh*) y la Rectified linear units (*ReLU*). En Goodfellow et al. (2016, cap.6), la función *ReLU* se expresa como

$$g(z) = \max\{0, z\},$$

mientras que la *tanh* como

$$g(z) = \tanh(z) = 2\sigma(2z) - 1,$$

con $g(z) = \sigma(z)$ representando la función de activación sigmoide logística.

Se define $\mathbf{g}_i(\cdot)$ como el vector que contiene a todas las funciones de activación en una capa $\{g_1^l(\cdot) \dots g_M^l(\cdot)\}'$ para cada nodo $1, \dots, M$. Por lo tanto, el primer MLP estará dado por la composición de las funciones de activación de cada nodo, en cada capa, de la red de encoding feedforward

$$\mathbf{f}_t = G(\mathbf{x}_t) = \mathbf{g}_L(\mathbf{g}_{L-1}(\dots(\mathbf{g}_1(\mathbf{x}_t))))). \quad (2.25)$$

Por otro lado, también se define al MLP que funciona como un decodificador de la red. Es decir, una secuencia de capas donde cada una contiene neuronas que operan como funciones de activación sobre la suma ponderada de las entradas, más un sesgo

$$\tilde{\mathbf{x}}_t = F(\mathbf{x}_t) = \tilde{\mathbf{g}}_{L'}(\tilde{\mathbf{g}}_{L'-1}(\dots(\tilde{\mathbf{g}}_1(\mathbf{f}_t)))) \quad (2.26)$$

donde $\tilde{\mathbf{g}}_{L'}$ es el vector de funciones de activación y L' es la cantidad de capas ocultas en la red de decoding. Cuando la cantidad de capas ocultas en el encoding es igual a la cantidad de capas ocultas en el decoding $L = L'$ se dice que el AE es simétrico, de lo contrario es asimétrico. Los AE asimétricos, por lo general, presentan varias capas en el encoding, pero una sola al decodificar, o sea, $L' = 1$. Esto implica que la red de decoding es un Perceptrón de una capa (SLP)¹.

La formulación (2.20) con (2.25) y (2.26) supone un mapeo estático para estimar los factores. Una forma de introducir un comportamiento dinámico en los DFM consiste en asumir que \mathbf{f}_t y \mathbf{e}_t son generados por ecuaciones de vectores estocásticos lineales en diferencias. Esto implica que tanto \mathbf{f}_t como \mathbf{e}_t se suponen procesos $VAR(p)$ y $VAR(d)$, respectivamente. En particular, Giannone et al. (2010), Bańbura y Modugno (2014) especifican el siguiente sistema (2.27)

$$\mathbf{x}_t = \Lambda \mathbf{f}_t + \mathbf{e}_t, \quad (2.27a)$$

$$\mathbf{f}_t = \mathbf{B}_1 \mathbf{f}_{t-1} + \dots + \mathbf{B}_p \mathbf{f}_{t-p} + \mathbf{u}_t, \quad \mathbf{u}_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{U}) \quad (2.27b)$$

$$\mathbf{e}_t = \Phi_1 \mathbf{e}_{t-1} + \dots + \Phi_d \mathbf{e}_{t-d} + \epsilon_t, \quad \epsilon_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{Q}) \quad (2.27c)$$

donde (2.27a) es una función lineal, $\mathbf{B}_1, \dots, \mathbf{B}_p$ es la matriz de coeficientes autorregresivos, de tamaño $r \times r$, de los factores y Φ_1, \dots, Φ_d es la matriz diagonal de coeficientes autorregresivos, de tamaño $n \times n$, del componente idiosincrático. Tomando estas ideas, Andreini et al. (2023a) expresan un DDFM en un esquema general al cual aplican restricciones a la hora de implementarlo empíricamente. En su forma general, el DDFM (2.28) planteado en Andreini et al. (2023a) se describe como

¹Por sus siglas en inglés relativas a Single Layer Perceptron.

$$\mathbf{f}_t = G(\mathbf{x}_t) \quad (2.28a)$$

$$\mathbf{x}_t = F(\mathbf{f}_t) + \mathbf{e}_t \quad (2.28b)$$

$$\mathbf{f}_t = \mathbf{B}_1 \mathbf{f}_{t-1} + \dots + \mathbf{B}_p \mathbf{f}_{t-p} + \mathbf{u}_t, \quad \mathbf{u}_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{U}) \quad (2.28c)$$

$$\mathbf{e}_t = \Phi_1 \mathbf{e}_{t-1} + \dots + \Phi_d \mathbf{e}_{t-d} + \epsilon_t, \quad \epsilon_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{Q}). \quad (2.28d)$$

El DDFM permite una estructura no lineal en (2.28a) y (2.28b) mientras que mantiene los supuestos de linealidad en la parte dinámica de las 2 últimas ecuaciones (2.28c) y (2.28d). Empíricamente, si se supone un $VAR(p = 2)$ para (2.28c), un $VAR(d = 1)$ para (2.28d) y la función de decoding (2.28b) bajo la forma $F(\mathbf{f}_t) = \mathbf{\Lambda} \mathbf{f}_t$, el modelo final implementado DDFM (2.29) puede expresarse como

$$\mathbf{f}_t = G(\mathbf{x}_t) \quad (2.29a)$$

$$\mathbf{x}_t = \mathbf{\Lambda} \mathbf{f}_t + \mathbf{e}_t \quad (2.29b)$$

$$\mathbf{f}_t = \mathbf{B}_1 \mathbf{f}_{t-1} + \mathbf{B}_2 \mathbf{f}_{t-2} + \mathbf{u}_t, \quad \mathbf{u}_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{U}) \quad (2.29c)$$

$$\mathbf{e}_t = \Phi_1 \mathbf{e}_{t-1} + \epsilon_t, \quad \epsilon_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \mathbf{Q}). \quad (2.29d)$$

Bajo este esquema, los parámetros del modelo $\boldsymbol{\theta} = (\mathbf{\Lambda}, \mathbf{B}(L), \Phi(L), \mathbf{Q})$ se pueden estimar con el Método de gradientes Monte Carlo¹. La implementación del modelo (2.29) se encuentra en el capítulo 3, sección 3.2.

¹Se emplea este método de estimación por presentar ventajas computacionales, es decir es un método rápido y confiable, entre otros motivos (Andreini et al. (2023a)).

FAVAR profundo

Klieber (2024) presenta una forma general para expresar un modelo FAVAR con la siguiente *ecuación de observación*

$$\mathbf{x}_t = g(\mathbf{f}_t, \mathbf{e}_t), \quad \mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_e) \quad (2.30)$$

donde \mathbf{x}_t es el vector de variables informacionales de tamaño N y g es una función genérica que se puede aproximar con técnicas de reducción de la dimensionalidad. Por su parte, la *ecuación de transición* se puede definir como

$$\begin{aligned} \mathbf{z}_t &= [\mathbf{f}'_t, \mathbf{y}'_t] \\ VAR(p) : \quad \mathbf{z}_t &= \mathbf{c} + \mathbf{A}_1 \mathbf{z}_{t-1} + \dots + \mathbf{A}_p \mathbf{z}_{t-p} + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_v) \end{aligned} \quad (2.31)$$

donde \mathbf{z}_t es un vector de variables endógenas de tamaño $K = r + M$ compuesta por los factores latentes \mathbf{f}_t de tamaño r a estimar y el factor observable \mathbf{y}_t de tamaño M . Además, $\mathbf{A}_1, \dots, \mathbf{A}_p$ son matrices de tamaño $K \times K$ con los coeficientes del modelo en su forma reducida y $\boldsymbol{\Sigma}_v$ es de $K \times K$.

El FAVAR profundo incorpora los factores estimados por el modelo DDFM (2.29), con los que se estima un VAR en $(\hat{\mathbf{f}}'_t, \mathbf{y}'_t)$. En este esquema, los factores estimados se representan de la siguiente forma

$$\hat{\mathbf{f}}_t = \left(h_1^{W(1), b_1} \circ \dots \circ h_L^{W(L), b_L} \right) (\mathbf{x}_t) \quad (2.32)$$

con

$$h_l^{W(l), b_l} = h_l \left(\sum_{i=1}^{m_l} \mathbf{W}_{.i}^{(l)} \hat{\mathbf{d}}_i^{(l)} + b_l \right), \quad 1 \leq l \leq L$$

donde $l \in \{1, \dots, L\}$ es la cantidad de transformaciones no lineales, L es la cantidad de capas ocultas en el encoder, m_l es la cantidad de neuronas por capa. La primer capa se expresa como $\hat{\mathbf{x}}^{(1)} = \mathbf{x}$ donde $\hat{\mathbf{d}}_i^{(l)}$ es la i -ésima columna de la matriz $\hat{\mathbf{x}}^{(l)}$. Por su parte, \mathbf{W} y b_l son los parámetros de las funciones de activación a estimar en cada capa l .

En la sección 3.2 se estima un modelo FAVAR profundo.

Capítulo 3

Estrategia empírica

3.1. Datos del estudio

Fuente del conjunto de datos

El conjunto de datos empleado en el presente estudio es el propuesto por McCracken y Ng (2016). Los datos surgen de un trabajo colaborativo con la Reserva Federal de St. Louis y se emplean Datos Económicos de la Reserva Federal (FRED)¹ para la economía de Estados Unidos. Como consecuencia, la *FRED-MD*² es una amplia base de datos macroeconómica con frecuencia mensual que tiene como objetivo establecer un punto de partida común para el análisis empírico en el marco de lo que se denomina *Big Data*.

La base de datos *FRED-MD* se asemeja a lo que se conoce en la literatura como *el conjunto de datos de Stock y Watson*. En este sentido, se cuenta con más de un centenar de series temporales que se agrupan en 8 categorías de variables: «Producto e ingresos», «Consumo, órdenes e inventarios», «Mercado de trabajo», «Vivienda», «Dinero y crédito», «Mercado de activos», «Tasas de interés y tipos de cambio» y «Precios» (McCracken y Ng (2016), Chen et al. (2023)).

La recolección del conjunto de datos está guiada por cuatro recomendaciones (Stock y Watson (1996)): primero, la muestra debería incorporar los principales agregados mensuales y sus indicadores relacionados; segundo, debería incluir como consecuencia los indicadores económicos considerados importantes y claves; tercero, los datos deberían representar un amplio conjunto de variables con distintas propiedades de series de tiempo; cuarto, los datos deberían tener definiciones históricas consistentes y, en caso contrario, que sean sujetas a revisiones y actualizaciones. Siguiendo estas sugerencias, a través de la *FRED-MD* se busca brindar una base de datos accesible, actualizada y común para la comunidad científica (McCracken y Ng (2016)).

¹Por sus siglas en inglés relativas a Federal Reserve Economic Data.

²Base de datos mensual para la investigación macroeconómica.

Preprocesamiento de las series de tiempo

Carga y transformación de las series temporales

Los datos se cargan y preprocesan usando distintos paquetes de R¹, principalmente el paquete `fbi`², propuesto por Chen et al. (2023), contiene funciones para estimar modelos de factores, para tratar datos atípicos, imputar datos faltantes y para cargar y preparar al conjunto de datos *FRED-MD*. Se emplea el conjunto de datos de ejemplo del paquete: se cargan los mismos y se manipulan realizando las transformaciones sugeridas con el objetivo de que las series temporales sean estacionarias. Cada serie temporal se asocia a un código de transformación denominado `tcode`. La descripción de cada columna de datos y sus transformaciones se pueden hallar en el Apéndice 1.

Se obtiene, en total, un conjunto de datos con 128 columnas relativas a 127 variables macroeconómicas y una columna temporal mensual. Como mencioné en la sección 3.1, las variables se clasifican en 8 categorías. Si bien su clasificación tiene como antecedente estudios de Stock y Watson (McCracken y Ng (2016)), las variables se agrupan en base a criterios similares a los que se encuentran en Ludvigson y Ng (2010)³.

Una vez que se consiguen los datos transformados, se pasa a remover los datos atípicos por fila, empleando la función `rm_outliers.fredmd`. Luego, se eliminan columnas con 5% o más de datos faltantes (NA). Como consecuencia, se pasa a tener una grilla de datos de 751 filas con 123 columnas. Posteriormente, se chequea la presencia de NA al inicio y al final de la grilla de datos y se eliminan filas hasta no tener más NA en los extremos de la misma. Una vez realizado lo anterior, se pasa a imputar los restantes NA empleando la función `na_kalman` del paquete `imputeTS`, propuesto por Moritz y Bartz-Beielstein (2017). Este paquete sirve para imputar datos faltantes en series univariadas. Para este estudio se emplearon los argumentos `model = "StructTS"`, `smooth = TRUE` de la función `na_kalman`, por lo que se ajusta a cada serie un modelo estructural por máxima verosimilitud y se suavizan los datos empleando el filtro de Kalman.

Siguiendo con el preprocesamiento de los datos, se pasa a chequear que la

¹R Core Team (2023)

²Factor-Based Imputation and FRED-MD/QD Data Set

³Mientras que en Stock y Watson (2005) y Stock y Watson (2006), se organizan los datos en 14 categorías, Ludvigson y Ng (2010) actualizan los datos de Stock y Watson y emplean categorías de datos más amplias (McCracken y Ng (2016)).

columna de fechas (`date`) sea completa, o sea, considerando la fecha mínima (t_{min}) y máxima (t_{max}) de la muestra, que no haya ninguna fecha intermedia t faltante. Realizado lo anterior, se obtiene un conjunto de datos con 722 filas y 123 columnas (incluyendo la columna temporal de fecha), por lo que se cuenta con 122 variables económicas en total. La fecha de inicio del estudio es 1960-01-01 y la fecha máxima es 2020-02-01 (ir al [código](#)).

Análisis de raíces unitarias y estacionariedad de cada serie temporal

Dado que la política monetaria en Estados Unidos ha cambiado significativamente en los últimos 20 años como resultado de varias crisis económicas y especulativas, y que los datos de este estudio abarcan un período de tiempo considerablemente largo, no sería esperable que las series sean estacionarias en todo el período analizado. En estudios posteriores, se podría realizar el análisis para ventanas de tiempo más reducidas asociadas a los distintos momentos históricos de la política monetaria estadounidense.

En la presente subsección se chequea que las series resultantes del preprocesamiento anterior sean estacionarias para todo el período considerado (para entender el concepto de estacionariedad ver Enders (2014, cap.3)). En la literatura existen varios métodos para testear la presencia de RU en series temporales. Sin embargo, en esta investigación, en un primer momento, se va a considerar la prueba Prueba de Dickey-Fuller aumentada (ADF)¹ para poner a prueba la hipótesis nula de existencia de RU en la serie temporal z_t (ver Anexo 1). Como desde el punto de vista empírico, las pruebas de RU pueden fallar en rechazar la hipótesis nula de RU cuando en realidad la misma se cumple (Shin et al. (1992)), también se va a emplear una prueba alternativa, denominada Prueba de Kwiatkowski, Phillips, Schmidt y Shin (KPSS). Para esta prueba, la hipótesis nula es de estacionariedad contra una alternativa de no estacionariedad para cada serie temporal (ver Anexo 1). Cuando los resultados de ambos test indican que una serie z_t es estacionaria para un determinado nivel de confianza, se puede concluir que los resultados son confirmatorios, en caso contrario son robustos.

Para llevar a cabo las pruebas para cada una de las series temporales del estudio, empleo el paquete `arch` de `python` desarrollado por Sheppard et al. (2023) para crear dos clases: `ModelsADF` y `KPSSAnalysis`. La clase `ModelsADF`

¹Por sus siglas en inglés relativas a Augmented Dickey-Fuller.

permite realizar las pruebas ADF considerando los modelos con constante y con tendencia (`ct`), con constante determinística (`c`) y sin constante ni tendencia (`n`). En cada caso se obtiene la cantidad de series que no presentarían RU (y las que efectivamente presentarían) con un 95 % de confianza y la lista de variables macroeconómicas (ir al [código](#)). La otra clase `KPSSAnalysis` sirve para realizar las pruebas KPSS para distintas formas de estimar la cantidad de rezagos y para las hipótesis nulas de estacionariedad alrededor de una tendencia (`ct`) y una constante (`c`). Para esta prueba se consigue también la cantidad de series no estacionarias, estacionarias y la lista de variables en cada caso, para el mismo nivel de confianza (ir al [código](#)).

Resultados de las pruebas ADF y KPSS: Partiendo de un modelo general (1.1) para la prueba ADF, se tiene que un total de 118 series no presentarían RU con un 95 % de confianza, por lo que se concluye que son estacionarias. Sólo cuatro series tendrían RU, entonces serían no estacionarias y las mismas son HOUSTNE (Housing Starts, Northeast), HOUSTMW (Housing Starts, Midwest), HOUSTS (Housing Starts, South), PERMITMW (New Private Housing Permits, Midwest (SAAR)). Los resultados se detallan en el Apéndice 2 y el código se encuentra disponible en [este enlace](#). Con respecto a los resultados de la prueba KPSS, se establecieron cuatro modelos para determinar la cantidad de rezagos (`legacy`, `auto`, `Newey West`, y `Schwert`) para las hipótesis nulas de estacionariedad alrededor de una constante y una tendencia (ver Apéndice 2 y el código se encuentra disponible en [este enlace](#)).

Dado que se espera que las series sean estacionarias porque se aplicaron las transformaciones sugeridas, como se indicó antes, en McCracken y Ng (2016), se elige un resultado de la prueba en base a la mayor cantidad de series que son estacionarias con un 95 % de confianza. Como conclusión de las pruebas realizadas, cuando la cantidad de rezagos se determina en base al criterio `legacy` es decir, cantidad de lags = $12 \times (T/100)^{(1/4)}$, con T siendo la cantidad de períodos, y la hipótesis nula es de serie estacionaria alrededor de una tendencia, se tiene que 105 series son estacionarias y 17 no lo serían. En la siguiente Tabla 3.1 se presenta la cantidad de series estacionarias y no estacionarias para cada una de las pruebas ADF y KPSS bajo las formas antes mencionadas.

Bajo este resultado, para las 4 series que no son estacionarias según ADF, 3 no son estacionarias según KPSS por lo cual, para 3 de 4 series los resultados confirman que no son estacionarias. En total, 15 series presentan resultados

Tabla 3.1: Cantidad de series estacionarias y no estacionarias según pruebas ADF y KPSS

	Estacionarias	No estacionarias
ADF	118	4
KPSS	105	17

contradictorios según los resultados de las pruebas. Dado que se siguieron las recomendaciones de transformación de las series de McCracken y Ng (2016) para que las series sean estacionarias, se opta por descartar del estudio únicamente a las 3 series de datos para las cuales los resultados de no estacionariedad son confirmatorios con ambos tests: HOUSTMW (Housing Starts, Midwest), HOUSTS (Housing Starts, South) y PERMITMW (New Private Housing Permits, Midwest (SAAR)). Los resultados de los modelos menos generales coinciden para un nivel de confianza del 95 % en que para estas series habría evidencia de presencia de RU. Se podría profundizar en el análisis de las pruebas ADF con los estadísticos F asociados a cada modelo pero al no ser el centro de este estudio y porque se parte de la idea de que las series deberían ser estacionarias, se trunca el análisis de las pruebas de RU en esta etapa.

Ordenamiento y división del conjunto de datos en entrenamiento y prueba

Tras eliminar las series temporales no estacionarias se procede a ordenar los datos siguiendo el esquema recursivo (2.19) (ir a código). Luego, se divide a la muestra en conjuntos de entrenamiento y prueba. Los datos de entrenamiento, considerando al índice temporal ordenado de manera ascendente, abarcan el 80 % de la muestra y el restante 20 % de los datos son el conjunto de prueba. Como se está en un contexto de series temporales, el conjunto de entrenamiento constituye los datos más antiguos mientras que el conjunto de datos de prueba está compuesto por datos más recientes. De esta manera, el conjunto de entrenamiento está compuesto por 577 filas con un rango de fechas de 1960-01-01 a 2008-01-01 y 118 columnas¹. Mientras que el conjunto de prueba está compuesto por 145 filas empezando el 2008-02-01 hasta 2020-

¹Dado que el año que divide al conjunto de entrenamiento y prueba coincide con la crisis financiera de 2008, como indiqué antes, es esperable un cambio de régimen para las series del estudio.

02-01 y 118 columnas ([código](#)). Una vez obtenido el panel de datos, se escala cada una de las series univariadas de manera que cada una tenga media igual a 0 y desviación típica igual a 1. Como resultado, se obtiene un conjunto de datos escalado para emplear como entrada de los modelos FAVAR y DDFM a estimar en la próxima sección [3.2](#) (ir a [código](#)).

3.2. Inferencia

Estimación de la cantidad óptima de factores

Siguiendo a Stock y Watson (2016), para estimar la cantidad de factores estáticos r se puede emplear una combinación de herramientas: conocimiento a priori, gráficos como el scree plot y criterios de información, entre otros.

Scree Plot: Esta visualización muestra la contribución marginal del k -ésimo componente principal al promedio de R^2 de los N regresores de X_t contra los primeros k componentes principales¹. Se emplea el paquete `dfms` de Krantz y Bagdziunas (2023) que permite obtener las siguientes Figuras 3.1 y 3.2 (código). De esta manera, el scree plot de la Figura 3.1, grafica los valores propios ordenados de $\hat{\Sigma}_X = \frac{1}{T} \sum_{t=1}^T X_t X_t'$ hallados en (3.1).

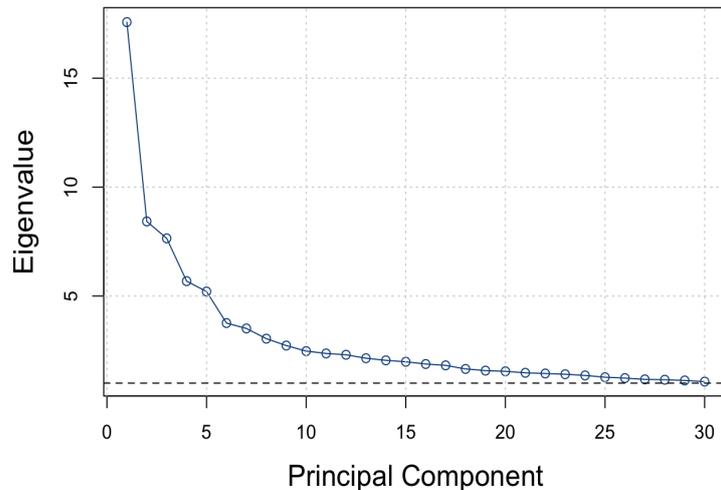


Figura 3.1: Scree plot de los valores propios ordenados

¹En este estudio, los regresores X_t se corresponden con la muestra de entrenamiento.

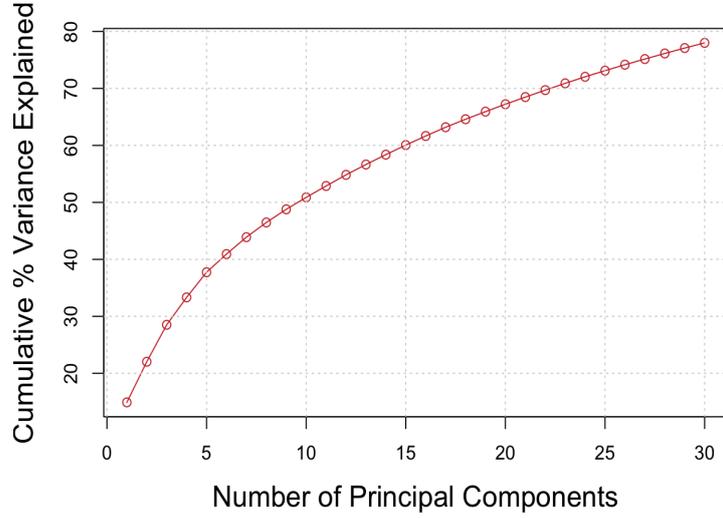


Figura 3.2: % de la varianza explicada acumulada

Gráficamente, para 5 factores se acumula casi el 40 % de la varianza explicada (Figura 3.2).

Criterios de información: Siguiendo a Stock y Watson (2016), la estimación de los factores con un enfoque de PC, resuelve el problema de Mínimos cuadrados ordinarios (MCO) para los cuales Λ y F son parámetros desconocidos a inferir

$$\min_{F_1, \dots, F_T, \Lambda} V_r(\Lambda, F) = \frac{1}{NT} \sum_{t=1}^T (X_t - \Lambda F_t)^T (X_t - \Lambda F_t) \quad (3.1)$$

sujeito a una normalización.

Bajo la normalización por PC, las columnas de Λ son ortogonales y se escalan para tener norma unitaria. La misma se expresa como $N^{-1}\Lambda\Lambda = I$ con Σ_F diagonal y $\Sigma_F = E(F_t^T F_t)$ o $F_t F_t^T / T = I$ cuando se restringen los factores. Además, los elementos de la diagonal de Σ_F son débilmente decrecientes.

Como no hay datos faltantes, la solución de mínimos cuadrados del problema (3.1) es el estimador de PC para los factores tal que $\hat{F}_t = \frac{1}{N} \hat{\Lambda}' X_t$, donde $\hat{\Lambda}$ es la matriz de vectores propios de la matriz de varianza muestral de X_t , $\hat{\Sigma}_X = \frac{1}{T} \sum_{t=1}^T X_t X_t'$, asociada a los r mayores valores propios de $\hat{\Sigma}_X$ (Stock y

Watson (2016, p.430)). En este marco, Bai y Ng (2002) extienden el criterio de información de *Akaike*: se penaliza a (3.1) compensando el beneficio de incluir parámetros adicionales contra el costo de incrementar la variabilidad de la muestra. Para esto, se incluye un factor adicional tal que

$$IC(r) = \ln V_r(\hat{\Lambda}, \hat{F}) + r \times g(N, T) \quad (3.2)$$

donde $V_r(\hat{\Lambda}, \hat{F})$ es la función objetivo de MCO en (3.1) evaluada en $(\hat{\Lambda}, \hat{F})$ y $g(N, T)$ es el factor de penalización que tiende a cero a medida que N y T tienden a infinito. Una función de penalización típica, por ser parsimoniosa, es la IC_{r2} para la cual $g(N, T) = [(N + T) / (NT)] \ln(\min(N, T))$. Entonces, se puede expresar IC_{r2} como

$$IC_{r2} = \ln V_r(\hat{\Lambda}, \hat{F}) + r \times \frac{N + T}{NT} \ln(\min(N, T)).$$

Con el paquete `dfms` de Krantz y Bagdziunas (2023) se pueden comparar tres Criterios de información (ICr)¹. La cantidad óptima de factores r^* se determina por el mínimo ICr (ver Anexo 2). Los tres ICr son equivalentes asintóticamente, aunque en muestras finitas puedan diferir: la penalidad de IC_{r2} es la mayor para muestras finitas. La Figura 3.3 muestra distintas cantidades óptimas de factores según el tipo de penalización del ICr en (2.1).

¹Por sus siglas en inglés relativas a Information criteria.

Optimal Number of Factors (r) from Bai and Ng (2002) Criteria

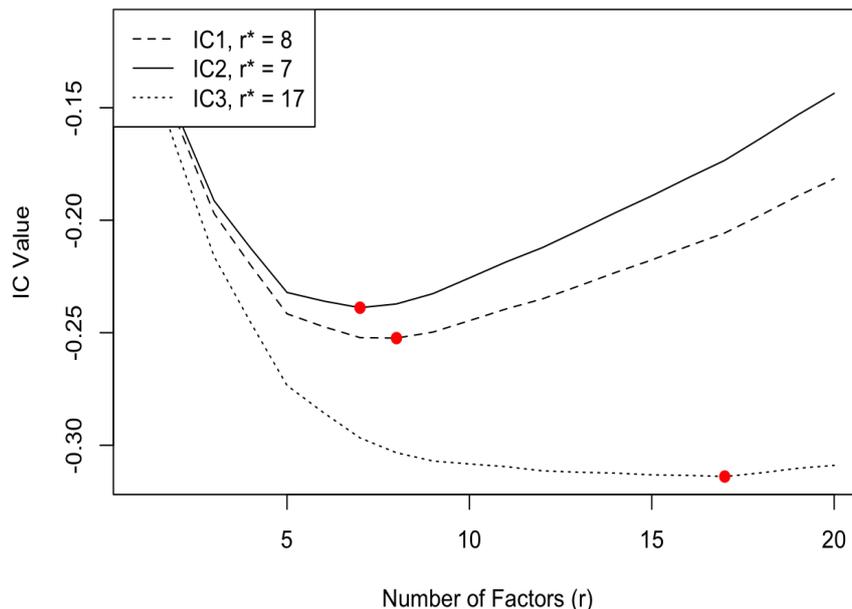


Figura 3.3: Criterios de información

Se observa que $IC_{r2} = 7$, sugiriendo que el número óptimo de factores es este mismo. Además, la Figura 3.2 muestra que con 5 factores se logra explicar cerca del 40% de la varianza, lo que motiva la estimación de modelos para distintas cantidades de factores y luego, comparar sus resultados.

Estimación del modelo FAVAR lineal

Como se está en el contexto de un DFM estático (2.13) exacto¹, la estimación del FAVAR se puede hacer mediante PC. Este tipo de estimación no paramétrica se basa en la idea de que cuando la cantidad de variables X_t es lo suficientemente grande, sólo con la variación que se da entre las mismas ya es posible estimar el espacio generado por los factores (Stock y Watson (2016, p. 429)). Entonces, para estimar el modelo FAVAR lineal se adopta el enfoque de PC en dos etapas empleado en Bernanke et al. (2005), que es análogo al de Stock y Watson (2002). Con este método no paramétrico, se busca estimar el espacio generado por los componentes comunes $C(F_t, Y_t)$ en (2.18).

¹Bajo un DFM exacto se supone que las varianzas idiosincrásicas son homogéneas y los factores son tratados como parámetros a estimar (Stock y Watson (2016)).

- Paso 1:** Se infiere $\hat{C}(F_t, Y_t)$ con los primeros $r + M$ componentes principales de X_t . En este paso no se explota el hecho de que Y_t sea una variable observada. Sin embargo, cuando N es grande y la cantidad de PC es al menos tan grande como la cantidad de factores verdaderos, esta estimación recupera consistentemente el espacio generado por F_t y Y_t (Stock y Watson (2002)). Como $\hat{C}(F_t, Y_t)$ es una combinación lineal arbitraria de sus argumentos, estimar F_t implica determinar la parte de $\hat{C}(F_t, Y_t)$ que no es generada por Y_t . En este sentido, como no se impone a Y_t explícitamente como un componente común, cualquier combinación lineal subyacente de $\hat{C}(F_t, Y_t)$ podría involucrar a la FFR (R_t), que está incluida en Y_t . Entonces, no se puede estimar un VAR en $\hat{C}(F_t, Y_t)$ y Y_t , y a la vez identificar el shock de PM recursivamente. Como consecuencia, es necesario remover la dependencia directa de $\hat{C}(F_t, Y_t)$ sobre R_t . Para lograr lo anterior, se estima la siguiente regresión

$$\hat{C}(F_t, Y_t) = b_{C^*} \hat{C}^*(F_t) + b_R R_t + e_t, \quad (3.3)$$

donde \hat{C}^* es una estimación de los componentes comunes distintos a R_t . Para obtener $\hat{C}^*(F_t)$ se sustraen los PC asociados a las variables lentas. Con (3.3) se obtiene

$$\hat{F}_t = \hat{C}(F_t, Y_t) - \hat{b}_R R_t. \quad (3.4)$$

Para asegurarse que la normalización sea una de las propuestas en Bernanke et al. (2005) y Stock y Watson (2016), se creó la función `pca_norm` para estimar por PC (código). De este modo, la normalización en este estudio es $C' C / N = I$ con $C' = [C(F_1, Y_1), \dots, C(F_T, Y_T)]$. A diferencia de Bernanke et al. (2005) que toma la normalización $C' C / T = I$, lo que implica que $\hat{C} = \sqrt{T} \hat{Z}$ donde \hat{Z} son los vectores propios vinculados a los r mayores valores propios de $X' X$.

- Paso 2:** Se estima un VAR con métodos estándares, reemplazando a F_t por su estimación \hat{F}_t en (3.4).

El código para implementar lo anterior, aplica y extiende el de Duarte (2020). En particular, para estimar el VAR en \hat{F}_t y Y_t se empleó el paquete `vars` en R propuesto por Pfaff (2008), considerando un bucle para distintos valores

de los factores de 5 a 7. Lo anterior se vincula con los resultados encontrados en la subsección 3.2, por lo que se considera que la cantidad óptima de factores podría estar entre estos valores.

La estimación mediante el método de PC presenta la ventaja de ser un procedimiento sencillo y fácil de implementar, ya que no requiere de supuestos estrictos sobre las distribuciones y tolera cierto nivel de correlación cruzada en el término e_t . No obstante, una desventaja significativa de este método es que conlleva la generación de regresores en la segunda etapa del proceso de estimación (Bernanke et al. (2005)).

Estimación del modelo DDFM

El método de estimación del modelo DDFM aplicado en este estudio, sigue al desarrollado por Andreini et al. (2023a)¹. Los autores indican que si bien la estimación de los parámetros de un modelo bajo la forma $X_t = F(\mathbf{f}_t) + \mathbf{e}_t$ podría realizarse mediante el método máxima verosimilitud, en la práctica, obtener las estimaciones de los parámetros, incluso para modelos lineales, es poco factible. Dado el esquema no lineal del modelo, los parámetros se infieren con un método de gradientes Monte Carlo porque tendría ventajas computacionales frente a otros métodos, sin dejar de lado la interpretación dinámica del mismo².

Entrenamiento del modelo

De modo general, en la fase de entrenamiento, se emplea el método denominado Back-propagation (Rumelhart et al. (1986)). Consiste en estimar, mediante actualizaciones repetidas, el vector de parámetros $\boldsymbol{\theta}$ que contiene el conjunto completo de pesos y sesgos, para obtener una función de costo $J(\boldsymbol{\theta}) = \mathcal{L}(\mathbf{x}_t, \tilde{\mathbf{x}}_t, \boldsymbol{\theta}) = \mathcal{L}(\mathbf{x}_t, F(G(\mathbf{x}_t)); \boldsymbol{\theta})$ mínima. Este método de optimización consiste en dos pases: en el pase hacia adelante, la entrada \mathbf{x} brinda la información inicial a propagar a lo largo de la red para obtener un resultado $\hat{\mathbf{x}}$. Durante el entrenamiento, la propagación hacia adelante puede continuar hasta obtener un costo escalar $J(\boldsymbol{\theta})$; el pase hacia atrás permite que la información fluya desde el costo hacia atrás para calcular el gradiente de la función de costo respecto a los parámetros $\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta})$, lo que permite actualizar los pesos de la red. Entendiendo a la red neuronal como un grafo computacional, Back-propagation es un algoritmo que emplea la regla de la cadena para efectuar operaciones en un orden específico, es decir, para hallar las derivadas de las funciones compuestas por otras funciones, cuyas derivadas son conocidas (ver Anexo 3).

En particular, para este trabajo, se sigue el algoritmo (1) desarrollado por Andreini et al. (2023b). Es un procedimiento en dos pasos para entrenar el modelo y obtener estimaciones. Primero, se estiman todos los parámetros del modelo. Segundo, se usan los resultados del encoding en el decoding, en un es-

¹Que se basa y extiende el propuesto por Bengio et al. (2013), denominado Generalised denoising autoencoders.

²Además, los resultados de la estimación se pueden pensar como una aproximación de un máximo a priori (Andreini et al. (2023a)).

quema de espacios de estados para hacer actualizaciones de los estados latentes (código).

Algorithm 1 MCMC para *DDFM* con componentes idiosincráticos estacionarios $AR(d)$, una estructura de *encoding* $G_{\theta_G}(\cdot)$ y de *decoding* $F_{\theta_F}(\cdot)$

Init: $\theta_G, \theta_F, \Phi, \Sigma_e, \mathbf{e}_t$
repeat
 1: $\tilde{\mathbf{x}}_t | (\mathbf{x}_t, \hat{\mathbf{e}}_t) = \mathbf{x}_t - \Phi(L)\mathbf{e}_t$
 2: **Loop** epochs, batches **Do**
 3: draw $\mathbf{e}_t^{(mc)} \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_e)$
 4: $\mathbf{x}_t^{(mc)} = \tilde{\mathbf{x}}_t | (\mathbf{x}_t, \hat{\mathbf{e}}_t) + \mathbf{e}_t^{(mc)}$
 5: θ_G, θ_F update by gradient based step on $\hat{\mathcal{L}}(\mathbf{x}_t, F_{\theta_F}(G_{\theta_G}(\mathbf{x}_t^{(mc)})))$
 6: **End Loop**
 7: $\mathbf{f}_t | \mathbf{x}_t^{(mc)} = \mathbb{E}_{\mathbf{x}_t^{(mc)} \sim \mathbf{x}_t, \hat{\mathbf{e}}_t} G_{\theta_G}(\mathbf{x}_t^{(mc)})$
 8: $\mathbf{e}_t | \mathbf{x}_t, \mathbf{f}_t = \mathbf{x}_t - F_{\theta_F}(\mathbf{f}_t | \mathbf{x}_t^{(mc)})$
 9: $\Phi \leftarrow$ stationary $AR(d)$ on \mathbf{e}_t
 10: $\Sigma_e \leftarrow$ from \mathbf{e}_t
until convergence on $\hat{\mathcal{L}}(\mathbf{x}_t, F_{\theta_F}(\mathbf{f}_t | \mathbf{x}_t^{(mc)}))$ in L_1 norm
return $\Sigma_e, \Phi, \mathbf{f}_t, F_{\theta_F}$

Paso 1: Se implementa el algoritmo (1): consiste en una Cadena de Markov Monte Carlo (MCMC)¹ y asume $AR(d)$ para los componentes idiosincráticos con una estructura de encoding $G_{\theta_G}(\cdot)$ y una estructura de decoding $F_{\theta_F}(\cdot)$. El paso 1 está compuesto, a su vez, por una secuencia de etapas (líneas):

- Se comienza inicializando los parámetros del modelo $(\theta_G, \theta_F, \Sigma_e, \Phi, \mathbf{e}_t)$.
- Se repite el algoritmo hasta alcanzar convergencia según el criterio de parada:

línea 1: Se filtran los datos de entrada \mathbf{x}_t con la media condicional de los componentes idiosincráticos $AR(d)$. Como resultado, se obtiene $\tilde{\mathbf{x}}_t$, o sea, se capturan los estados latentes.

líneas 2 a 6: Se lleva a cabo el paso Monte Carlo y la actualización de los gradientes para cada época y lote². Para esto se emplea $\tilde{\mathbf{x}}_t$ y se

¹Por sus siglas en inglés relativas a Markov Chain Monte Carlo. Estos métodos permiten aproximar integrales de funciones continuas, devolviendo su valor esperado empírico mediante el muestreo de una distribución de probabilidad de entrenamiento. En particular, asumiendo un espacio de cadenas de Markov se supone que se parte de un estado x a partir de un valor arbitrario (Goodfellow et al. (2016, cap.17)).

²Ambos son hiperparámetros del modelo, una época significa una pasada completa del conjunto de datos de entrenamiento por el algoritmo. Mientras que el tamaño del lote define el número de muestras con las que se va a trabajar antes de actualizar los parámetros internos del modelo (Andreini et al. (2023a)).

inyecta ruido gaussiano a partir de e_t , bajo la forma de denoising, para obtener unas observaciones ruidosas $\mathbf{x}_t^{(mc)}$ (Bengio et al. (2013)).

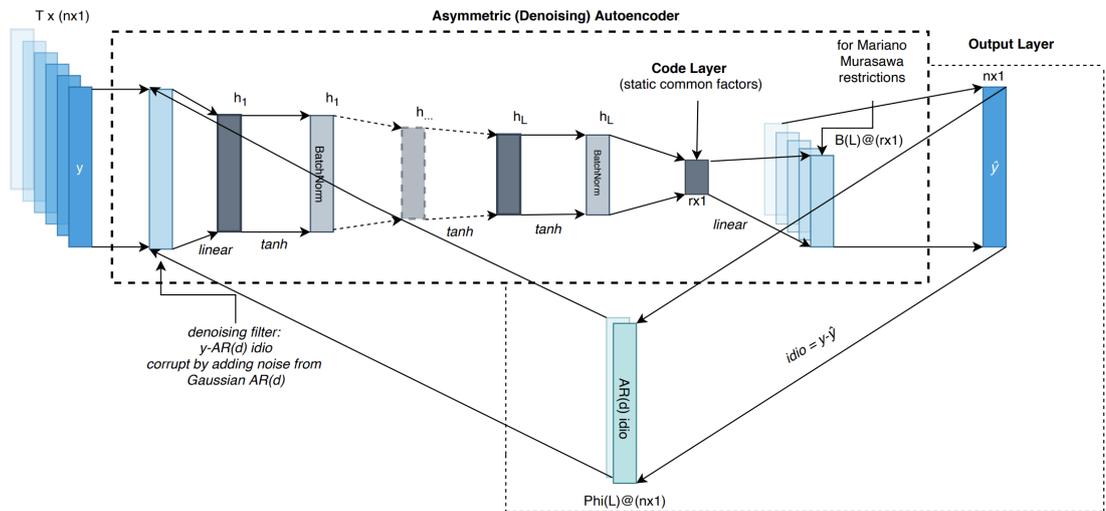
línea 7: Se extraen los estados latentes \mathbf{f}_t de la red de encoding con una aproximación Monte Carlo.

líneas 8 a 10: Se actualizan los parámetros del modelo del componente idiosincrático e_t condicional a los factores y a las variables observables.

- **Criterio de parada:** Se adopta una función de pérdida MSE dada por $\hat{\mathcal{L}}(\mathbf{x}_t, F_{\theta_F}(\mathbf{f}_t | \mathbf{x}_t^{(mc)}))$. Se alcanza convergencia cuando $\|MSE_k - MSE_{k-1}\|_1 < \delta = 0.0005$, con k siendo la iteración.
- Se obtienen las estimaciones de $F_{\theta_F}, \Sigma_e, \Phi, \mathbf{f}_t$.

Paso 2: Se lleva el resultado del algoritmo a las ecuaciones de espacio de estados (2.29c)-(2.29d). Las actualizaciones de los estados en presencia de un decodificador lineal se realizan con un filtro de Kalman (código).

En la siguiente Figura 3.4 se describe el proceso de entrenamiento¹.



Fuente: Andreini et al. (2023a, p.16)

Figura 3.4: Proceso de entrenamiento del modelo DDFM

¹La expresión *idio* hace referencia a los componentes idiosincráticos.

Elección de los hiperparámetros del modelo

Siguiendo a Goodfellow et al. (2016, cap.11), los hiperparámetros son todos aquellos parámetros de un modelo que se determinan por fuera del proceso de entrenamiento. En general, los algoritmos de aprendizaje profundo precisan especificar un conjunto de hiperparámetros para controlar ciertos aspectos del comportamiento del algoritmo. Esto implica que pueden afectar el tiempo y el costo de memoria en la ejecución del algoritmo, la calidad del modelo recuperado por el proceso de entrenamiento y su capacidad para ofrecer resultados correctos cuando se implementa en nuevas entradas.

A grandes rasgos, los hiperparámetros se pueden clasificar en dos categorías. La primera, se relaciona con la estructura de la red neuronal e incluye al tipo de capas ocultas, a la cantidad de neuronas por cada capa oculta, a los coeficientes de penalización, a las capas de dropout, a las capas de batch normalization, a la elección de las funciones de activación, entre otros. La segunda categoría, se vincula con el algoritmo de optimización: el tamaño de los minibatches (Apéndice 3), la cantidad de épocas, la tasa de aprendizaje, el método de optimización, entre otros. Además, los modelos típicos de factores para series temporales incluyen otros hiperparámetros: la cantidad de estados latentes comunes, de rezagos en las variables de entrada, de rezagos en los estados latentes comunes y del componente idiosincrático¹ (Andreini et al. (2023a, p.19-20)).

La elección de los hiperparámetros puede ser automática o manual. Cuando la elección es manual es necesario saber qué hace cada hiperparámetro y cómo el algoritmo de ML alcanza un buen desempeño, es decir, hay que entender la relación entre hiperparámetros, error de entrenamiento, error de generalización y recursos computacionales². El objetivo de la búsqueda manual, consiste en encontrar el menor error de generalización sujeto a un presupuesto de tiempo de ejecución y memoria. Por lo tanto, se busca ajustar la capacidad efectiva del modelo a la complejidad de la tarea a realizar³. En ausencia de estas

¹En la literatura asociada a series temporales, este tipo de hiperparámetro o bien se fija a priori o se emplea algún criterio de información en vez de emplear algoritmos de grid search.

²Esto es, memoria y tiempo de ejecución.

³La capacidad efectiva está limitada por tres factores: la capacidad de representación del modelo, la capacidad del algoritmo de aprendizaje para minimizar con éxito la función de costo utilizada para entrenar el modelo y el grado en que la función de costo y el proceso de entrenamiento regularizan el modelo.

referencias, los algoritmos automatizados pueden encontrar valores útiles de hiperparámetros al optimizar una función objetivo mediante la exploración de diversas opciones (Goodfellow et al. (2016, cap.11)).

En este trabajo, se va a emplear el método de grid search (búsqueda por grilla) para determinar algunos hiperparámetros del modelo. Este método consiste en seleccionar una cantidad finita de valores a investigar. Luego, se entrena el modelo para cada combinación de hiperparámetros y el experimento que presente menor valor de la función de pérdida determinará el conjunto de hiperparámetros elegido. Para este método, la elección de los hiperparámetros se realiza, en general, con respecto a una pérdida de validación, que se estima mediante un proceso llamado Validación cruzada (VC), detallado en el Apéndice 3. Sin embargo, como se indica en Andreini et al. (2023a), este método se asocia a un costo computacional creciente a medida que aumenta la cantidad de hiperparámetros. Si hay m hiperparámetros, y cada uno toma n valores, entonces la cantidad de entrenamientos y pruebas de evaluación crece con $O(n^m)$.

Tomando en cuenta lo anterior, sigo estos pasos:

- a) Elijo un punto de partida inicial para fijar una grilla de hiperparámetros basándome en la experiencia de Andreini et al. (2023a) y Klieber (2024). Además, tomando como referencia el resultado de la subsección 3.2 anterior, considero 5, 6 y 7 factores latentes como hiperparámetros.
- b) Ajusto el modelo DDFM desarrollado en Keras¹ por Andreini et al. (2023a), para las distintas combinaciones de hiperparámetros.
- c) Selecciono aquellas combinaciones de hiperparámetros para las cuales, del resultado de entrenar el DDFM, se obtiene el mínimo MSE y comparo resultados.

En total, puse a prueba $3 \times 2 \times 2 \times 2 \times 2 \times 2 = 96$ combinaciones de hiperparámetros: cantidad de factores (3), estructura del encoder (2), si adicionar un sesgo en la última capa de decodificación o no (2), el tipo de función de activación a emplear (2), el número de rezagos de las entradas en el codificador (2) y el tamaño del minibatch (2). Con esto, se ajusta para cada combinación un modelo DDFM y se ejecuta el método de VC para un solo pliegue² ($K = 1$).

En la Tabla 3.2 se detallan las elecciones iniciales de hiperparámetros. Co-

¹Es una API de alto nivel de la plataforma TensorFlow (Martín Abadi et al. (2015)).

²Por cuestiones de tiempo computacional y de la complejidad para realizar la tarea para $K > 1$. En estudios posteriores sería interesante realizar VC con más pliegues.

mo la cantidad neuronas en cada capa oculta no está predefinida se sigue el criterio de realizar una reducción progresiva de las dimensiones en cada capa del encoder, manteniendo un factor de escala constante r hasta llegar a la capa de código (Klieber (2024)). De esta forma, las opciones para la estructura del encoder son: $(r \times 12, r \times 6, r \times 4, r)$, $(r \times 24, r \times 12, r \times 8, r)$. En relación a las funciones de activación, en Andreini et al. (2023a) se emplea la función \tanh mientras que en Klieber (2024) la función $ReLU$. Las opciones iniciales del tamaño del minibatch se vinculan con la duración en meses del ciclo económico: en Andreini et al. (2023a), se establece un tamaño igual a 100 indicando que un ciclo económico dura 8 años y por lo tanto, aproximadamente 100 meses; en Klieber (2024), se señala que para Estados Unidos es de 6 años, por lo tanto, estaría en el entorno de los 72 meses.

Tabla 3.2: Elección inicial de hiperparámetros

Hiperparámetros	Opciones
Q de neuronas en cada capa	Reducción progresiva de las dimensiones
Función de activación	\tanh , $ReLU$
Sesgo en la capa de decoder	True, False
Tamaño de los minibatches	100, 72
Q de factores latentes	5, 6, 7
Q de rezagos en las variables de entrada	0, 1

Nota: Q refiere a Cantidad

Además, como en Andreini et al. (2023a), en el entrenamiento de los modelos DDFM se incluye lo siguiente:

- En la parte de encoding, dos capas de normalización por lotes (Batch-Norm) para que haya cierta regularización y control sobre el potencial Desplazamiento de covariación interna (ICS)¹. El ICS se refiere al cambio en la distribución de las entradas de la capa causado por actualizaciones en las capas precedentes. Se dice que dicho cambio continuo impacta negativamente en el entrenamiento. Por este motivo, el objetivo de Batch-Norm es reducir el potencial ICS y así remediar el mencionado efecto (Santurkar et al. (2019)).
- El método de optimización Estimación adaptativa de momentos (Adam) para calcular las tasas de aprendizaje adaptativas de cada parámetro (Apéndice 3).

¹En inglés se denomina Internal covariate shift.

- Cien épocas tanto en el preentrenamiento como en el entrenamiento del modelo.
- Se inicializan los parámetros con un enfoque en dos pasos. Primero, se emplea la inicialización de Xavier (Anexo 3) y luego se preentrena el modelo para hacer *entrar en calor* a la cadena (como en Andreini et al. (2023a)).
- Hiperparámetros relacionados con la literatura de series temporales: la cantidad de estados comunes latentes, la cantidad de rezagos de las variables de entrada, la cantidad de rezagos de los estados latentes comunes y del componente idiosincrático. Salvo por la cantidad de rezagos de las variables de entrada que se selecciona por grid search (ver Tabla 3.2), el resto se fija a priori siguiendo a Bańbura y Modugno (2014), Giannone et al. (2010) y Andreini et al. (2023a, p.20).

En la Tabla 3.3 se presenta la elección final de hiperparámetros considerando el valor mínimo del MSE que surge de entrenar el modelo DDFM para cada combinación de los mismos con $r = 5, 6, 7$ factores (código).

Tabla 3.3: Elección final de hiperparámetros

Cantidad de factores	5	6	7
Estructura encoder	120, 60, 40	144, 72, 48	168, 84, 56
Rezagos de las entradas	1	1	1
Sesgo en decoder	FALSE	TRUE	TRUE
Función de activación	tanh	tanh	tanh
Tamaño minibatch	100	100	100
MSE	0.62	0.59	0.56

Estimación del modelo FAVAR profundo

A partir de la inferencia de los modelos DDFM empleando los argumentos de la Tabla 3.3, se construyen tres modelos FAVAR profundos. Los mismos incorporan los factores latentes estimados $\hat{\mathbf{f}}_{r=5,6,7}$ junto a un factor observable \mathbf{y}_t para estimar tres modelos VAR (como en Klieber (2024)). En este contexto de factores no lineales, para poder realizar el análisis de respuesta al impulso del capítulo 4, sección 4.3, es necesario que exista un mapeo entre los factores \mathbf{f}_t y las variables \mathbf{x}_t . Entonces, como los factores estimados son no lineales, para poder medir el efecto de los mismos sobre las variables informacionales, se calcula la inversa generalizada de Moore-Penrose de la librería MASS de R (Venables y Ripley (2002)). Con esto se obtienen unos coeficientes linearizados tal que

$$\hat{\boldsymbol{\beta}} = Proj(\mathbf{f}, \mathbf{x}) = \mathbf{f}^+ \mathbf{x}$$

donde \mathbf{f}^+ es la pseudoinversa de Moore-Penrose de \mathbf{f}_t (código). La pseudoinversa de Moore-Penrose va a permitir la existencia de un vínculo entre los factores no lineales y las variables informacionales aún cuando \mathbf{f}^{-1} no exista. Además, cuando \mathbf{f} sea de rango completo $\hat{\boldsymbol{\beta}}$ van a ser los estimadores MCO de $\boldsymbol{\beta}$ (Crawford et al. (2018, 2019), Theodoridis (2015)). Por consiguiente, como en la sección 3.2, se estima un VAR en $\hat{\mathbf{f}}_t^+$ y \mathbf{y}_t , se siguen los mismos pasos para determinar la cantidad de rezagos del modelo, y se obtienen las predicciones de las series univariadas 145 pasos hacia adelante.

Capítulo 4

Resultados empíricos

4.1. Desempeño de los modelos estimados

Medidas de error de generalización: Como en Andreini et al. (2023a) y en Klieber (2024), se emplea el Error de predicción cuadrático medio (RMSFE)¹ como medida de desempeño predictivo de referencia. Se calcula además, el Error cuadrático medio (MSE)² y el Error absoluto medio (MAE)³. Entonces, para un tamaño del conjunto de prueba igual a $T = 145$ y para cada variable i con $i = 1, \dots, N$, se tienen las siguientes métricas

- $MAE_i = \frac{1}{T} \sum_{t=1}^T |x_{i,t} - \hat{x}_{i,t}|$
- $MSE_i = \frac{1}{T} \sum_{t=1}^T (x_{i,t} - \hat{x}_{i,t})^2$
- $RMSFE_i = \sqrt{\frac{1}{T} \sum_{t=1}^T (x_{i,t} - \hat{x}_{i,t})^2}$

La evaluación de los modelos se realiza tomando los promedios de cada métrica

$$\begin{aligned} RMSFE &= \frac{1}{N} \sum_{i=1}^N RMSFE_i \\ MSE &= \frac{1}{N} \sum_{i=1}^N MSE_i \\ MAE &= \frac{1}{N} \sum_{i=1}^N MAE_i \end{aligned}$$

En la evaluación del rendimiento predictivo para los modelos FAVAR lineal, DDFM y FAVAR profundo se comparan los resultados de las medidas de error de generalización promedio para el caso de 5, 6 y 7 factores latentes estimados. En las Tablas 4.1, 4.2 y 4.3 se observan los resultados.

¹Por sus siglas en inglés relativas a Root Mean Squared Forecast Error.

²Por sus siglas en inglés relativas a Mean Squared Error.

³Por sus siglas en inglés relativas a Mean Absolute Error.

FAVAR lineal:

Tabla 4.1: MAE, MSE y RMSFE promedio para el FAVAR lineal con 5, 6 y 7 factores latentes estimados (\hat{F}_t) y 1 observado (Y_t)

\hat{F}_t	Y_t	MAE	MSE	RMSFE
7	1	0.684	0.956	0.977
6	1	0.688	0.967	0.982
5	1	0.685	0.960	0.979

DDFM:

Tabla 4.2: MAE, MSE y RMSFE promedio para el DDFM para 5, 6 y 7 factores latentes estimados (\hat{F}_t)

\hat{F}_t	MAE	MSE	RMSFE
7	0.684	0.956	0.978
6	0.684	0.958	0.979
5	0.686	0.964	0.982

FAVAR profundo:

Tabla 4.3: MAE, MSE y RMSFE promedio para el FAVAR profundo con 5, 6 y 7 factores latentes estimados (\hat{F}_t) y 1 observado (Y_t)

\hat{F}_t	Y_t	MAE	MSE	RMSFE
7	1	0.678	0.941	0.969
6	1	0.686	0.958	0.978
5	1	0.685	0.961	0.979

Para los 3 modelos comparados, los resultados son similares. Cabe resaltar que el modelo *DDFM* no considera factores observables en sus estimaciones y predicciones a diferencia de los FAVAR, que incorporan un factor observable (Y_t) además de los latentes. Si bien no hay diferencias significativas a la hora de generalizar los modelos en nuevas entradas, los resultados para los modelos FAVAR coinciden en arrojar un menor MSE para el caso de 7 factores inobservables estimados (\hat{F}_t). Por lo tanto, en el análisis de las próximas secciones 4.2 y 4.3 se toman los resultados de los modelos FAVAR lineal y profundo para 7 factores latentes.

4.2. Importancia de las variables y correlación entre factores

A continuación, se describen las técnicas empleadas para inferir la importancia de las variables informacionales sobre los factores estimados mediante el FAVAR lineal y profundo. En una segunda instancia se grafican los resultados y por último, se realiza un análisis de los mismos.

Importancia de las variables sobre los factores lineales: Para calcular la importancia de las variables para el caso de los factores lineales, se toma el valor absoluto de los pesos de la matriz $\hat{\Lambda}^f$, se los ordena de mayor a menor (Despois y Doz (2023), Klieber (2024)). En este estudio se seleccionan y visualizan las primeras 15 variables con mayores pesos absolutos para cada uno de los factores (ver Figuras 4.3, 4.6, 4.9 para los primeros 4 factores, los otros casos se presentan en el Apéndice 3).

Importancia de las variables sobre los factores no lineales: Siguiendo a Klieber (2024), en la literatura se han criticado las técnicas de ML y DL por la dificultad a la hora de interpretar sus resultados. En respuesta a esto, una estrategia para evaluar la importancia de las variables sobre los factores, radica en la aplicación de una metodología basada en los valores de Shapley (Shapley (1953), Strumbelj y Kononenko (2010), Lundberg y Lee (2017)). A grandes rasgos, los valores de Shapley son una herramienta vinculada a la teoría de juegos cooperativos que asigna un valor a cada jugador en función de su contribución a un resultado global. La misma se ha convertido en una técnica para interpretar los modelos de ML. En particular, la literatura reciente vinculada a la macroeconomía y a las finanzas toma este tipo de medidas para poder interpretar los resultados de técnicas no lineales (Bluwstein et al. (2021), Borup et al. (2022), Coulombe y Goebel (2024)). En este contexto, se busca identificar a las fuerzas principales asociadas a los factores estimados y darles un sentido económico. Se define el valor de Shapley de cada variable n en el momento t como $\phi_{t,n}$, para $n = 1, \dots, N$ y $t = 1, \dots, T$. Cada factor estimado $\hat{F}_{t,r}$ con $r = 1, \dots, R$ se puede descomponer en

$$\hat{F}_{t,r} = \sum_{n=1}^N \phi_{t,n} + \phi_0$$

donde ϕ_0 es el valor medio estimado en el conjunto de entrenamiento, o sea el intercepto. El cálculo del valor Shapley de cada variable n implica derivar su contribución marginal a la predicción del factor. Lo anterior se realiza comparando las retribuciones (payoffs en inglés) de todas las posibles cooperaciones de variables¹ de los regresores, dadas por

$$\phi_{t,n} = \sum_{S \subseteq N/n} \frac{S!(N-S-1)!}{N!} (f(S \cup \{n\}) - f(S))$$

donde el payoff de una cooperación $S \subseteq N$ se expresa como $f(S)$ y el payoff de la misma cooperación combinado con el regresor n es $f(S \cup \{n\})$. La diferencia entre estas dos medidas arroja como resultado la contribución marginal de la variable n a la cooperación. Dado que en grandes volúmenes de datos, el costo computacional de calcular estos valores es creciente y considerable, se emplea un algoritmo híbrido como el planteado en Covert y Lee (2021). En este estudio, se usa la librería `kernelshap` de R para calcular la importancia de las variables en los factores (código). Kernel SHAP es un método que utiliza una regresión lineal ponderada particular para calcular la importancia de cada variable. Los valores de importancia calculados son los valores de Shapley de la teoría de juegos cooperativos y son los coeficientes de una regresión lineal local (Mayer y Watson (2023)). En las Figuras 4.4, 4.7, 4.10 se visualizan las primeras 15 variables más influyentes sobre factores estimados no lineales.

Máxima correlación absoluta entre factores estimados: Siguiendo a Klieber (2024), primero los factores obtenidos por los dos modelos FAVAR se escalan para que cada columna tenga media igual a cero y desviación típica igual a uno. Además, para lograr la comparación de los dos tipos de factores, los mismos se normalizan en el recorrido $[-1, 1]$ y se calcula la matriz de correlaciones en valores absolutos entre factores lineales (`fi_l`) y profundos (`fi_p`) con $r = 1, \dots, 7$, como se muestra en la Figura 4.1 (código).

¹Una cooperación es un subconjunto de variables S del conjunto N . Se dice que N es la gran cooperación de todos los jugadores (Strumbelj y Kononenko (2010)).

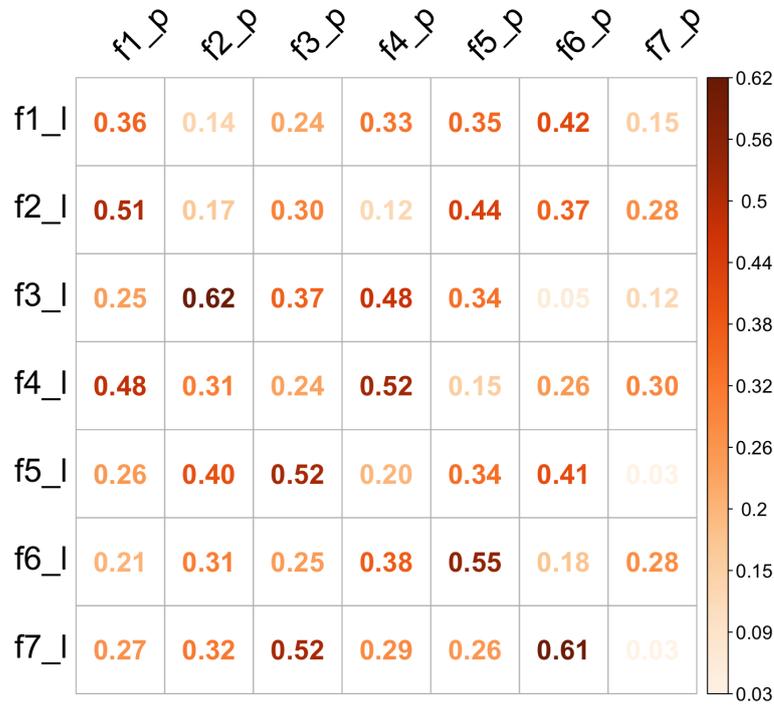


Figura 4.1: Correlación absoluta entre factores estimados

En las próximas Figuras 4.2, 4.5, 4.8 y 4.11 se visualizan los primeros 4 factores lineales de acuerdo a su correlación máxima absoluta con cada factor profundo (como en Klieber (2024)). Además, en las Figuras 4.3, 4.4, 4.6, 4.7, 4.9, 4.10, 4.12, 4.13 se grafican las 15 variables que más influyen en los factores estimados, categorizadas en los siguientes grupos de regresores (Tabla 4.4)¹.

Tabla 4.4: Grupos de variables por color

Color	Grupo de variables
Blue	Producto e ingresos
Orange	Consumo, órdenes e inventarios
Yellow	Mercado de trabajo
Red	Vivienda
Purple	Dinero y crédito
Grey	Mercado de activos
Pink	Tasas de interés y tipos de cambio
Cyan	Precios

¹El resto de los gráficos se hallan en el Apéndice 3.

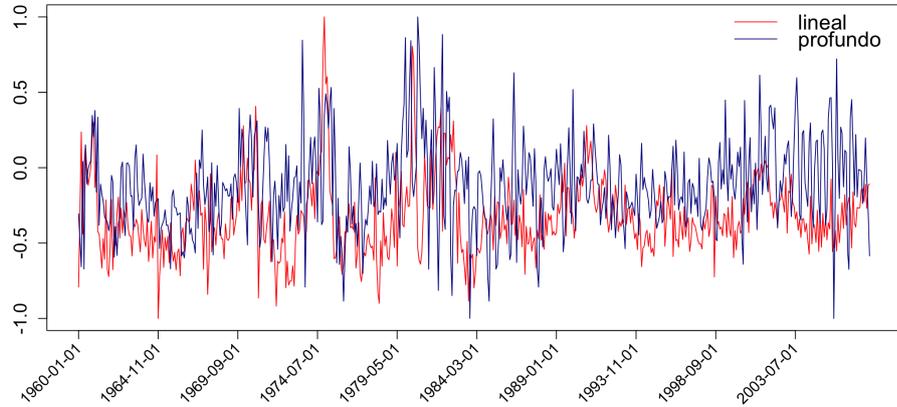


Figura 4.2: El primer factor lineal \hat{F}_1^L (línea roja) se correlaciona con el sexto factor profundo \hat{F}_6^P (línea azul)

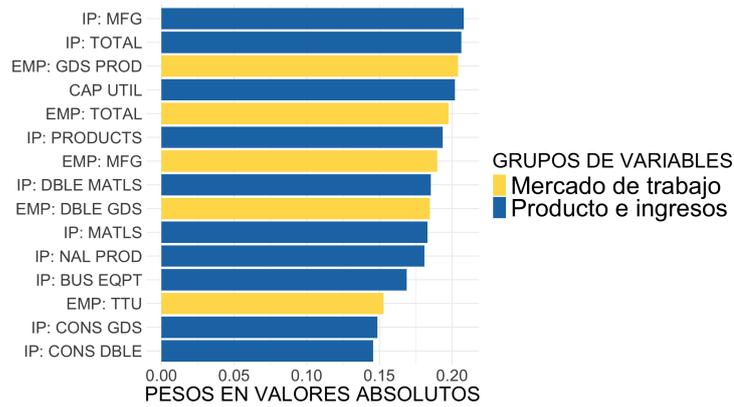


Figura 4.3: Primeras 15 variables que contribuyen a \hat{F}_1^L



Figura 4.4: Primeras 15 variables que contribuyen a \hat{F}_6^P

Nota: Descripción de las variables en Apéndice 3, Tablas 3.1 y 3.13. Los grupos de variables por color se detallan en la Tabla 4.4.

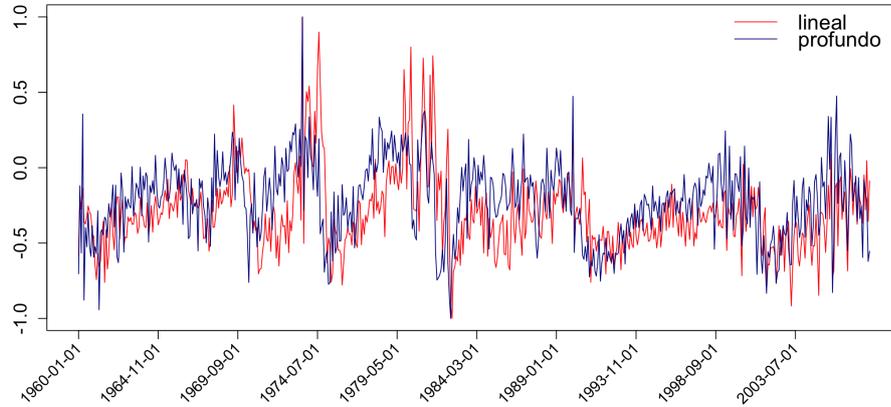


Figura 4.5: El segundo factor lineal \hat{F}_2^L (línea roja) se correlaciona con el primer profundo \hat{F}_1^P (línea azul)



Figura 4.6: Primeras 15 variables que contribuyen a \hat{F}_2^L

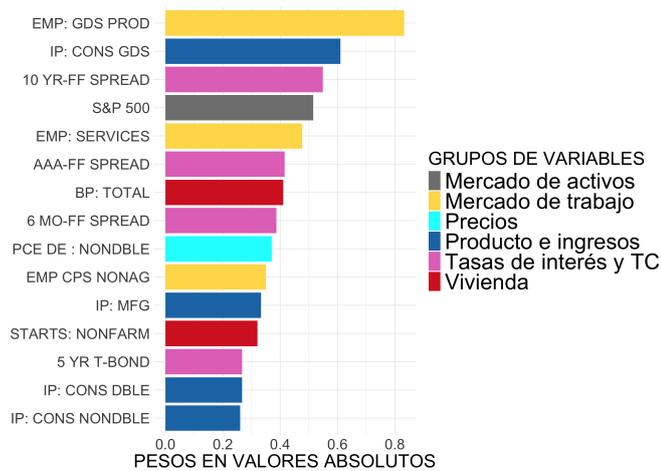


Figura 4.7: Primeras 15 variables que contribuyen a \hat{F}_1^P

Nota: Descripción de las variables en Apéndice 3, Tablas 3.2 y 3.8. Los grupos de variables por color se detallan en la Tabla 4.4.

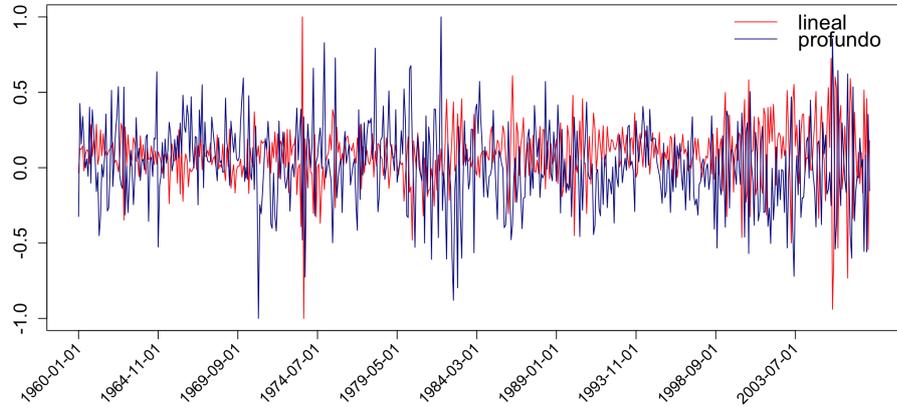


Figura 4.8: El tercer factor lineal \hat{F}_3^L (línea roja) se correlaciona con el segundo factor profundo \hat{F}_2^P (línea azul)

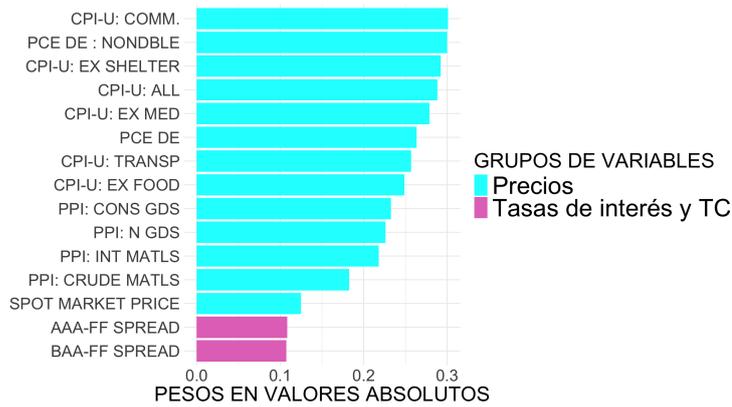


Figura 4.9: Primeras 15 variables que contribuyen a \hat{F}_3^L

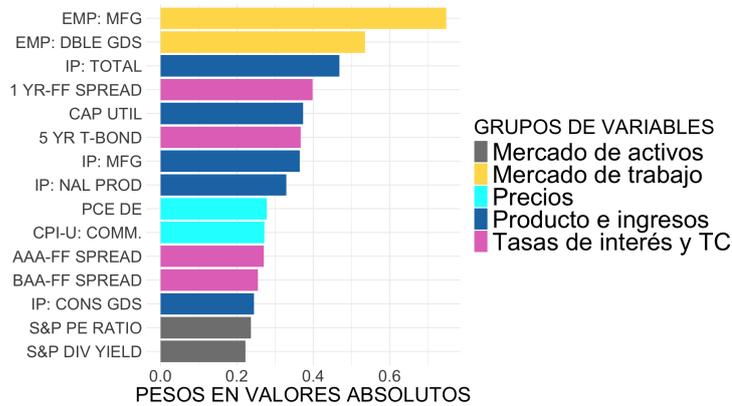


Figura 4.10: Primeras 15 variables que contribuyen a \hat{F}_2^P

Nota: Descripción de las variables en Apéndice 3, Tablas 3.3 y 3.9. Los grupos de variables por color se detallan en la Tabla 4.4.

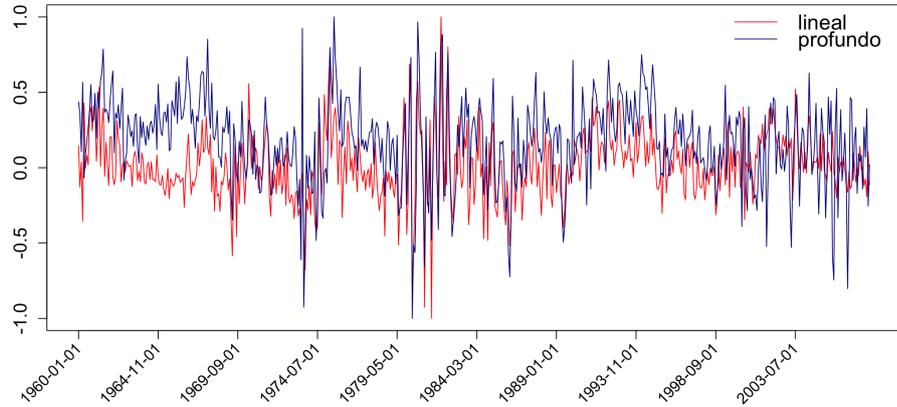


Figura 4.11: El cuarto factor lineal \hat{F}_4^L (línea roja) se correlaciona con el cuarto factor profundo \hat{F}_4^P (línea azul)

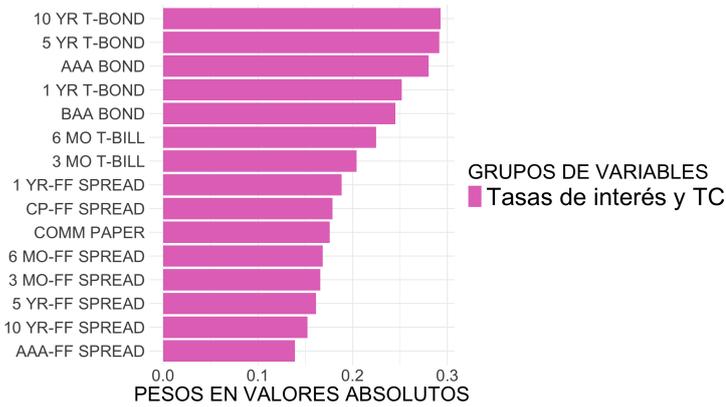


Figura 4.12: Primeras 15 variables que contribuyen a \hat{F}_4^L



Figura 4.13: Primeras 15 variables que contribuyen a \hat{F}_4^P

Nota: Descripción de las variables en Apéndice 3, Tablas 3.4 y 3.11. Los grupos de variables por color se detallan en la Tabla 4.4.

Análisis gráfico de los resultados: Tomando a las variables por grupos (Tabla 4.4), en términos generales, los factores lineales se relacionan con grupos más homogéneos de variables comparado con los profundos. En particular, considerando los primeros 4 factores lineales estimados, \hat{F}_1 se relaciona principalmente con variables del grupo de «Producto e ingresos» y «Mercado de Trabajo». \hat{F}_2 se vincula básicamente con variables asociadas a las «Tasas de interés y tipos de cambio». \hat{F}_3 se relaciona principalmente con variables del grupo «Precios» y el \hat{F}_4 también se encuentra vinculado en sus primeras 15 variables con el grupo de «Tasas de interés y tipos de cambio». Por su lado, los factores profundos, para todos los casos, presentan una relación más variada entre grupos de variables. El grupo de variables en «Dinero y crédito» es el que menos influye a los factores estimados tanto lineales como profundos considerando las primeras 15 variables más importantes en cada caso.

4.3. Análisis de impacto de política monetaria

Shock de PM y sus efectos sobre la economía

En Christiano et al. (1999), se consideran las siguientes tres interpretaciones económicas de los shocks de PM: primero, podrían vincularse a decisiones de la autoridad monetaria dadas ciertas condiciones en el nivel de empleo e inflación; segundo, podrían surgir de resoluciones internas a la FED para controlar las expectativas de los agentes; tercero, podrían darse por errores de medida en los datos preliminares disponibles del Comité Federal de Mercado Abierto (FOMC) a la hora de tomar ciertas decisiones políticas¹.

Se plantea que si bien en la literatura no se ha llegado a definir, de manera consensuada, un conjunto particular de supuestos para identificar los efectos de una perturbación exógena de PM, hay un acuerdo generalizado sobre sus efectos cualitativos. En este sentido, considerando algunos agregados macroeconómicos, tras una perturbación contractiva de PM se esperan los siguientes resultados: las tasas de interés de corto plazo deberían crecer; el producto agregado, el empleo, los beneficios y los agregados monetarios deberían decrecer; el nivel general de precios debería responder de manera lenta; las medidas de los salarios deberían caer en cantidades modestas; los tipos de cambio deberían depreciarse inicialmente. Además, hay acuerdo en que los shocks de PM son responsables sólo de una parte pequeña del porcentaje de volatilidad del producto agregado y de una cantidad incluso menor para los movimientos del nivel de precios generalizado (Christiano et al. (1999), Bernanke et al. (2005)).

Funciones de respuesta al impulso

Por un lado, tomo como instrumento de PM a la Tasa de Fondos Federales (FFR) porque en la literatura se la considera una buena aproximación de los cambios de PM y además, sirve para informar sobre los movimientos futuros de otras variables macroeconómicas (ver Bernanke y Blinder (1992) y Balke y Emery (1994)). Por otro lado, considero un subconjunto de variables informacionales que fueron seleccionadas siguiendo a Duarte (2020) y Bernanke et al. (2005). En este esquema, aplico una política monetaria contractiva, aumentando la FFR en 25 puntos básicos, siguiendo a Duarte (2020). Entonces, se

¹En el Anexo 4 se describen los cometidos de la PM de la Reserva Federal de Estados Unidos (FED).

analizan los efectos de la innovación sobre los 19 indicadores macroeconómicos² de la Tabla 4.5. Se calculan las FIR que se grafican en las Figuras 4.14, 4.15 y las FEVD³ que se expresan en las Tablas 4.6 y 4.7 (código: FIR FEVD lineal y FIR FEVD profundo).

Tabla 4.5: Variables para el análisis de impacto de PM

Variables	Grupos de variables
Fed Funds Rate	Tasas de interés y tipos de cambio
IP Index	Producto e ingresos
CPI	Precios
3-Month Treasury Bill	Tasas de interés y tipos de cambio
5-Year Treasury Rate	Tasas de interés y tipos de cambio
M1 Money Stock	Dinero y crédito
M2 Money Stock	Dinero y crédito
Exchange rate: Japan/U.S.	Tasas de interés y tipos de cambio
CPI : Commodities	Precios
Capacity Utilization	Producto e ingresos
Real personal consump.	Consumo, órdenes e inventarios
Durable goods consump.	Precios
Nondurable goods consump.	Precios
Unemployment Rate	Mercado de trabajo
Employment	Mercado de trabajo
Avg Hourly Earnings	Mercado de trabajo
Housing Starts	Vivienda
New Orders	Consumo, órdenes e inventarios
Dividend Yield	Mercado de activos

En las Figuras 4.14 y 4.15 comparo visualmente los movimientos de las respuestas de las variables frente a la PM contractiva con las figuras de Bernanke et al. (2005). Las curvas negras representan las funciones de respuesta al impulso de cada una de las 19 variables macroeconómicas clave frente a un aumento de 0.25 % de la Tasa de Fondos Federales (FFR). Las líneas punteadas rojas son los intervalos de confianza del 95 % asociados a los coeficientes de D_h en (2.9). Por su parte, el eje de las abscisas muestra los $h = 49$ pasos hacia adelante una vez que se genera el impacto de PM en el momento $t = 0$, es decir, cuando la FFR aumenta.

²Cabe mencionar que las variable en este estudio no son exactamente las mismas que las de Bernanke et al. (2005), con lo cual es esperable que los resultados difieran.

³Por sus siglas en inglés relativas a Forecast error variance decomposition.

De modo general, las tasas de interés de corto plazo deberían crecer; el producto agregado, el empleo, los beneficios y los agregados monetarios deberían decrecer; el nivel general de precios debería responder de manera lenta; las medidas de los salarios deberían caer en cantidades modestas; los tipos de cambio deberían depreciarse inicialmente (Bernanke et al. (2005)).

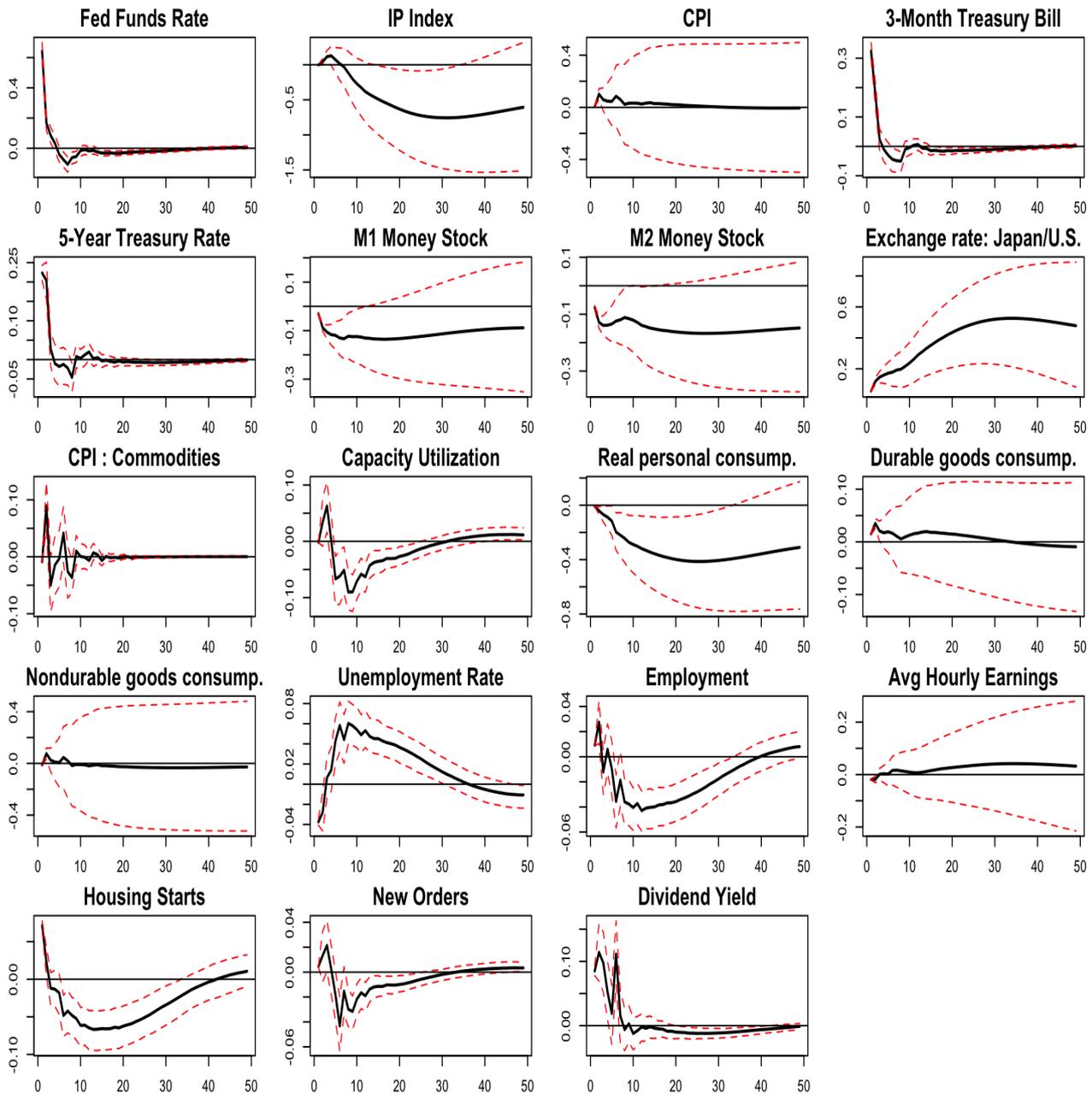


Figura 4.14: Funciones de respuesta al impulso en FAVAR lineal

Algunas variables presentan comportamientos similares a los esperados y presentados en Bernanke et al. (2005). Las mismas son IP index, CPI, 3-Month Treasury Bill, Capacity Utilization, Real personal Consumption, Non durable good consumption, Unemployment Rate, Employment, Housing Starts y New orders.

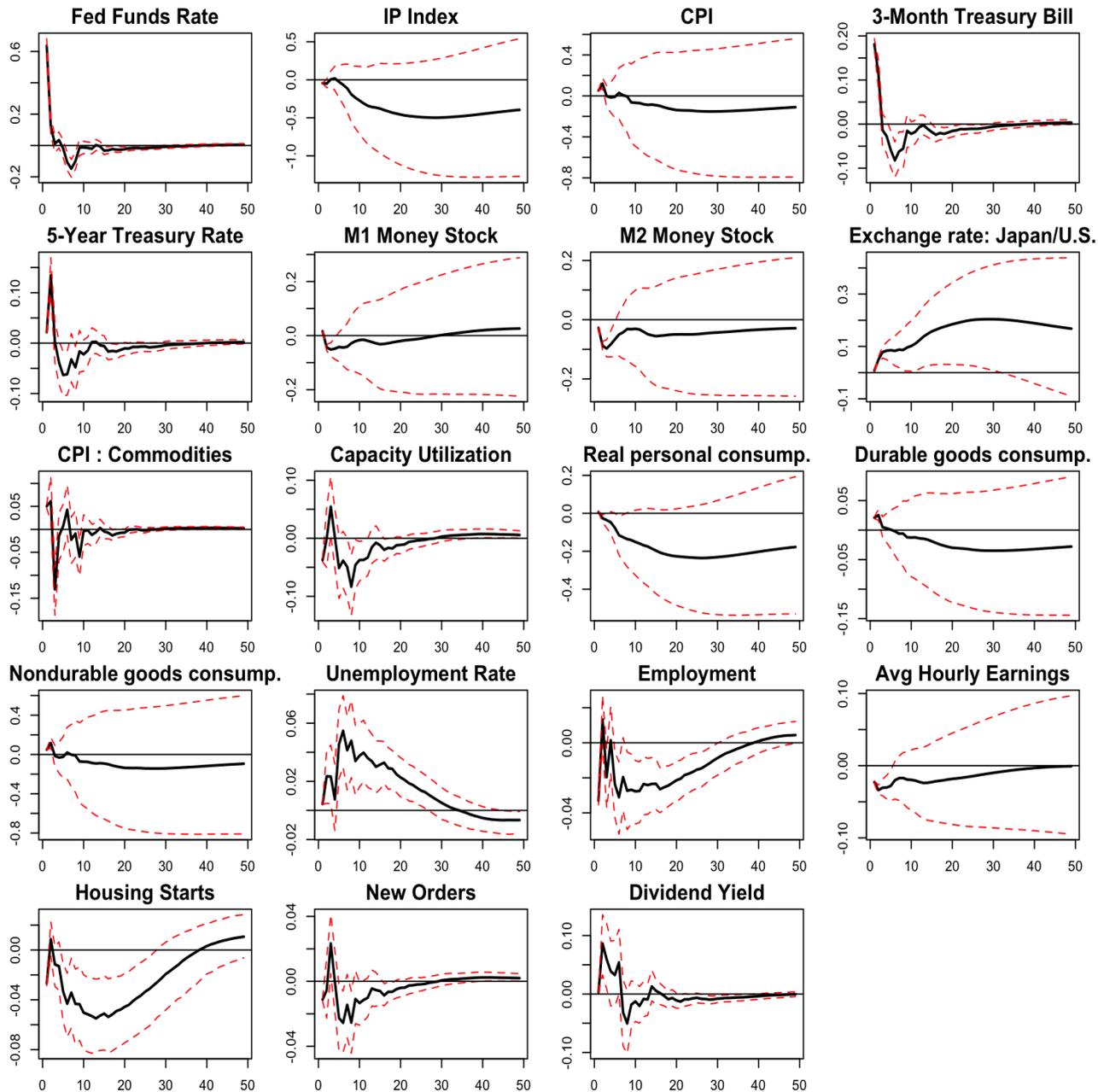


Figura 4.15: Funciones de respuesta al impulso en FAVAR profundo

Para el caso no lineal los resultados son similares al caso lineal (Figura 4.14). Las variables que se comportan de acuerdo a lo esperado son IP index, CPI, 3-Month Treasury Bill, M1, Capacity Utilization, Real personal consumption, Durable goods consumption, Nondurable goods consumption, Unemployment rate, Employment, Avg Hourly Earnings, Housing Starts y New Orders.

Descomposición de la varianza del error de predicción

Siguiendo a Bernanke et al. (2005), la FEVD mide la fracción del error de predicción de una determinada variable que se puede atribuir a un determinado shock, para un horizonte temporal dado. Los resultados de la FEVD surgen de los coeficientes de la representación Medias móviles (MA) del VAR y de la varianza de las perturbaciones estructurales. Si se considera a la fracción de la varianza de $X_{t+s} - \hat{X}_{t+s}$ dada por las perturbaciones de política monetaria ε_t^{PM} , se tiene

$$\frac{\text{var}(X_{t+s} - \hat{X}_{t+s|t} | \varepsilon_t^{PM})}{\text{var}(X_{t+s} - \hat{X}_{t+s|t})}. \quad (4.1)$$

Se plantea que un resultado típico en la literatura asociada a los *VAR* es que las perturbaciones de PM explican una fracción relativamente modesta del error de predicción de las medidas de actividad real de la economía y de la inflación.

En las siguientes Tablas 4.6 y 4.7, la primera columna mide la contribución del shock a la varianza del error de predicción, en un horizonte temporal de 60 meses, para cada una de las variables de interés. La segunda columna, contiene el R^2 del componente común (\hat{F}_t, Y_t) para cada una de las 19 variables. Es decir, el coeficiente de determinación (R^2) mide la fracción de la varianza de cada variable explicada por los componentes comunes (\hat{F}_t, Y_t). Como la FFR es un factor observable, su R^2 es igual a 1 por construcción.

Tabla 4.6: Contribución del shock de PM a la varianza del componente común (\hat{F}'_t, Y'_t) para el modelo FAVAR lineal

Variables	Descomposición de la Varianza	R^2
Fed Funds Rate	0.721	1.000
IP Index	0.084	0.924
Capacity Utilization	0.092	0.903
Housing Starts	0.174	0.876
5-Year Treasury Rate	0.144	0.834
CPI : Commodities	0.022	0.805
Nondurable goods consump.	0.021	0.796
Dividend Yield	0.076	0.794
CPI	0.023	0.760
3-Month Treasury Bill	0.227	0.641
Unemployment Rate	0.077	0.367
Employment	0.052	0.314
Real personal consump.	0.022	0.271
New Orders	0.012	0.177
Exchange rate: Japan/U.S.	0.024	0.135
M1 Money Stock	0.008	0.076
M2 Money Stock	0.014	0.075
Avg Hourly Earnings	0.002	0.059
Durable goods consump.	0.002	0.031

Por un lado, la contribución del shock de PM para el modelo FAVAR lineal (Tabla 4.6) oscila entre 0.2 % y 22.7 %. Comparando con los resultados de Bernanke et al. (2005, p.413) que varían entre 0 % y 10.3 %, el efecto de la PM sobre el conjunto de variables es mayor. En particular, el shock explica un 22.7 % para 3-Month Treasury Bill (tasa de interés de corto plazo) , 17.4 % para Housing Starts (inversión inmobiliaria) , 14.4 % para 5-Year Treasury Rate (tasa de interés de largo plazo), 9.2 % Capacity Utilization (capacidad utilizada) y 7.7 % para Unemployment Rate (desempleo).

Por otro lado, analizando el R^2 del componente común, los factores explican un fracción considerable de la varianza para estos indicadores macroeconómicos. Los indicadores que más resaltan son IP Index (índice de producción industrial) con 92.4 %, Capacity Utilization (capacidad utilizada) con 90.3 %, Housing Starts (inversión inmobiliaria) con 87.6 %, 5-Year Treasury Rate (tasa de interés de largo plazo) con 83.4 % , y CPI : Commodities (IPC commodities) con 80.5 %. Como en Bernanke et al. (2005), estos resultados confirmarían que el enfoque FAVAR estimado por PC permite capturar una dimensión considerable del ciclo de negocios. Sin embargo, el R^2 para el componente común es

bajo para Durable goods consump. (consumo de bienes durables) con 0.3% , Avg Hourly Earnings (salario promedio por hora) 0.6% , M2 Money Stock con 0.8%. Esto indica que los resultados asociados a las funciones de respuesta al impulso para estas variables también deberían ser poco confiables.

Tabla 4.7: Contribución del shock de PM a la varianza del componente común (\hat{F}'_t, Y'_t) para el modelo FAVAR profundo

Variables	Descomposición de la Varianza	R^2
Fed Funds Rate	0.560	1.000
Housing Starts	0.067	0.924
IP Index	0.033	0.915
Capacity Utilization	0.034	0.908
Dividend Yield	0.027	0.892
Nondurable goods consump.	0.039	0.798
CPI : Commodities	0.038	0.788
CPI	0.036	0.717
5-Year Treasury Rate	0.043	0.627
3-Month Treasury Bill	0.087	0.521
Unemployment Rate	0.035	0.315
Employment	0.021	0.221
Real personal consump.	0.007	0.135
New Orders	0.005	0.128
M1 Money Stock	0.006	0.068
M2 Money Stock	0.007	0.066
Exchange rate: Japan/U.S.	0.005	0.057
Durable goods consump.	0.001	0.019
Avg Hourly Earnings	0.001	0.009

La contribución del shock de PM para el modelo FAVAR profundo oscila entre 0.1% y 8.7%. Comparando con los resultados de Bernanke et al. (2005, p.413) que varían entre 0% y 10.3%, el efecto de la PM sobre el conjunto de variables es menor. En particular, el shock explica un 8.7% para 3-Month Treasury Bill (tasa de interés de corto plazo), 6.7% para Housing Starts (inversión inmobiliaria), 4.3% para 5-Year Treasury Rate (tasa de interés de largo plazo). Para el resto de las variables las contribuciones son menores.

Con respecto al R^2 para el componente común, los factores explican un fraccción considerable de la varianza de algunas variables: Housing Starts (92.4%), IP Index (91.5%), Capacity Utilization (90.8%), Dividend Yield (89.2%). Los resultados son magros específicamente para Avg Hourly Earnings (0.9%), Durable goods consump. (1.9%) y Exchange rate: Japan/U.S. (5.7%).

Capítulo 5

Consideraciones finales

5.1. Conclusiones y trabajos futuros

Se estimaron tres modelos (FAVAR lineal, DDFM y FAVAR profundo) para, inicialmente, evaluar su rendimiento predictivo en el contexto de modelos dinámicos de factores lineales y de aprendizaje profundo. Los resultados indican que los tres modelos presentan desempeños predictivos similares. Si bien el error de generalización es menor para el modelo FAVAR profundo, las diferencias entre modelos son marginales. Al igual que Klieber (2024), en este estudio se propone extender el modelo FAVAR a un esquema más general donde se puedan incorporar factores estimados bajo distintas representaciones.

En una segunda etapa, se examinó la importancia de las variables informativas para relacionarse con los factores estimados por los modelos FAVAR lineal y profundo, encontrando diferencias significativas en los resultados. Mientras que para el caso lineal, los factores estarían vinculados a grupos más homogéneos de variables, para el caso profundo los resultados son más heterogéneos.

Por último, se llevó a cabo un análisis de impacto de PM contractiva en un conjunto de variables clave para investigar la respuesta de cada una frente a un shock. Con esto se busca avanzar en el debate sobre la integración de modelos no lineales en las estimaciones para la implementación de políticas monetarias. Del estudio de las funciones de respuesta al impulso y de la descomposición de la varianza del error de predicción, se obtienen resultados similares con ambos modelos.

Este trabajo busca arrojar luz sobre la pertinencia de ir en la dirección de modelos de cajas negras a la hora de aplicarlos en contextos socialmente sensibles como puede ser la política monetaria. Andreini et al. (2023a) indican que su contribución del modelo *DDFM* tiene como ventaja de que al ser lineal en la parte del decoder, no se pierde interpretabilidad de los resultados. Los autores citan a Rudin (2019), quien plantea dejar de usar modelos de cajas negras¹ para la toma de decisiones que implican un alto riesgo para la sociedad. En este marco, se cuestiona la tendencia reciente vinculada al desarrollo de modelos de Aprendizaje Automático que muchas veces dejan de ser interpretables, explicativos y transparentes para los humanos. La discusión

¹Se dice que un modelo es de caja negra cuando contiene funciones complicadas de entender por los humanos, cuando las funciones son propietarias o en ambos casos. Los modelos de DL tienden a ser de cajas negras por el primer caso, o sea, por usar funciones complicadas (Rudin (2019)).

toma un sentido ético a la hora de seleccionar modelos que puedan impactar negativamente sobre la vida humana. Critica la idea generalizada de que haya un compromiso entre precisión e interpretabilidad: un modelo más complejo no necesariamente es más preciso. En particular, lo anterior se cumple cuando los datos están bien estructurados y representan correctamente cierta realidad. La autora afirma que cuando se da lo anterior, tiende a haber poca diferencia en los resultados de los algoritmos. Agrega que la creencia generalizada en la literatura, de que existe un compromiso entre precisión e interpretabilidad ha llevado a la investigación en Aprendizaje Automático a poner el foco en los modelos de aprendizaje profundo descartando modelos más interpretables².

Este estudio evidencia que tanto el modelo FAVAR lineal como profundo arrojan resultados comparables e interpretables. Se podría avanzar en estudios posteriores tomando distintos momentos de la política monetaria de Estados Unidos y comparar también lo que ocurre para momentos de tiempos considerados estables y de crisis como en Klieber (2024). Además, se podrían incorporar otros métodos de estimación de los factores como el de Esperanza-Maximización (EM) implementado en Bańbura y Modugno (2014) y complejizar el análisis incorporando datos de frecuencias mixtas con datos faltantes.

²Un modelo es interpretable en la medida que se sepa y entienda su funcionamiento (Rudin (2019)).

Referencias bibliográficas

- Andreini, P., Izzo, C., y Ricco, G. (2023a). Deep Dynamic Factor Models. *arXiv*. <https://arxiv.org/abs/2007.11887>
- Andreini, P., Izzo, C., y Ricco, G. (2023b). Deep Dynamic Factor Models [GitHub repository]. <https://github.com/cosimoizzo/DDFM>
- Bai, J., y Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70(1), 191-221.
- Balke, N. S., y Emery, K. M. (1994). The federal funds rate as an indicator of monetary policy: evidence from the 1980s. *Economic and Financial Policy Review*, (Q I), 1-15.
- Bañbura, M., y Modugno, M. (2014). Maximum Likelihood Estimation Of Factor Models On Datasets With Arbitrary Pattern Of Missing Data. *Journal of Applied Econometrics*, 29(1), 133-160.
- Bellman, R. (1957). *Dynamic Programming* (1.^a ed.). Princeton University Press.
- Bengio, Y., Yao, L., Alain, G., y Vincent, P. (2013). Generalized Denoising Auto-Encoders as Generative Models.
- Bernanke, B. S. (1986). Alternative explanations of the money-income correlation. *Carnegie-Rochester Conference Series on Public Policy*, 25, 49-99.
- Bernanke, B. S., y Blinder, A. S. (1992). The Federal Funds Rate and the Channels of Monetary Transmission. *American Economic Review*, 82(4), 901-921.
- Bernanke, B. S., y Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50(3), 525-546.
- Bernanke, B. S., Boivin, J., y Elias, P. (2005). Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *The Quarterly Journal of Economics*, 120, 387-422.
- Blanchard, O. J., y Watson, M. W. (1984). *Are business cycles all alike?* National Bureau of Economic Research Cambridge, MA.

- Bluwstein, K., Buckmann, M., Joseph, A., Kapadia, S., y Simsek, Ö. (2021, noviembre). *Credit Growth, the Yield Curve and Financial Crisis Prediction: Evidence from a Machine Learning Approach* (inf. téc. N.º 2021/2614). ECB Working Paper.
- Borup, D., Goulet Coulombe, P., Rapach, D., Schütte, E. C. M., y Schwenk-Nebbe, S. (2022, noviembre). *The Anatomy of Out-of-Sample Forecasting Accuracy* (inf. téc. N.º 2022-16). FRB Atlanta Working Paper.
- Bucacos, E. (2015). *Impact of international monetary policy in Uruguay: a FAVAR approach* (Documentos de trabajo). Banco Central del Uruguay.
- Chen, Y. B., Ng, S., y Bai, J. (2023). *fbi: Factor-Based Imputation and FRED-MD/QD Data Set* [R package version 0.7.0]. <https://github.com/cykbennie/fbi>
- Christiano, L. J., Eichenbaum, M., y Evans, C. L. (1997). Sticky price and limited participation models of money: A comparison. *European Economic Review*, 41(6), 1201-1249.
- Christiano, L. J., Eichenbaum, M., y Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? En J. B. Taylor y M. Woodford (Eds.), *Handbook of Macroeconomics* (pp. 65-148, Vol. 1). Elsevier.
- Coulombe, P. G., y Goebel, M. (2024). Maximally Machine-Learnable Portfolios. *arXiv*. <https://arxiv.org/abs/2306.05568>
- Covert, I., y Lee, S.-I. (2021, 13–15 Apr). Improving KernelSHAP: Practical Shapley Value Estimation Using Linear Regression. En A. Banerjee y K. Fukumizu (Eds.), *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics* (pp. 3457-3465, Vol. 130). PMLR. <https://proceedings.mlr.press/v130/covert21a.html>
- Crawford, L., Flaxman, S. R., Runcie, D. E., y West, M. (2019). Variable Prioritization in Nonlinear Black Box Methods: A Genetic Association Case Study. *The Annals of Applied Statistics*, 13(2), 958-989. <https://doi.org/10.1214/18-aos1222>
- Crawford, L., Wood, K. C., Zhou, X., y Mukherjee, S. (2018). Bayesian Approximate Kernel Regression With Variable Selection. *Journal of the American Statistical Association*, 113(524), 1710-1721.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 2(4), 303-314.

- Despois, T., y Doz, C. (2023). Identifying and interpreting the factors in factor models via sparsity: Different approaches. *Journal of Applied Econometrics*, 38(4).
- Dickey, D. A., y Fuller, W. A. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with Unit Root. *Econometrica*, 49, 1057-1072.
- Duarte, J. (2020). FAVAR BBE(2005) Replication in R [GitHub repository]. https://github.com/jbduarte/blog/tree/master/_notebooks
- Enders, W. (2014). *Applied Econometric Time Series* (4.^a ed.). Wiley.
- FED. (2018). Monetary Policy Principles and Practice [Página Web]. <https://www.federalreserve.gov/monetarypolicy/monetary-policy-principles-and-practice.htm>
- Giannone, D., Reichlin, L., y Bańbura, M. (2010, diciembre). *Nowcasting* (Working Paper Series). European Central Bank.
- Glorot, X., y Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research - Proceedings Track*, 9, 249-256.
- Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep Learning* [<http://www.deeplearningbook.org>]. MIT Press.
- Hamilton, J. D. (1994). Lag Operators. En *Time Series Analysis* (pp. 25-42). Princeton University Press.
- Hornik, K., Stinchcombe, M., y White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 359-366.
- Hornik, K., Stinchcombe, M., y White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*, 3(5), 551-560.
- Kingma, D. P., y Ba, J. (2017). Adam: A Method for Stochastic Optimization.
- Klieber, K. (2024). Non-linear dimension reduction in factor-augmented vector autoregressions. *Journal of Economic Dynamics and Control*, 159, 104800. <https://doi.org/https://doi.org/10.1016/j.jedc.2023.104800>
- Krantz, S., y Bagdziunas, R. (2023). *dfms: Dynamic Factor Models* [R package version 0.2.1]. <https://CRAN.R-project.org/package=dfms>
- Lecun, Y., y Soulie Fogelman, F. (1987). Modeles connexionnistes de l'apprentissage. *Intellectica, special issue apprentissage et machine*, 2.
- Lu, Z., Pu, H., Wang, F., Hu, Z., y Wang, L. (2017). The Expressive Power of Neural Networks: A View from the Width.

- Lucas, R. E. (1980). Methods and Problems in Business Cycle Theory. *Journal of Money, Credit and Banking*, 12(4), 696-715. <http://www.jstor.org/stable/1992030>
- Ludvigson, S., y Ng, S. (2010). A factor analysis of bond risk premia. En A. Ulah y D. Giles (Eds.), *Handbook of empirical economics and finance* (pp. 313-372, Vol. 1). Chapman; Hall, Boca Raton, FL.
- Lundberg, S. M., y Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. En I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan y R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc.
- Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Jia, Y., Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, ... Xiaoqiang Zheng. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems [Software available from tensorflow.org]. <https://www.tensorflow.org/>
- Mayer, M., y Watson, D. (2023). *kernelshap: Kernel SHAP* [R package version 0.4.1]. <https://CRAN.R-project.org/package=kernelshap>
- McCracken, M. W., y Ng, S. (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics*, 34(4), 574-589. <https://research.stlouisfed.org/econ/mccracken/sel/>
- Moritz, S., y Bartz-Beielstein, T. (2017). imputeTS: Time Series Missing Value Imputation in R. *The R Journal*, 9(1), 207-218. <https://steffenmoritz.github.io/imputeTS>
- Pfaff, B. (2008). VAR, SVAR and SVEC Models: Implementation Within R Package vars. *Journal of Statistical Software*, 27(4).
- Python Software Foundation. (2023). Python Language Reference, version 3.11.6.
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Ramey, V. A. (2016). Macroeconomic Shocks and Their Propagation. En J. B. Taylor y H. Uhlig (Eds.), *Handbook of Macroeconomics* (pp. 71-162, Vol. 2A). Elsevier.

- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1, 206-215.
- Rumelhart, D. E., Hinton, G. E., y Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533-536.
- Santurkar, S., Tsipras, D., Ilyas, A., y Madry, A. (2019). How Does Batch Normalization Help Optimization? *arXiv*. <https://arxiv.org/abs/1805.11604>
- Shapley, L. (1953). A Value for n-Person Games. En H. Kuhn y A. Tucker (Eds.), *Contributions to the Theory of Games II* (pp. 307-317). Princeton University Press.
- Sheppard, K., Khrapov, S., Lipták, G., mikedeltalima, Capellini, R., alejandrocermeno, Hugle, esvhd, bot, S., Fortin, A., JPN, Judell, M., Russell, R., Li, W., 645775992, Adams, A., jbrockmendel, Migrator, L., Rabbaba, M., ... Çelik, B. (2023, septiembre). *bashtage/arch: Release 6.2.0* (Ver. v6.2.0). Zenodo. <https://doi.org/10.5281/zenodo.8380532>
- Shin, Y., Kwiatkowski, D., Schmidt, P., y Phillips, P. (1992). Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Are Nonstationary? *Journal of Econometrics*, 54(1-3), 159-178.
- Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1), 1-48.
- Stanford University. (2023). CS231n: Convolutional Neural Networks for Visual Recognition. <http://cs231n.stanford.edu/>
- Stock, J. H., y Watson, M. W. (1996). Evidence on Structural Instability in Macroeconomic Time Series Relations. *Journal of Business and Economic Statistics*, 14, 11-30.
- Stock, J. H., y Watson, M. W. (2002). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics*, 20(2), 147-162.
- Stock, J. H., y Watson, M. W. (2005). *Implications of Dynamic Factor Models for VAR Analysis* (inf. téc. N.º 11467). National Bureau of Economic Research.
- Stock, J. H., y Watson, M. W. (2006). Forecasting with Many Predictors. En *Handbook of Forecasting*. North Holland.
- Stock, J. H., y Watson, M. W. (2011, julio). Dynamic Factor Models. En *The Oxford Handbook of Economic Forecasting*. Oxford University Press.

- Stock, J. H., y Watson, M. W. (2015). *Factor Models for Macroeconomics* (J. B. Taylor y H. Uhlig, Eds.; Vol. 2). North Holland.
- Stock, J. H., y Watson, M. W. (2016). Chapter 8 - Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics. En J. B. Taylor y H. Uhlig (Eds.). Elsevier.
- Strumbelj, E., y Kononenko, I. (2010). An Efficient Explanation of Individual Classifications using Game Theory. *The Journal of Machine Learning Research*, 11, 1-18.
- Theodoridis, S. (2015). *Machine Learning: A Bayesian and Optimization Perspective*. Academic Press.
- Valk, S. D., Mattos, D. D., y Ferreira, P. (2019). Nowcasting: An R Package for Predicting Economic Variables Using Dynamic Factor Models. *The R Journal*, 11(1), 230-244.
- Venables, W., y Ripley, B. (2002). *Modern Applied Statistics With S* (4.^a ed.). Springer.
- Wold, H. (1954). Causality and Econometrics. *Econometrica*, 22(2), 162-177.
- Wolpert, D., y Macready, W. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), 67-82.

APÉNDICES

Apéndice 1

Conjunto de datos

Códigos de transformación de las series de datos

Siendo x_t cada variable original, se tienen las siguientes transformaciones:

$$z_t = \begin{cases} x_t & tcode = 1 \\ \Delta x_t & tcode = 2 \\ \Delta^2 x_t & tcode = 3 \\ \log(x_t) & tcode = 4 \\ \Delta \log(x_t) & tcode = 5 \\ \Delta^2 \log(x_t) & tcode = 6 \\ \Delta (x_t/x_{t-1} - 1) & tcode = 7 \end{cases} \quad (1.1)$$

Variables, descripción, grupo, categorización en lenta o rápida y código de transformación

A continuación, en las Tablas 1.1, 1.2 y 1.3 se muestran las variables del estudio junto con sus descripciones, su clasificación en rápidas o lentas, el grupo al que pertenece cada serie, así como el código de transformación para que sea estacionaria. La clasificación de las variables en rápidas o lentas se realiza siguiendo el criterio descrito por Bernanke et al. (2005). Además, se incluye el grupo al que pertenece cada variable y el código de transformación (tcode) correspondiente de acuerdo a Chen et al. (2023). En la ecuación (1.1) se detalla el tipo de transformación de cada tcode.

Tabla 1.1: Variables, descripción, grupo de variables, categorización en lenta (1) o rápida (0) y código de transformación (tcode)

Variable	Descripción	Grupos de variables	Categoría	tcode
DPCERA3M086SBEA	Real personal consumption expenditures	Consumo, órdenes e inventarios	1	5
CMRMTSPLx	Real Manu. and Trade Industries Sales	Consumo, órdenes e inventarios	0	5
RETAILx	Retail and Food Services Sales	Consumo, órdenes e inventarios	0	5
AMDMNOx	New Orders for Durable Goods	Consumo, órdenes e inventarios	0	5
AMDMUOx	Un lled Orders for Durable Goods	Consumo, órdenes e inventarios	0	5
BUSINVx	Total Business Inventories	Consumo, órdenes e inventarios	0	5
ISRATIOx	Total Business: Inventories to Sales Ratio	Consumo, órdenes e inventarios	0	2
HOUST	Housing Starts: Total New Privately Owned	Vivienda	0	4
HOUSTNE	Housing Starts, Northeast	Vivienda	0	4
HOUSTW	Housing Starts, West	Vivienda	0	4
PERMIT	New Private Housing Permits (SAAR)	Vivienda	0	4
PERMITNE	New Private Housing Permits, Northeast (SAAR)	Vivienda	0	4
PERMITS	New Private Housing Permits, South (SAAR)	Vivienda	0	4
PERMITW	New Private Housing Permits, West (SAAR)	Vivienda	0	4
FEDFUNDS	Effective Federal Funds Rate	Tasas de interes y tipos de cambio	0	2
CP3Mx	3-Month AA Financial Commercial Paper Rate	Tasas de interes y tipos de cambio	0	2
TB3MS	3-Month Treasury Bill:	Tasas de interes y tipos de cambio	0	2
TB6MS	6-Month Treasury Bill:	Tasas de interes y tipos de cambio	0	2
GS1	1-Year Treasury Rate	Tasas de interes y tipos de cambio	0	2
GS5	5-Year Treasury Rate	Tasas de interes y tipos de cambio	0	2
GS10	10-Year Treasury Rate	Tasas de interes y tipos de cambio	0	2
AAA	Moody s Seasoned Aaa Corporate Bond Yield	Tasas de interes y tipos de cambio	0	2
BAA	Moody s Seasoned Baa Corporate Bond Yield	Tasas de interes y tipos de cambio	0	2
COMPAPFFx	3-Month Commercial Paper Minus FEDFUNDS	Tasas de interes y tipos de cambio	0	1
TB3SMFFM	3-Month Treasury C Minus FEDFUNDS	Tasas de interes y tipos de cambio	0	1
TB6SMFFM	6-Month Treasury C Minus FEDFUNDS	Tasas de interes y tipos de cambio	0	1
T1YFFM	1-Year Treasury C Minus FEDFUNDS	Tasas de interes y tipos de cambio	0	1
T5YFFM	5-Year Treasury C Minus FEDFUNDS	Tasas de interes y tipos de cambio	0	1
T10YFFM	10-Year Treasury C Minus FEDFUNDS	Tasas de interes y tipos de cambio	0	1
AAAFFM	Moody s Aaa Corporate Bond Minus FEDFUNDS	Tasas de interes y tipos de cambio	0	1
BAAFFM	Moody s Baa Corporate Bond Minus FEDFUNDS	Tasas de interes y tipos de cambio	0	1
EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	Tasas de interes y tipos de cambio	0	5
EXJPUSx	Japan / U.S. Foreign Exchange Rate	Tasas de interes y tipos de cambio	0	5
EXUSUKx	U.S. / U.K. Foreign Exchange Rate	Tasas de interes y tipos de cambio	0	5
EXCAUSx	Canada / U.S. Foreign Exchange Rate	Tasas de interes y tipos de cambio	0	5

Tabla 1.2: Variables, descripción, grupo de variables, categorización en lenta (1) o rápida (0) y código de transformación (tcode)

Variable	Descripción	Grupos de variables	Categoría	tcode
HWI	Help-Wanted Index for United States	Mercado de trabajo	1	2
HWIURATIO	Ratio of Help Wanted/No. Unemployed	Mercado de trabajo	1	2
CLF16OV	Civilian Labor Force	Mercado de trabajo	1	5
CE16OV	Civilian Employment	Mercado de trabajo	1	5
UNRATE	Civilian Unemployment Rate	Mercado de trabajo	1	2
UEMPMEAN	Average Duration of Unemployment (Weeks)	Mercado de trabajo	1	2
UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	Mercado de trabajo	1	5
UEMP5TO14	Civilians Unemployed for 5-14 Weeks	Mercado de trabajo	1	5
UEMP15OV	Civilians Unemployed - 15 Weeks & Over	Mercado de trabajo	1	5
UEMP15T26	Civilians Unemployed for 15-26 Weeks	Mercado de trabajo	1	5
UEMP27OV	Civilians Unemployed for 27 Weeks and Over	Mercado de trabajo	1	5
CLAIMSx	Initial Claims	Mercado de trabajo	1	5
PAYEMS	All Employees: Total nonfarm	Mercado de trabajo	1	5
USGOOD	All Employees: Goods-Producing Industries	Mercado de trabajo	1	5
CES1021000001	All Employees: Mining and Logging: Mining	Mercado de trabajo	1	5
USCONS	All Employees: Construction	Mercado de trabajo	1	5
MANEMP	All Employees: Manufacturing	Mercado de trabajo	1	5
DMANEMP	All Employees: Durable goods	Mercado de trabajo	1	5
NDMANEMP	All Employees: Nondurable goods	Mercado de trabajo	1	5
SRVPRD	All Employees: Service-Providing Industries	Mercado de trabajo	1	5
USTPU	All Employees: Trade, Transportation & Utilities	Mercado de trabajo	1	5
USWTRADE	All Employees: Wholesale Trade	Mercado de trabajo	1	5
USTRADE	All Employees: Retail Trade	Mercado de trabajo	1	5
USFIRE	All Employees: Financial Activities	Mercado de trabajo	1	5
USGOVT	All Employees: Government	Mercado de trabajo	1	5
CES0600000007	Avg Weekly Hours : Goods-Producing	Mercado de trabajo	1	1
AWOTMAN	Avg Weekly Overtime Hours : Manufacturing	Mercado de trabajo	1	2
AWHMAN	Avg Weekly Hours : Manufacturing	Mercado de trabajo	1	1
CES0600000008	Avg Hourly Earnings : Goods-Producing	Mercado de trabajo	1	6
CES2000000008	Avg Hourly Earnings : Construction	Mercado de trabajo	1	6
CES3000000008	Avg Hourly Earnings : Manufacturing	Mercado de trabajo	1	6
M1SL	M1 Money Stock	Dinero y crédito	0	6
M2SL	M2 Money Stock	Dinero y crédito	0	6
M2REAL	Real M2 Money Stock	Dinero y crédito	0	5
TOTRESNS	Total Reserves of Depository Institutions	Dinero y crédito	0	6
NONBORRES	Reserves Of Depository Institutions	Dinero y crédito	0	7
BUSLOANS	Commercial and Industrial Loans	Dinero y crédito	0	6
REALLN	Real Estate Loans at All Commercial Banks	Dinero y crédito	0	6
NONREVSL	Total Nonrevolving Credit	Dinero y crédito	0	6
CONSPI	Nonrevolving consumer credit to Personal Income	Dinero y crédito	0	2
DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding	Dinero y crédito	0	6
DTCTHFNM	Total Consumer Loans and Leases Outstanding	Dinero y crédito	0	6
INVEST	Securities in Bank Credit at All Commercial Banks	Dinero y crédito	0	6

Tabla 1.3: Variables, descripción, grupo de variables, categorización en lenta (1) o rápida (0) y código de transformación (tcode)

Variable	Descripción	Grupos de variables	Categoría	tcode
RPI	Real Personal Income	Producto e ingresos	1	5
W875RX1	Real personal income ex transfer receipts	Producto e ingresos	1	5
INDPRO	IP Index	Producto e ingresos	1	5
IPFPNSS	IP: Final Products and Nonindustrial Supplies	Producto e ingresos	1	5
IPFINAL	IP: Final Products (Market Group)	Producto e ingresos	1	5
IPCONGD	IP: Consumer Goods	Producto e ingresos	1	5
IPDCONGD	IP: Durable Consumer Goods	Producto e ingresos	1	5
IPNCONGD	IP: Nondurable Consumer Goods	Producto e ingresos	1	5
IPBUSEQ	IP: Business Equipment	Producto e ingresos	1	5
IPMAT	IP: Materials	Producto e ingresos	1	5
IPDMAT	IP: Durable Materials	Producto e ingresos	1	5
IPNMAT	IP: Nondurable Materials	Producto e ingresos	1	5
IPMANSICS	IP: Manufacturing (SIC)	Producto e ingresos	1	5
IPB51222S	IP: Residential Utilities	Producto e ingresos	1	5
IPFUELS	IP: Fuels	Producto e ingresos	1	5
CUMFNS	Capacity Utilization: Manufacturing	Producto e ingresos	1	2
WPSFD49207	PPI: Finished Goods	Precios	1	6
WPSFD49502	PPI: Finished Consumer Goods	Precios	1	6
WPSID61	PPI: Intermediate Materials	Precios	1	6
WPSID62	PPI: Crude Materials	Precios	1	6
OILPRICEx	Crude Oil, spliced WTI and Cushing	Precios	1	6
PPICMM	PPI: Metals and metal products:	Precios	1	6
CPIAUCSL	CPI : All Items	Precios	1	6
CPIAPPSL	CPI : Apparel	Precios	1	6
CPITRNSL	CPI : Transportation	Precios	1	6
CPIMEDSL	CPI : Medical Care	Precios	1	6
CUSR0000SAC	CPI : Commodities	Precios	1	6
CUSR0000SAD	CPI : Durables	Precios	1	6
CUSR0000SAS	CPI : Services	Precios	1	6
CPIULFSL	CPI : All Items Less Food	Precios	1	6
CUSR0000SA0L2	CPI : All items less shelter	Precios	1	6
CUSR0000SA0L5	CPI : All items less medical care	Precios	1	6
PCEPI	Personal Cons. Expend.: Chain Index	Precios	1	6
DDURRG3M086SBEA	Personal Cons. Exp: Durable goods	Precios	1	6
DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable goods	Precios	1	6
DSERRG3M086SBEA	Personal Cons. Exp: Services	Precios	1	6
S&P 500	S&P s Common Stock Price Index: Composite	Mercado de activos	0	5
S&P: indust	S&P s Common Stock Price Index: Industrials	Mercado de activos	0	5
S&P div yield	S&P s Composite Common Stock: Dividend Yield	Mercado de activos	0	2
S&P PE ratio	S&P s Composite Common Stock: Price-Earnings Ratio	Mercado de activos	0	5

Apéndice 2

Resultados de las pruebas ADF y KPSS

Resultados de las pruebas ADF para un nivel de confianza del 95%

Tabla 2.1: Resultados de las pruebas ADF: modelo ct

```

-----
-----
Results for trend 'ct':
ADF Test Results:
-----

Stationary Series Count: 118
Non-Stationary Series Count: 4
-----

Stationary Series as a List:
['RPI', 'W875RX1', 'DPCERA3M086SBEA', 'CMRMTSPLx',
'RETAILx', 'INDPRO', 'IPFPNSS', 'IPFINAL', 'IPCONGD',
'IPDCONGD', 'IPNCONGD', 'IPBUSEQ', 'IPMAT',
'IPDMAT', 'IPNMAT', 'IPMANSICS', 'IPB51222S',
'IPFUELS', 'CUMFNS', 'HWI', 'HWIURATIO', 'CLF160V',
'CE160V', 'UNRATE', 'UEMPMEAN', 'UEMPLT5',
'UEMP5T014', 'UEMP150V', 'UEMP15T26', 'UEMP270V',
'CLAIMSx', 'PAYEMS', 'USGOOD', 'CES1021000001',
'USCONS', 'MANEMP', 'DMANEMP', 'NDMANEMP', 'SRVPRD',
'USTPU', 'USWTRADE', 'USTRAD', 'USFIRE', 'USGOVT',
'CES0600000007', 'AWOTMAN', 'AWHMAN', 'HOUST',
'HOUSTW', 'PERMIT', 'PERMITNE', 'PERMITS',
'PERMITW', 'AMDMNOx', 'AMDMUOx', 'BUSINVx',
'ISRATIOx', 'M1SL', 'M2SL', 'M2REAL', 'BOGMBASE',
'TOTRESNS', 'NONBORRES', 'BUSLOANS', 'REALLN',
'NONREVSL', 'CONSPI', 'S.P.500', 'S.P..indust',
'S.P.div.yield', 'S.P.PE.ratio', 'FEDFUNDS', 'CP3Mx',
'TB3MS', 'TB6MS', 'GS1', 'GS5', 'GS10', 'AAA', 'BAA',
'COMPAPFFx', 'TB3SMFFM', 'TB6SMFFM', 'T1YFFM',
'T5YFFM', 'T10YFFM', 'AAAFFM', 'BAAFFM', 'EXSZUSx',
'EXJPUSx', 'EXUSUKx', 'EXCAUSx', 'WPSFD49207',
'WPSFD49502', 'WPSID61', 'WPSID62', 'OILPRICEx',
'PPICMM', 'CPIAUCSL', 'CPIAPPSL', 'CPITRNSL', 'CPIMEDSL',
'CUSR0000SAC', 'CUSR0000SAD', 'CUSR0000SAS', 'CPIULFSL',
'CUSR0000SA0L2', 'CUSR0000SA0L5', 'PCEPI', 'DDURRG3M086SBEA',
'DNDGRG3M086SBEA', 'DSERRG3M086SBEA', 'CES0600000008',
'CES2000000008', 'CES3000000008', 'DTCOLNVHFM', 'DTCTHFM',
'INVEST']

Non-Stationary Series as a List:
['HOUSTNE', 'HOUSTMw', 'HOUSTS', 'PERMITMW']

```

Tabla 2.2: Resultados de las pruebas ADF: modelo c

```
-----  
-----  
Results for trend 'c':  
ADF Test Results:  
-----  
Stationary Series Count: 118  
Non-Stationary Series Count: 4  
-----  
Stationary Series as a List:  
['RPI', 'W875RX1', 'DPCERA3M086SBEA', 'CMRMTSPLx',  
'RETAILx', 'INDPRO', 'IPFPNSS', 'IPFINAL', 'IPCONGD',  
, 'IPDCONGD', 'IPNCONGD', 'IPBUSEQ', 'IPMAT', 'IPDMAT',  
'IPNMAT', 'IPMANSICS', 'IPB51222S', 'IPFUELS', 'CUMFNS',  
, 'HWI', 'HWIURATIO', 'CLF160V', 'CE160V', 'UNRATE',  
'UEMPMEAN', 'UEMPLT5', 'UEMP5T014', 'UEMP150V',  
'UEMP15T26', 'UEMP270V', 'CLAIMSx', 'PAYEMS',  
'USGOOD', 'CES1021000001', 'USCONS', 'MANEMP',  
'DMANEMP', 'NDMANEMP', 'SRVPRD', 'USTPU',  
'USWTRADE', 'USTRADE', 'USFIRE', 'USGOVT',  
'CES0600000007', 'AWOTMAN', 'AWHMAN',  
'HOUST', 'HOUSTS', 'HOUSTW', 'PERMIT',  
'PERMITS', 'PERMITW', 'AMDMNOx', 'AMDMUOx',  
'BUSINVx', 'ISRATIOx', 'M1SL', 'M2SL', 'M2REAL',  
'BOGMBASE', 'TOTRESNS', 'NONBORRES', 'BUSLOANS',  
'REALLN', 'NONREVSL', 'CONSPI', 'S.P.500', 'S.P..indust',  
, 'S.P.div.yield', 'S.P.PE.ratio', 'FEDFUNDS', 'CP3Mx', 'TB3MS',  
'TB6MS', 'GS1', 'GS5', 'GS10', 'AAA', 'BAA', 'COMPAPFFx',  
'TB3SMFFM', 'TB6SMFFM', 'T1YFFM', 'T5YFFM', 'T10YFFM',  
'AAAFFM', 'BAAFFM', 'EXSZUSx', 'EXJPUSx', 'EXUSUKx',  
'EXCAUSx', 'WPSFD49207', 'WPSFD49502', 'WPSID61',  
, 'WPSID62', 'OILPRICEx', 'PPICMM', 'CPIAUCSL', 'CPIAPPSL',  
'CPITRNSL', 'CPIMEDSL', 'CUSR0000SAC', 'CUSR0000SAD',  
'CUSR0000SAS', 'CPIULFSL', 'CUSR0000SA0L2', 'CUSR0000SA0L5',  
'PCEPI', 'DDURRG3M086SBEA', 'DNDGRG3M086SBEA', 'DSERRG3M086SBEA',  
'CES0600000008', 'CES2000000008', 'CES3000000008', 'DTCOLNVHFM',  
'DTCTHFM', 'INVEST']  
  
Non-Stationary Series as a List:  
['HOUSTNE', 'HOUSTMW', 'PERMITNE', 'PERMITMW']
```

Tabla 2.3: Resultados de las pruebas ADF: modelo n

Results for trend 'n':

ADF Test Results:

Stationary Series Count: 109

Non-Stationary Series Count: 13

Stationary Series as a List:

['RPI', 'W875RX1', 'DPCERA3M086SBEA', 'CMRMTSPLx',
'RETAILx', 'INDPRO', 'IPFPNSS', 'IPFINAL', 'IPCONGD',
'IPDCONGD', 'IPNCONGD', 'IPBUSEQ', 'IPMAT', 'IPDMAT',
, 'IPNMAT', 'IPMANSICS', 'IPB51222S', 'IPFUELS', 'CUMFNS',
, 'HWI', 'HWIURATIO', 'CE160V', 'UNRATE', 'UEMPMEAN',
'UEMPLT5', 'UEMP5T014', 'UEMP150V', 'UEMP15T26',
'UEMP270V', 'CLAIMSx', 'PAYEMS', 'USGOOD',
'CES1021000001', 'USCONS', 'MANEMP', 'DMANEMP',
'NDMANEMP', 'SRVPRD', 'USTPU', 'USWTRADE', 'USTRAD',
'USFIRE', 'USGOVT', 'AWOTMAN', 'AMDMNOx', 'AMDMUOX',
'BUSINVx', 'ISRATIOx', 'M1SL', 'M2SL', 'M2REAL', 'BOGMBASE',
'TOTRESNS', 'NONBORRES', 'BUSLOANS', 'REALLN', 'NONREVS',
'CONSPI', 'S.P.500', 'S.P..indust', 'S.P.div.yield', 'S.P.PE.ratio',
'FEDFUNDS', 'CP3Mx', 'TB3MS', 'TB6MS', 'GS1', 'GS5', 'GS10', 'AAA',
'BAA', 'COMPAPFFx', 'TB3SMFFM', 'TB6SMFFM', 'T1YFFM', 'T5YFFM',
'T10YFFM', 'AAAFFM', 'BAAFFM', 'EXSZUSx', 'EXJPUSx', 'EXUSUKx',
'EXCAUSx', 'WPSFD49207', 'WPSFD49502', 'WPSID61', 'WPSID62',
'OILPRICEx', 'PPICMM', 'CPIAUCSL', 'CPIAPPSL', 'CPITRNSL', 'CPIMEDSL',
'CUSR0000SAC', 'CUSR0000SAD', 'CUSR0000SAS', 'CPIULFSL',
'CUSR0000SA0L2', 'CUSR0000SA0L5', 'PCEPI', 'DDURRG3M086SBEA',
'DNDGRG3M086SBEA', 'DSERRG3M086SBEA', 'CES0600000008',
'CES2000000008', 'CES3000000008', 'DTCOLNVHFM', 'DTCTHFM',
'INVEST']

Non-Stationary Series as a List:

['CLF160V', 'CES0600000007', 'AWHMAN', 'HOUST', 'HOUSTNE',
'HOUSTMW', 'HOUSTS', 'HOUSTW', 'PERMIT', 'PERMITNE',
'PERMITMW', 'PERMITS', 'PERMITW']

Resultados de las pruebas KPSS

Tabla 2.4: Resultados de las pruebas KPSS para un nivel de confianza del 95 %

```
-----  
---- TESTS KPSS -----  
-----  
Results for model=ct and nlags method=legacy:  
Count of Stationary series for model=ct and nlags method=legacy: 105  
Count of Non-Stationary series for model=ct and nlags method=legacy: 17  
-----  
Results for model=ct and nlags method=auto:  
Count of Stationary series for model=ct and nlags method=auto: 100  
Count of Non-Stationary series for model=ct and nlags method=auto: 22  
-----  
Results for model=ct and nlags method=Newey_West:  
Count of Stationary series for model=ct and nlags method=Newey_West: 100  
Count of Non-Stationary series for model=ct and nlags method=Newey_West: 22  
-----  
Results for model=ct and nlags method=Schwert:  
Count of Stationary series for model=ct and nlags method=Schwert: 102  
Count of Non-Stationary series for model=ct and nlags method=Schwert: 20  
-----  
Results for model=c and nlags method=legacy:  
Count of Stationary series for model=c and nlags method=legacy: 91  
Count of Non-Stationary series for model=c and nlags method=legacy: 31  
-----  
Results for model=c and nlags method=auto:  
Count of Stationary series for model=c and nlags method=auto: 79  
Count of Non-Stationary series for model=c and nlags method=auto: 43  
-----  
Results for model=c and nlags method=Newey_West:  
Count of Stationary series for model=c and nlags method=Newey_West: 79  
Count of Non-Stationary series for model=c and nlags method=Newey_West: 43  
-----  
Results for model=c and nlags method=Schwert:  
Count of Stationary series for model=c and nlags method=Schwert: 83  
Count of Non-Stationary series for model=c and nlags method=Schwert: 39  
-----  
-----  
Best model and lags combination with the highest stationary count:  
Model=ct, NLags=legacy, Count_stationaries=105, Count_non_stationaries=17
```

Tabla 2.5: Series en el mejor modelo KPSS para un nivel de confianza del 95 %

Best model and lags combination with the highest stationary count:

Model=ct, NLags=legacy, Count_stationaries=105, Count_non_stationaries=17

List of Stationary series for the best combination:

```
['RPI', 'W875RX1', 'DPCERA3M086SBEA', 'CMRMTSPLx',  
'RETAILx', 'INDPRO', 'IPFPNSS', 'IPFINAL', 'IPCONGD',  
'IPDCONGD', 'IPNCONGD', 'IPBUSEQ', 'IPMAT', 'IPDMAT',  
'IPNMAT', 'IPMANSICS', 'IPB51222S', 'IPFUELS', 'CUMFNS',  
, 'HWI', 'HWIURATIO', 'CE160V', 'UNRATE', 'UEMPMEAN',  
'UEMPLT5', 'UEMP5T014', 'UEMP150V', 'UEMP15T26',  
'UEMP270V', 'CLAIMSx', 'PAYEMS', 'USGOOD',  
'CES1021000001', 'USCONS', 'MANEMP', 'DMANEMP',  
'SRVPRD', 'USTPU', 'USWTRADE', 'USTRAD', 'USFIRE',  
'AWOTMAN', 'HOUSTNE', 'PERMITNE', 'AMDMNOx',  
'AMDMUOx', 'BUSINVx', 'ISRATIOx', 'M1SL', 'M2SL',  
'BOGMBASE', 'TOTRESNS', 'NONBORRES', 'BUSLOANS',  
'REALLN', 'NONREVSL', 'CONSPI', 'S.P.500', 'S.P..indust',  
'S.P.div.yield', 'S.P.PE.ratio', 'FEDFUNDS', 'CP3Mx', 'TB3MS',  
'TB6MS', 'GS1', 'GS5', 'GS10', 'AAA', 'BAA', 'T1YFFM', 'T5YFFM',  
'T10YFFM', 'AAAFFM', 'BAAFFM', 'EXSZUSx', 'EXJPUSx', 'EXUSUKx',  
'EXCAUSx', 'WPSFD49207', 'WPSFD49502', 'WPSID61', 'WPSID62',  
'OILPRICEx', 'PPICMM', 'CPIAUCSL', 'CPIAPPSL', 'CPITRNSL', 'CPIMEDSL',  
'CUSR0000SAC', 'CUSR0000SAD', 'CUSR0000SAS', 'CPIULFSL', 'CUSR0000SAOL2',  
'CUSR0000SAOL5', 'PCEPI', 'DDURRG3M086SBEA', 'DNDGRG3M086SBEA',  
'DSERRG3M086SBEA', 'CES0600000008', 'CES2000000008', 'CES3000000008',  
'DTCOLNVHFM', 'DTCTHFM', 'INVEST']
```

List of Non-Stationary series for the best combination:

```
['CLF160V', 'NDMANEMP', 'USGOVT', 'CES0600000007',  
'AWHMAN', 'HOUST', 'HOUSTMW', 'HOUSTS', 'HOUSTW', 'PERMIT',  
'PERMITMW', 'PERMITS', 'PERMITW', 'M2REAL', 'COMPAPFFx', 'TB3SMFFM',  
'TB6SMFFM']
```

Apéndice 3

Importancia de las variables sobre los factores

Correlación entre factores e importancia de las variables

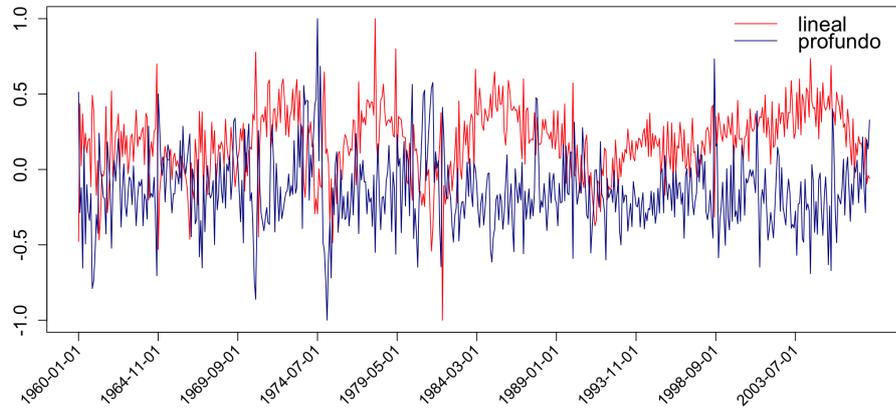


Figura 3.1: El quinto factor lineal \hat{F}_5^L (línea roja) se correlaciona con el tercer factor profundo \hat{F}_3^P (línea azul)

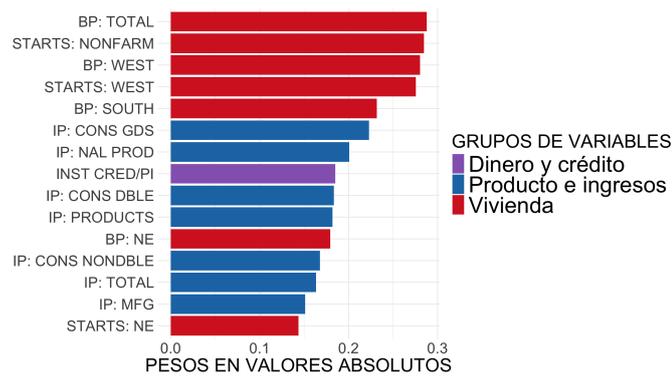


Figura 3.2: Primeras 15 variables que contribuyen a \hat{F}_5^L



Figura 3.3: Primeras 15 variables que contribuyen a \hat{F}_3^P

Nota: Descripción de las variables en Tablas 3.5 y 3.10. Los grupos de variables por color se ubican en la Tabla 4.4.

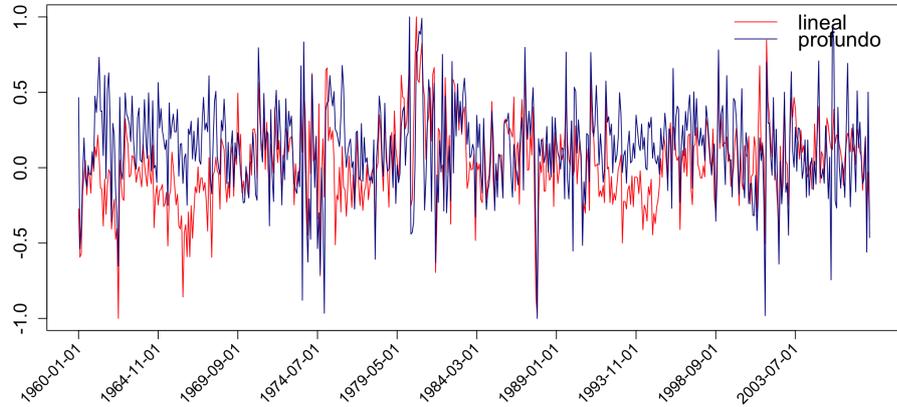


Figura 3.4: El sexto factor lineal \hat{F}_6^L (línea roja) se correlaciona con el quinto factor profundo \hat{F}_5^P (línea azul)

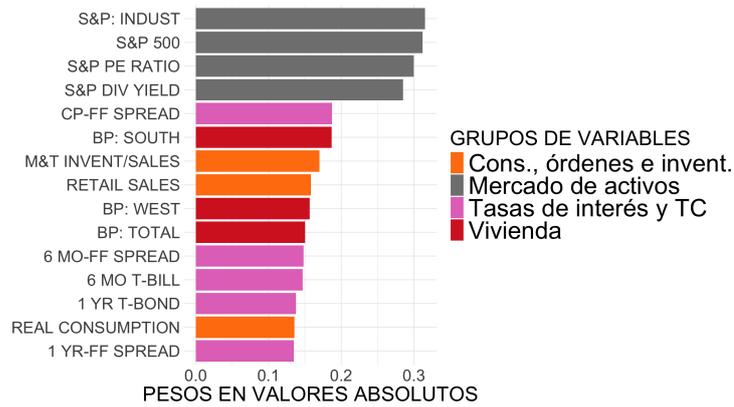


Figura 3.5: Primeras 15 variables que contribuyen a \hat{F}_6^L



Figura 3.6: Primeras 15 variables que contribuyen a \hat{F}_5^P

Nota: Descripción de las variables en Tablas 3.6 y 3.12. Los grupos de variables por color se ubican en la Tabla 4.4.

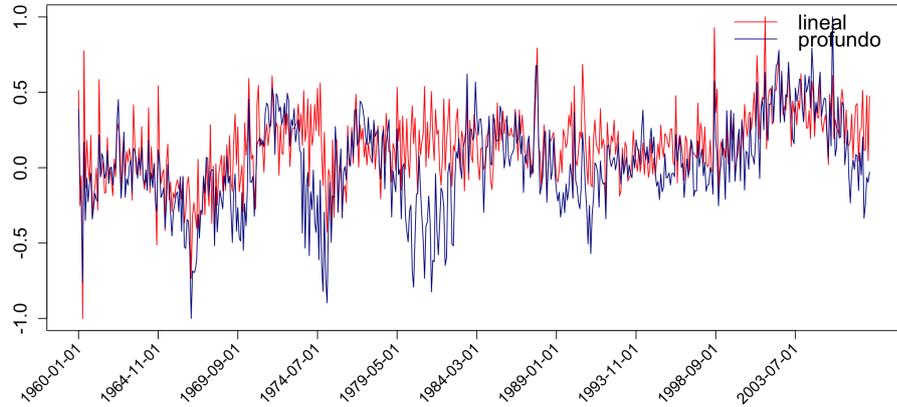


Figura 3.7: El séptimo factor lineal \hat{F}_7^L (línea roja) se correlaciona con el sexto factor profundo \hat{F}_2^P (línea azul)

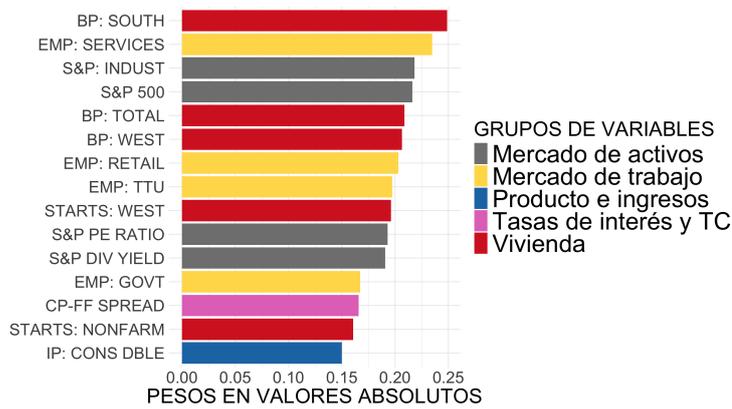


Figura 3.8: Primeras 15 variables que contribuyen a \hat{F}_7^L



Figura 3.9: Primeras 15 variables que contribuyen a \hat{F}_6^P

Nota: Descripción de las variables en Tablas 3.7 y 3.13. Los grupos de variables por color se ubican en la Tabla 4.4.

Descripción de las primeras 15 variables asociadas a los factores estimados

Factores lineales

Tabla 3.1: Primeras 15 variables y su descripción asociadas al factor 1 lineal

Variables	Descripción
IP: mfg	IP: Manufacturing (SIC)
IP: total	IP Index
Emp: gds prod	All Employees: Goods-Producing Industries
Cap util	Capacity Utilization: Manufacturing
Emp: total	All Employees: Total nonfarm
IP: products	IP: Final Products and Nonindustrial Supplies
Emp: mfg	All Employees: Manufacturing
IP: dble matls	IP: Durable Materials
Emp: dble gds	All Employees: Durable goods
IP: matls	IP: Materials
IP: nal prod	IP: Final Products (Market Group)
IP: bus eqpt	IP: Business Equipment
Emp: TTU	All Employees: Trade, Transportation & Utilities
IP: cons gds	IP: Consumer Goods
IP: cons dble	IP: Durable Consumer Goods

Tabla 3.2: Primeras 15 variables y su descripción asociadas al factor 2 lineal

Variables	Descripción
Aaa-FF spread	Moody s Aaa Corporate Bond Minus FEDFUNDS
Baa-FF spread	Moody s Baa Corporate Bond Minus FEDFUNDS
10 yr-FF spread	10-Year Treasury C Minus FEDFUNDS
5 yr-FF spread	5-Year Treasury C Minus FEDFUNDS
Baa bond	Moody s Seasoned Baa Corporate Bond Yield
6 mo-FF spread	6-Month Treasury C Minus FEDFUNDS
1 yr-FF spread	1-Year Treasury C Minus FEDFUNDS
3 mo-FF spread	3-Month Treasury C Minus FEDFUNDS
Aaa bond	Moody s Seasoned Aaa Corporate Bond Yield
M&T invent	Total Business Inventories
S&P PE ratio	S&P s Composite Common Stock: Price-Earnings Ratio
M2 (real)	Real M2 Money Stock
S&P div yield	S&P s Composite Common Stock: Dividend Yield
10 yr T-bond	10-Year Treasury Rate
5 yr T-bond	5-Year Treasury Rate

Tabla 3.3: Primeras 15 variables y su descripción asociadas al factor 3 lineal

Variables	Descripción
CPI-U: comm.	CPI : Commodities
PCE de : nondble	Personal Cons. Exp: Nondurable goods
CPI-U: ex shelter	CPI : All items less shelter
CPI-U: all	CPI : All Items
CPI-U: ex med	CPI : All items less medical care
PCE de	Personal Cons. Expend.: Chain Index
CPI-U: transp	CPI : Transportation
CPI-U: ex food	CPI : All Items Less Food
PPI: cons gds	PPI: Finished Consumer Goods
PPI: n gds	PPI: Finished Goods
PPI: int matls	PPI: Intermediate Materials
PPI: crude matls	PPI: Crude Materials
Spot market price	Crude Oil, spliced WTI and Cushing
Aaa-FF spread	Moody s Aaa Corporate Bond Minus FEDFUNDS
Baa-FF spread	Moody s Baa Corporate Bond Minus FEDFUNDS

Tabla 3.4: Primeras 15 variables y su descripción asociadas al factor 4 lineal

Variables	Descripción
10 yr T-bond	10-Year Treasury Rate
5 yr T-bond	5-Year Treasury Rate
Aaa bond	Moody s Seasoned Aaa Corporate Bond Yield
1 yr T-bond	1-Year Treasury Rate
Baa bond	Moody s Seasoned Baa Corporate Bond Yield
6 mo T-bill	6-Month Treasury Bill:
3 mo T-bill	3-Month Treasury Bill:
1 yr-FF spread	1-Year Treasury C Minus FEDFUNDS
CP-FF spread	3-Month Commercial Paper Minus FEDFUNDS
Comm paper	3-Month AA Financial Commercial Paper Rate
6 mo-FF spread	6-Month Treasury C Minus FEDFUNDS
3 mo-FF spread	3-Month Treasury C Minus FEDFUNDS
5 yr-FF spread	5-Year Treasury C Minus FEDFUNDS
10 yr-FF spread	10-Year Treasury C Minus FEDFUNDS
Aaa-FF spread	Moody s Aaa Corporate Bond Minus FEDFUNDS

Tabla 3.5: Primeras 15 variables y su descripción asociadas al factor 5 lineal

Variables	Descripción
BP: total	New Private Housing Permits (SAAR)
Starts: nonfarm	Housing Starts: Total New Privately Owned
BP: West	New Private Housing Permits, West (SAAR)
Starts: West	Housing Starts, West
BP: South	New Private Housing Permits, South (SAAR)
IP: cons gds	IP: Consumer Goods
IP: nal prod	IP: Final Products (Market Group)
Inst cred/PI	Nonrevolving consumer credit to Personal Income
IP: cons dble	IP: Durable Consumer Goods
IP: products	IP: Final Products and Nonindustrial Supplies
BP: NE	New Private Housing Permits, Northeast (SAAR)
IP: cons nondble	IP: Nondurable Consumer Goods
IP: total	IP Index
IP: mfg	IP: Manufacturing (SIC)
Starts: NE	Housing Starts, Northeast

Tabla 3.6: Primeras 15 variables y su descripción asociadas al factor 6 lineal

Variables	Descripción
S&P: indust	S&P s Common Stock Price Index: Industrials
S&P 500	S&P s Common Stock Price Index: Composite
S&P PE ratio	S&P s Composite Common Stock: Price-Earnings Ratio
S&P div yield	S&P s Composite Common Stock: Dividend Yield
CP-FF spread	3-Month Commercial Paper Minus FEDFUNDS
BP: South	New Private Housing Permits, South (SAAR)
M&T invent/sales	Total Business: Inventories to Sales Ratio
Retail sales	Retail and Food Services Sales
BP: West	New Private Housing Permits, West (SAAR)
BP: total	New Private Housing Permits (SAAR)
6 mo-FF spread	6-Month Treasury C Minus FEDFUNDS
6 mo T-bill	6-Month Treasury Bill:
1 yr T-bond	1-Year Treasury Rate
Real Consumption	Real personal consumption expenditures
1 yr-FF spread	1-Year Treasury C Minus FEDFUNDS

Tabla 3.7: Primeras 15 variables y su descripción asociadas al factor 7 lineal

Variables	Descripción
BP: South	New Private Housing Permits, South (SAAR)
Emp: services	All Employees: Service-Providing Industries
S&P: indust	S&P s Common Stock Price Index: Industrials
S&P 500	S&P s Common Stock Price Index: Composite
BP: total	New Private Housing Permits (SAAR)
BP: West	New Private Housing Permits, West (SAAR)
Emp: retail	All Employees: Retail Trade
Emp: TTU	All Employees: Trade, Transportation & Utilities
Starts: West	Housing Starts, West
S&P PE ratio	S&P s Composite Common Stock: Price-Earnings Ratio
S&P div yield	S&P s Composite Common Stock: Dividend Yield
Emp: Govt	All Employees: Government
CP-FF spread	3-Month Commercial Paper Minus FEDFUNDS
Starts: nonfarm	Housing Starts: Total New Privately Owned
IP: cons dble	IP: Durable Consumer Goods

Factores no lineales

Tabla 3.8: Primeras 15 variables y su descripción asociadas al factor 1 no lineal

Variables	Descripción
EMP: GDS PROD	All Employees: Goods-Producing Industries
IP: CONS GDS	IP: Consumer Goods
10 YR-FF SPREAD	10-Year Treasury C Minus FEDFUNDS
S&P 500	S&P s Common Stock Price Index: Composite
EMP: SERVICES	All Employees: Service-Providing Industries
AAA-FF SPREAD	Moody s Aaa Corporate Bond Minus FEDFUNDS
BP: TOTAL	New Private Housing Permits (SAAR)
6 MO-FF SPREAD	6-Month Treasury C Minus FEDFUNDS
PCE DE : NONDBLE	Personal Cons. Exp: Nondurable goods
EMP CPS NONAG	Civilian Employment
IP: MFG	IP: Manufacturing (SIC)
STARTS: NONFARM	Housing Starts: Total New Privately Owned
5 YR T-BOND	5-Year Treasury Rate
IP: CONS DBLE	IP: Durable Consumer Goods
IP: CONS NONDBLE	IP: Nondurable Consumer Goods

Tabla 3.9: Primeras 15 variables y su descripción asociadas al factor 2 no lineal

Variables	Descripción
EMP: MFG	All Employees: Manufacturing
EMP: DBLE GDS	All Employees: Durable goods
IP: TOTAL	IP Index
1 YR-FF SPREAD	1-Year Treasury C Minus FEDFUNDS
CAP UTIL	Capacity Utilization: Manufacturing
5 YR T-BOND	5-Year Treasury Rate
IP: MFG	IP: Manufacturing (SIC)
IP: NAL PROD	IP: Final Products (Market Group)
PCE DE	Personal Cons. Expend.: Chain Index
CPI-U: COMM.	CPI : Commodities
AAA-FF SPREAD	Moody s Aaa Corporate Bond Minus FEDFUNDS
BAA-FF SPREAD	Moody s Baa Corporate Bond Minus FEDFUNDS
IP: CONS GDS	IP: Consumer Goods
S&P PE RATIO	S&P s Composite Common Stock: Price-Earnings Ratio
S&P DIV YIELD	S&P s Composite Common Stock: Dividend Yield

Tabla 3.10: Primeras 15 variables y su descripción asociadas al factor 3 no lineal

Variables	Descripción
IP: CONS GDS	IP: Consumer Goods
5 YR-FF SPREAD	5-Year Treasury C Minus FEDFUNDS
EMP: GDS PROD	All Employees: Goods-Producing Industries
IP: MFG	IP: Manufacturing (SIC)
EMP: MFG	All Employees: Manufacturing
S&P 500	S&P s Common Stock Price Index: Composite
10 YR-FF SPREAD	10-Year Treasury C Minus FEDFUNDS
IP: CONS DBLE	IP: Durable Consumer Goods
EMP: SERVICES	All Employees: Service-Providing Industries
IP: CONS NONDBLE	IP: Nondurable Consumer Goods
3 MO-FF SPREAD	3-Month Treasury C Minus FEDFUNDS
BAA-FF SPREAD	Moody s Baa Corporate Bond Minus FEDFUNDS
PCE DE : NONDBLE	Personal Cons. Exp: Nondurable goods
IP: NAL PROD	IP: Final Products (Market Group)
S&P PE RATIO	S&P s Composite Common Stock: Price-Earnings Ratio

Tabla 3.11: Primeras 15 variables y su descripción asociadas al factor 4 no lineal

Variables	Descripción
5 YR-FF SPREAD	5-Year Treasury C Minus FEDFUNDS
EMP: GDS PROD	All Employees: Goods-Producing Industries
IP: CONS GDS	IP: Consumer Goods
EMP: MFG	All Employees: Manufacturing
10 YR-FF SPREAD	10-Year Treasury C Minus FEDFUNDS
IP: CONS DBLE	IP: Durable Consumer Goods
IP: CONS NONDBLE	IP: Nondurable Consumer Goods
PCE DE : NONDBLE	Personal Cons. Exp: Nondurable goods
IP: TOTAL	IP Index
3 MO-FF SPREAD	3-Month Treasury C Minus FEDFUNDS
EMP: SERVICES	All Employees: Service-Providing Industries
BP: WEST	New Private Housing Permits, West (SAAR)
BAA-FF SPREAD	Moody s Baa Corporate Bond Minus FEDFUNDS
STARTS: NONFARM	Housing Starts: Total New Privately Owned
EMP CPS NONAG	Civilian Employment

Tabla 3.12: Primeras 15 variables y su descripción asociadas al factor 5 no lineal

Variables	Descripción
EMP: MFG	All Employees: Manufacturing
S&P PE RATIO	S&P s Composite Common Stock: Price-Earnings Ratio
EMP: DBLE GDS	All Employees: Durable goods
S&P 500	S&P s Common Stock Price Index: Composite
S&P: INDUST	S&P s Common Stock Price Index: Industrials
S&P DIV YIELD	S&P s Composite Common Stock: Dividend Yield
IP: CONS GDS	IP: Consumer Goods
IP: MFG	IP: Manufacturing (SIC)
5 YR T-BOND	5-Year Treasury Rate
PCE DE : NONDBLE	Personal Cons. Exp: Nondurable goods
1 YR T-BOND	1-Year Treasury Rate
IP: PRODUCTS	IP: Final Products and Nonindustrial Supplies
IP: TOTAL	IP Index
EMP: NONDBLES	All Employees: Nondurable goods
CPI-U: EX SHELTER	CPI : All items less shelter

Tabla 3.13: Primeras 15 variables y su descripción asociadas al factor 6 no lineal

Variables	Descripción
S&P 500	S&P s Common Stock Price Index: Composite
BP: TOTAL	New Private Housing Permits (SAAR)
EMP: MFG	All Employees: Manufacturing
EMP: DBLE GDS	All Employees: Durable goods
10 YR-FF SPREAD	10-Year Treasury C Minus FEDFUNDS
AAA-FF SPREAD	Moody s Aaa Corporate Bond Minus FEDFUNDS
STARTS: NONFARM	Housing Starts: Total New Privately Owned
BP: WEST	New Private Housing Permits, West (SAAR)
STARTS: WEST	Housing Starts, West
BP: SOUTH	New Private Housing Permits, South (SAAR)
6 MO-FF SPREAD	6-Month Treasury C Minus FEDFUNDS
CAP UTIL	Capacity Utilization: Manufacturing
S&P: INDUST	S&P s Common Stock Price Index: Industrials
EMP: SERVICES	All Employees: Service-Providing Industries
S&P PE RATIO	S&P s Composite Common Stock: Price-Earnings Ratio

Tabla 3.14: Primeras 15 variables y su descripción asociadas al factor 7 no lineal

Variables	Descripción
EMP: MFG	All Employees: Manufacturing
IP: CONS GDS	IP: Consumer Goods
EMP: GDS PROD	All Employees: Goods-Producing Industries
10 YR-FF SPREAD	10-Year Treasury C Minus FEDFUNDS
IP: CONS DBLE	IP: Durable Consumer Goods
5 YR-FF SPREAD	5-Year Treasury C Minus FEDFUNDS
IP: CONS NONDBLE	IP: Nondurable Consumer Goods
IP: TOTAL	IP Index
AVG HRS	Avg Weekly Hours : Goods-Producing
IP: MATLS	IP: Materials
EMP: DBLE GDS	All Employees: Durable goods
EMP: SERVICES	All Employees: Service-Providing Industries
PCE DE : NONDBLE	Personal Cons. Exp: Nondurable goods
U: ALL	Civilian Unemployment Rate
EMP CPS NONAG	Civilian Employment

Tabla 3.15: Primeras 15 variables y su descripción asociadas al factor 8 no lineal

Variable	Descripcion
1 EMP: MFG	All Employees: Manufacturing
2 EMP: DBLE GDS	All Employees: Durable goods
3 S&P: INDUST	S&P s Common Stock Price Index: Industrials
4 BAA-FF SPREAD	Moody s Baa Corporate Bond Minus FEDFUNDS
5 CAP UTIL	Capacity Utilization: Manufacturing
6 EMP: NONDBLES	All Employees: Nondurable goods
7 IP: MFG	IP: Manufacturing (SIC)
8 S&P 500	S&P s Common Stock Price Index: Composite
9 AAA-FF SPREAD	Moody s Aaa Corporate Bond Minus FEDFUNDS
10 IP: NAL PROD	IP: Final Products (Market Group)
11 PCE DE : NONDBLE	Personal Cons. Exp: Nondurable goods
12 10 YR-FF SPREAD	10-Year Treasury C Minus FEDFUNDS
13 6 MO-FF SPREAD	6-Month Treasury C Minus FEDFUNDS
14 IP: TOTAL	IP Index
15 S&P PE RATIO	S&P s Composite Common Stock: Price-Earnings Ratio

ANEXOS

Anexo 1

Estacionariedad de las series temporales

Pruebas ADF

Se testea para cada serie de datos

$$H_0) \gamma = 0$$

$$H_1) \gamma \neq 0$$

En este sentido, se consideran tres ecuaciones de regresión distintas para poner a prueba H_0 ,

$$\Delta z_t = \gamma z_{t-1} + \sum_{i=2}^p \beta_i \Delta z_{t-i+1} + \varepsilon_t \quad (1.1)$$

$$\Delta z_t = a_0 + \gamma z_{t-1} + \sum_{i=2}^p \beta_i \Delta z_{t-i+1} + \varepsilon_t \quad (1.2)$$

$$\Delta z_t = a_0 + \gamma z_{t-1} + a_2 t + \sum_{i=2}^p \beta_i \Delta z_{t-i+1} + \varepsilon_t \quad (1.3)$$

Existen tres estadísticos τ , τ_μ , τ_τ para probar la hipótesis nula $H_0) \gamma = 0$ en cada caso y además, se tienen otros tres F -estadísticos ϕ_1 , ϕ_2 , ϕ_3 para hacer pruebas conjuntas sobre los coeficientes.

Los estadísticos ϕ_1 , ϕ_2 , ϕ_3 se construyen como pruebas F :

$$\phi_i = \frac{[SSR_{restringido} - SSR_{no\ restringido}] / r}{SSR_{no\ restringido} / (t - k)}$$

donde SSR es la suma de los cuadrados de los residuos en los modelos restringidos y no restringidos, r es la cantidad de restricciones, T es la cantidad de observaciones, k es la cantidad de parámetros estimados en el modelo irrestricto, $i = 1, 2, 3$. A su vez, $T - k$ van a ser los grados de libertad del modelos sin restricciones. Los valores de los coeficientes estimados se van a comparar con los valores críticos de tablas reportados por Dickey y Fuller (1981). La hipótesis nula indica que el proceso de generación de los datos es la del modelo restringido contra la hipótesis alternativa de que los datos son generados por el modelo sin restringir. Cuando los valores de ϕ_1 , ϕ_2 , ϕ_3 sean mayores a los valores críticos reportados por Dickey y Fuller (1981) se rechaza la hipótesis nula, cuando sean menores a los valores críticos entonces no se rechaza la hipótesis

nula.

En el siguiente cuadro 1.1 se consideran los tres modelos y cada una de las hipótesis a testear con sus respectivos estadísticos.

Tabla 1.1: Modelos del test ADF

	Modelo	H_0)	Estadístico de prueba
c	$\Delta z_t = a_0 + \gamma z_{t-1} + a_2 t + \sum_{i=2}^p \beta_i \Delta z_{t-i+1} + \varepsilon_t$	$\gamma = 0$ $\gamma = a_2 = 0$ $\gamma = a_2 = a_0 = 0$	τ_τ ϕ_3 ϕ_2
b	$\Delta z_t = a_0 + \gamma z_{t-1} + \sum_{i=2}^p \beta_i \Delta z_{t-i+1} + \varepsilon_t$	$\gamma = 0$ $\gamma = a_0 = 0$	τ_μ ϕ_1
a	$\Delta z_t = \gamma z_{t-1} + \sum_{i=2}^p \beta_i \Delta z_{t-i+1} + \varepsilon_t$	$\gamma = 0$	τ

Cuando no se conozca el proceso de generación de los datos, se sugiere realizar las pruebas de Dickey-Fuller Aumentado partiendo del modelo menos restrictivo para cada serie temporal a uno más particular. Si bien las pruebas de ADF son útiles para detectar la presencia de raíces unitarias, los mismos tienen sus limitaciones. Partiendo del modelo general al particular, cada prueba está condicionada a que las pruebas anteriores sean correctas. Cuando se empieza por el primer paso, es decir, con el modelo (c) con constante y con tendencia, se hace más difícil rechazar H_0), por lo tanto, cuando se rechaza la hipótesis nula en un modelo (c) se tiende a rechazar también la hipótesis nula cuando no se incluyen los términos deterministas. A su vez, establece que el problema principal de las pruebas de Dickey-Fuller es que tanto el intercepto como la pendiente de la tendencia son, con frecuencia, estimados de manera *pobre* bajo la presencia de raíces unitarias. En general, se tiende a no rechazar la hipótesis nula de raíz unitaria incluso cuando el verdadero valor de γ no es cero. Además, la prueba presenta limitaciones también frente a cambios de régimen (Enders (2014)).

Pruebas KPSS

En Shin et al. (1992), se parte de una representación cada serie temporal como la suma de un componente de tendencia determinística, un paseo aleatorio y un error estacionario. En este contexto, se pone a prueba

H_0) la serie es estacionaria alrededor de una tendencia

H_1) la serie es no estacionaria

que se corresponde con la hipótesis de que la varianza del paseo aleatorio (*random walk*) es igual a cero.

Se emplea un estadístico de Multiplicadores de Lagrange (ML) para testear la hipótesis nula de estacionariedad. De esta manera, siendo z_t con $t = 1, 2, \dots, T$ las series a las que se les quiere aplicar el test, se asume que se puede descomponer a la serie en la suma de un componente de tendencia determinística, un paseo aleatorio y un error estacionario se tiene que,

$$z_t = \xi t + r_t + \varepsilon_t \quad (1.4)$$

Donde r_t es un paseo aleatorio:

$$r_t = r_{t-1} + u_t, \quad (1.5)$$

donde u_t es $iid(0, \sigma_u^2)$. El valor inicial r_0 es fijo y sirve se intercepto. La hipótesis de estacionariedad es $\sigma_u^2 = 0$ y como se asume que ε_t es estacionario, bajo la hipótesis nula z_t es estacionaria alrededor de una tendencia.

En el caso particular de que en el modelo (1.4) se tenga $\xi = 0$, bajo la hipótesis nula z_t va a ser estacionaria alrededor de una constante (r_0).

Sean e_t con $t = 1, 2, \dots, T$, los residuos de la regresión z con un intercepto y tendencia. A su vez, sea $\hat{\sigma}_\varepsilon^2$ la estimación del error de la varianza de la regresión (suma de los residuos al cuadrado dividida T). Con lo anterior, se define el proceso de suma parcial de los residuos como

$$S_t = \sum_{i=1}^t r_i, \quad t = 1, \dots, T \quad (1.6)$$

Entonces el estadístico ML es

$$ML = \sum_{t=1}^T S_t^2 / \hat{\sigma}_\varepsilon^2 \quad (1.7)$$

En el caso de que se quiera poner a prueba la hipótesis nula de estacionariedad alrededor de una constante se define e_t como los residuos de la regresión z sobre un intercepto ($e_t = z_t - \bar{z}$). Cabe resaltar que es una prueba de cola superior y se reportan los valores críticos. Además, para este caso se asume que los errores $\varepsilon_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$. Sin embargo, se puede extender la prueba con supuestos más débiles sobre la distribución de los errores dado que el supuesto anterior puede ser poco realista.

Anexo 2

Cantidad óptima de factores

Criterios de información

Siguiendo a Krantz y Bagdziunas (2023), en el paquete *dfms* de R, se emplea la función `ICr` que minimiza 3 criterios de información propuestos por Bai y Ng (2002) para determinar la cantidad óptima de factores r^* en modelos de factores aproximados. En este sentido, siguiendo también a Valk et al. (2019), sea $NSSR(r)$ la suma normalizada de los residuos al cuadrado $SSR(r)/(n \times T)$ cuando los r son estimados por componentes principales. Entonces los IC se pueden expresar como

$$\begin{aligned} IC_{r_1} &= \ln(NSSR(r)) + r \frac{N+T}{NT} + \ln\left(\frac{NT}{N+T}\right) \\ IC_{r_2} &= \ln(NSSR(r)) + r \frac{N+T}{NT} + \ln(\min(N, T)) \\ IC_{r_3} &= \ln(NSSR(r)) + r \left(\frac{\ln(\min(N, T))}{\min(N, T)} \right) \end{aligned} \quad (2.1)$$

Anexo 3

Modelo dinámico de factores profundo

Optimización y entrenamiento en Aprendizaje profundo

Siguiendo a Goodfellow et al. (2016, cap.4), optimizar involucra la tarea de minimizar o maximizar determinada función objetivo¹ $f(\mathbf{x})$ cuando se altera \mathbf{x} . Se dice que el valor que minimiza una función es

$$x^* = \operatorname{argmin} f(\mathbf{x})$$

En este contexto de minimizar funciones, se incluye el concepto de derivada $f'(\mathbf{x})$ porque indica en qué medida se debería alterar a \mathbf{x} para obtener pequeñas mejoras en $\mathbf{y} = f(\mathbf{x})$. Se emplea la técnica de descenso por gradiente cuando se reduce $f(\mathbf{x})$ en pequeños pasos de \mathbf{x} con el signo opuesto de la derivada.

Cuando las función de interés cuente con varias entradas $f : \mathbb{R}^n \rightarrow \mathbb{R}$, aparece el concepto de derivadas parciales

$$\frac{\delta f(\mathbf{x})}{\delta x_i}$$

Se mide el cambio en f cuando x_i aumenta en un punto de \mathbf{x} . El concepto de gradiente generaliza la noción de derivada para el caso vectorial. En este sentido, el gradiente de f es el vector que contiene todas las derivadas parciales de $\Delta_{\mathbf{x}} f(\mathbf{x})$.

La derivada direccional en la dirección de un vector unidad \mathbf{u} es la pendiente de la función f en la dirección de \mathbf{u} , o sea, es la derivada de la función $f(\mathbf{x} + \alpha \mathbf{u})$ evaluada en $\mathbf{u}' \Delta_{\mathbf{x}} f(\mathbf{x})$ cuando $\alpha = 0$. Para minimizar a f se busca una dirección para la cual f decrece lo más rápido posible. Se puede realizar lo anterior con la derivada direccional tal que

$$\begin{aligned} \min_{\mathbf{u}, \mathbf{u}'\mathbf{u}=1} \mathbf{u}' \Delta f(\mathbf{x}) \\ \min_{\mathbf{u}, \mathbf{u}'\mathbf{u}=1} \|\mathbf{u}\|_2 \|\Delta_{\mathbf{x}} f(\mathbf{x})\|_2 \cos \theta \end{aligned}$$

donde θ es el ángulo entre \mathbf{u} y el gradiente. Cuando $\|\mathbf{u}\|_2 = 1$ y se ignoren los factores que no dependen de \mathbf{u} , el problema se reduce a $\min_{\mathbf{u}} \cos \theta$. Lo anterior ocurre cuando \mathbf{u} apunta en la dirección opuesta al gradiente. Se llama

¹También llamada de costo o pérdida

a este método como Descenso por gradiente² y se representa como

$$\mathbf{x}' = \mathbf{x} - \delta \Delta \mathbf{x} f(\mathbf{x})$$

donde δ es la tasa de aprendizaje, un escalar positivo, que determina el tamaño del paso.

El Descenso por gradiente estocástico (SGD¹) es una extensión del algoritmo de descenso por gradiente. Un problema recurrente en los problemas de ML se da con conjuntos de entrenamiento grandes: si bien son buenos para generalizar son, contrariamente, computacionalmente intensivos. En general, la función de costo para datos de entrenamiento grandes se puede descomponer en sumas de entre ejemplos de funciones de costo (en inglés *per-example loss function*). En Goodfellow et al. (2016, p.149) se toma el caso de una función de log-verosimilitud para los datos de entrenamiento tal que

$$J(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{x}, y \sim \hat{p}_{data}} L(\mathbf{x}, y, \boldsymbol{\theta}) = \frac{1}{m} \sum_{i=1}^m L(\mathbf{x}^{(i)}, y^{(i)}, \boldsymbol{\theta}) \quad (3.1)$$

donde L es la pérdida *per-example* $L(\mathbf{x}, y, \boldsymbol{\theta}) = -\log p(y|\mathbf{x}; \boldsymbol{\theta})$. El costo computacional de la operación es $O(m)$. A medida que la muestra de entrenamiento crece, el tiempo de cada paso del gradiente aumenta considerablemente. Entonces, el gradiente pasa a ser una expectativa que se puede aproximar con muestras más pequeñas. Se denominan a estas muestras *minibatches* de ejemplos $\mathbb{B} = \{x^{(1)}, \dots, x^{(m')}\}$, con m' fijo, extraídos uniformemente a partir del conjunto de entrenamiento. Entonces, el gradiente estimado se puede calcular como

$$J(\boldsymbol{\theta}) = \frac{1}{m'} \Delta_{\boldsymbol{\theta}} \sum_{i=1}^{m'} L(\mathbf{x}^{(i)}, y^{(i)}, \boldsymbol{\theta}) \quad (3.2)$$

El algoritmo de Descenso por gradiente estocástico (SGD) entonces sigue la trayectoria descendente

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \delta g$$

con δ representando la tasa de aprendizaje.

Según Goodfellow et al. (2016, cap.8), de todos los problemas de optimiza-

²También llamado en inglés Method of steepest descent o gradient descent.

¹Por sus siglas en inglés relativas a Stochastic gradient descent.

ción que se pueden hallar en DL, el más difícil es el relacionado al entrenamiento del modelo. Por esto es que se desarrollaron técnicas para la optimización del mismo. Para esto se pone el foco en encontrar los parámetros $\boldsymbol{\theta}$ de una red neuronal que reducen significativamente una función de costo $J(\boldsymbol{\theta})$, que normalmente incluye una medida de rendimiento evaluada en todo el conjunto de entrenamiento, así como términos de regularización adicionales.

Se puede expresar a la función de costo como la media del conjunto de entrenamiento, tal que

$$J(\boldsymbol{\theta}) = \mathbb{E}_{(\mathbf{x}, y) \sim \hat{p}_{data}} \mathcal{L}(f(\mathbf{x}, \boldsymbol{\theta}), y)$$

donde \mathcal{L} es la función de pérdida, $f(\mathbf{x}, \boldsymbol{\theta})$ es el resultado de predecir cuando las entradas son \mathbf{x} y \hat{p}_{data} es la distribución empírica de los datos. Con esto se define una función objetivo con respecto al conjunto de entrenamiento finito. Por lo general, se busca minimizar la función objetivo cuando la expectativa se toma en relación a la distribución de generación de datos p_{data} ¹, tal que

$$J^*(\boldsymbol{\theta}) = \mathbb{E}_{(\mathbf{x}, y) \sim p_{data}} \mathcal{L}(f(\mathbf{x}, \boldsymbol{\theta}), y) \quad (3.3)$$

El objetivo de un algoritmo de aprendizaje automático es reducir el error de generalización esperado dado por la ecuación (3.3), que se conoce como *riesgo*. Como sólo se cuenta con un conjunto de muestras de entrenamiento, se trata de minimizar un riesgo empírico

$$\mathbb{E}_{\mathbf{x}, y \sim \hat{p}_{data}(\mathbf{x}, y)} [\mathcal{L}(f(\mathbf{x}, \boldsymbol{\theta}), y)] = \frac{1}{m} \sum_{i=1}^m \mathcal{L}(f(\mathbf{x}^{(i)}; \boldsymbol{\theta}), y^{(i)}) \quad (3.4)$$

donde m es la cantidad de muestras de entrenamiento. El proceso de entrenamiento basado en la minimización de este error medio de entrenamiento se conoce como minimización empírica del riesgo. Aunque la minimización empírica del riesgo es propensa al sobreajuste, los modelos que tengan una gran capacidad van a ser capaces de memorizar el conjunto de entrenamiento. Como esta medida de riesgo no es factible en general en la práctica, se propone optimizar mediante el método de descenso por gradiente².

Los algoritmos de optimización para el aprendizaje automático suelen cal-

¹En lugar de sólo a través del conjunto de entrenamiento finito.

²Siempre que se puedan emplear funciones de pérdida útiles, es decir, que se puedan optimizar eficientemente y no sea necesario recurrir a función de pérdida sustitutas

cular cada actualización de los parámetros basándose en un valor esperado de la función de costo estimada utilizando sólo un subconjunto de los términos de la función de costo completa. Por lo que en la práctica, se calculan las expectativas tomando muestras aleatorias de un pequeño número de ejemplos del conjunto de datos y , a continuación, calculando la media sólo de esos ejemplos. Se dice que el tamaño del *batch* refiere al tamaño del *minibatch*. Cuando el tamaño del minibatch sea mayor, se espera que las estimaciones de los gradientes sean más precisas pero que los retornos lineales sean menores.

Back-propagation y regla de la cadena

Siguiendo a Goodfellow et al. (2016, p.200-203), si x es un número real y f y g son dos funciones de mapeo entre números reales hacia otros también reales. Si se asume que $y = g(x)$ and $z = f(g(x)) = f(y)$ entonces la regla de la cadena establece que

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx}$$

Se puede generalizar este concepto más allá del caso escalar, suponiendo que $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$, g mapea de \mathbb{R}^m a \mathbb{R}^n , y f mapea de \mathbb{R}^n a \mathbb{R} . Si $\mathbf{y} = g(\mathbf{x})$ y $z = f(\mathbf{y})$ entonces

$$= \sum_j \frac{\partial z}{\partial y_j} \frac{\partial y_j}{\partial x_j} \quad (3.5)$$

En notación vectorial

$$\Delta_{\mathbf{x}z} = \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right)' \Delta_{\mathbf{y}z}$$

donde $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ es la matriz jacobiana de tamaño $n \times m$ de g .

El algoritmo de propagación hacia atrás consiste en realizar un producto de gradiente jacobiano para cada operación en el grafo. Si se entiende a una red neuronal como un circuito complejo compuesto por muchas capas y unidades de procesamiento llamadas neuronas, durante la fase de entrenamiento, la red aprende a ajustar sus pesos y sesgos para minimizar una función de pérdida que cuantifica la discrepancia entre las predicciones del modelo y los valores reales. La propagación hacia atrás opera en dos fases principales: el pase hacia

adelante y el pase hacia atrás.

Durante el pase hacia adelante, los datos de entrada se propagan a través de la red, pasando por cada capa de neuronas. Cada neurona realiza dos cálculos: calcula su salida basada en las entradas recibidas y también calcula la tasa de cambio local de su salida con respecto a sus entradas. Después de completar el pase hacia adelante y calcular la salida de la red, comienza el pase hacia atrás (*Back-propagation*). Durante este proceso, se calculan los gradientes de la función de pérdida con respecto a los pesos de la red (\mathbf{W}), utilizando la regla de la cadena. La propagación hacia atrás distribuye este gradiente de error desde la salida de la red hacia atrás, capa por capa, ajustando los pesos en la dirección que minimiza la función de pérdida. Es un algoritmo que permite a las redes neuronales aprender de los datos, ajustando sus pesos de manera iterativa para mejorar su rendimiento en tareas específicas (ver Stanford University (2023), Lecun y Soulie Fogelman (1987)).

Adam

Con el método SGD y sus variantes es posible obtener una estimación no sesgada del gradiente tomando el gradiente promedio en minibatches de m ejemplos¹. Un parámetro crucial para el algoritmo SGD es la tasa de aprendizaje. En la práctica, es necesario que la misma disminuya gradualmente con el tiempo. La tasa de aprendizaje puede ser elegida mediante prueba y error o monitoreando las curvas de aprendizaje que representan la función objetivo en relación al tiempo. El SGD puede ser lento por lo que, como alternativa, se pueden emplear métodos de momentos.

El algoritmo de *momentum* acumula decaimientos exponenciales de medias móviles de los gradientes pasados y se mueve en su dirección. En este esquema, se introduce a la variable velocidad v , que representa la dirección y velocidad para las cuales los parámetros se mueven en el espacio paramétrico (ver Goodfellow et al. (2016, cap.8)). En este marco, surgen distintos algoritmos de optimización de tasas de aprendizaje adaptativas. Uno de ellos es el método Adam, empleado en este estudio. En los cursos de Stanford University (2023) se recomienda, en la práctica, emplear este algoritmo como el de defecto para el entrenamiento.

¹Extraídos de forma independiente e idénticamente distribuida de la distribución generadora de datos.

Adam es un método de optimización estocástica utilizado en el entrenamiento de redes neuronales en aprendizaje profundo. Más precisamente, es un algoritmo que optimiza funciones objetivo estocásticas utilizando gradientes de primer orden, y se fundamenta en estimaciones adaptativas de momentos de órdenes inferiores (ver Kingma y Ba (2017)). El método calcula tasas de aprendizaje adaptativas individuales para diferentes parámetros a partir de estimaciones de los primeros y segundos momentos de los gradientes. Es una combinación de dos métodos: *AdaGrad* y *RMSPprop*. La principal idea detrás de Adam es mantener una tasa de aprendizaje adaptativa para cada parámetro.

Como ventajas del método se destaca que las magnitudes de las actualizaciones de los parámetros son invariantes a reescalas del gradiente, sus tamaños de paso están limitados por el hiperparámetro de tamaño de paso, no requiere de un objetivo estacionario, funciona con gradientes dispersos, entre otros.

Figura 3.1: Algoritmo Adam extraído de Kingma y Ba (2017)

Algorithm 1: *Adam*, our proposed algorithm for stochastic optimization. See section 2 for details, and for a slightly more efficient (but less clear) order of computation. g_t^2 indicates the elementwise square $g_t \odot g_t$. Good default settings for the tested machine learning problems are $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. All operations on vectors are element-wise. With β_1^t and β_2^t we denote β_1 and β_2 to the power t .

Require: α : Stepsize

Require: $\beta_1, \beta_2 \in [0, 1)$: Exponential decay rates for the moment estimates

Require: $f(\theta)$: Stochastic objective function with parameters θ

Require: θ_0 : Initial parameter vector

$m_0 \leftarrow 0$ (Initialize 1st moment vector)

$v_0 \leftarrow 0$ (Initialize 2nd moment vector)

$t \leftarrow 0$ (Initialize timestep)

while θ_t not converged **do**

$t \leftarrow t + 1$

$g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$ (Get gradients w.r.t. stochastic objective at timestep t)

$m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ (Update biased first moment estimate)

$v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$ (Update biased second raw moment estimate)

$\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$ (Compute bias-corrected first moment estimate)

$\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$ (Compute bias-corrected second raw moment estimate)

$\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$ (Update parameters)

end while

return θ_t (Resulting parameters)

El algoritmo Adam de la figura 3.1, optimiza funciones objetivo ruidosas utilizando promedios móviles exponenciales del gradiente y su cuadrado, ajustados por hiperparámetros de decaimiento exponencial. Corrige sesgos de inicialización y ofrece eficiencia mediante una actualización mejorada de los parámetros.

Inicialización Xavier

Los parámetros se inicializan empleando la *Inicialización Xavier* para la cual los pesos de las funciones de activación se muestrean a partir de una distribución gaussiana con media cero y varianza $2/(n_{in} + n_{out})$ donde n_{in} es la cantidad de unidades de entrada y n_{out} es la cantidad de unidades de salida (ver Glorot y Bengio (2010) y Andreini et al. (2023a)).

Validación cruzada para la selección de hiperparámetros

Para realizar la validación por sección cruzada en Andreini et al. (2023a) se sigue un enfoque *out-of-sample*. Esto implica dividir al conjunto de observaciones disponibles hasta un cierto punto en el tiempo T , entre una muestra de entrenamiento $[0, T - k \times h - 1]$ y otra de validación $[T - k \times h, T - (k - 1) \times h]$, donde h es el largo del conjunto y $k = K, \dots, 1$ con $K \ll \frac{T-1}{h}$. Entonces, se toma el promedio de las pérdidas en cada conjunto de validación K para alcanzar una pérdida de validación media. Lo anterior implica estimar un modelo dado con hiperparámetros fijos K veces para cada combinación de hiperparámetros. Con un método de búsqueda determinístico el costo computacional es exponencial a la dimensionalidad de los hiperparámetros.

Anexo 4

Cometidos de la política monetaria

Política Monetaria de la Reserva Federal de Estados Unidos

La FED (2018) a través de la política monetaria tiene como objetivo promover de manera efectiva un máximo nivel de empleo, estabilidad de precios, así como también moderar las tasas de interés de largo plazo. En este sentido, se dice que la FED tiene un mandato dual: en la medida que los individuos se encuentren trabajando¹ y que el nivel de precios se mantenga estable y en niveles bajos, van a estar dadas las condiciones para fijar tasas de interés en niveles moderados. Las decisiones de política monetaria se llevan a cabo anualmente, considerando los mencionados dos objetivos, teniendo como instrumento las metas para la FFR. Los canales de transmisión de la política monetaria se dan a través de las condiciones crediticias generales de la economía, esto es, tasas de interés, precios de los activos, tipos de cambio y expectativas de los agentes. Se espera que los resultados de la mencionada política económica afecten la demanda general de bienes y servicios.

Tres principios guían la PM de FED. Primero, la PM tiene que ser comunicada y sistemática. Segundo, el Banco Central debe proporcionar estímulo a la política monetaria cuando la actividad económica esté por debajo del nivel asociado con la utilización plena de recursos y la inflación esté por debajo de su meta. Por el contrario, se debe implementar una política monetaria contractiva cuando la economía esté sobrecalentada y la inflación esté por encima de su objetivo declarado. Un tercer principio es que el Banco Central debe aumentar la tasa de interés de política monetaria, con el tiempo, en más de un punto porcentual en respuesta a un aumento persistente en la inflación y disminuir la tasa de política más de un punto porcentual en respuesta a una disminución persistente en la inflación.

¹O puedan encontrar trabajo de manera rápida.