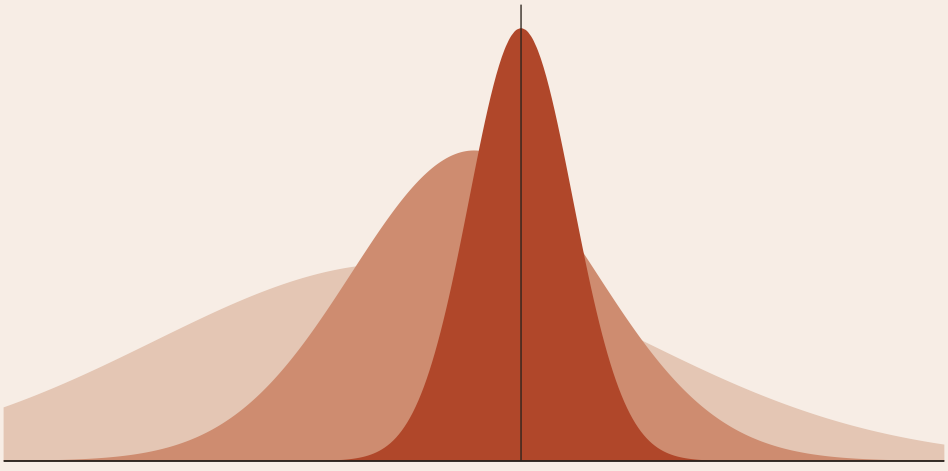
LIBRO ABIERTO DE POSGRADO · EDICIÓN 2026

Macroeconometría aplicada con Python

De los datos macro a la inferencia estructural y el pronóstico



Renato Vassallo

VAR · BVAR · SVAR
espacio de estados · ML

Tabla de contenidos

Presentación

Macroeconometría aplicada con Python es un libro de posgrado sobre los modelos que se usan para leer y pronosticar una economía: vectores autorregresivos, identificación estructural, métodos bayesianos, espacio de estados, volatilidad cambiante y aprendizaje automático. Cada modelo aparece como respuesta a una pregunta empírica concreta, con la derivación, el algoritmo, el código y la aplicación en la misma página.

Está escrito para estudiantes de maestría y de los primeros años de doctorado, y para economistas de bancos centrales y de instituciones de política que necesitan resultados reproducibles y no sólo salidas de software.

Cómo está organizado

1. **Datos macroeconómicos.** Qué mide cada serie y qué le hicimos antes de modelarla.
2. **Modelos univariados.** Memoria, estacionariedad y la referencia difícil de batir.
3. **Vectores autorregresivos.** Dinámica conjunta sin imponer una teoría completa.
4. **VAR estructurales.** Identificación: lo que los datos no pueden decidir por nosotros.
5. **VAR bayesianos.** Encoger para poder estimar, y pronosticar con densidades.
6. **Pronóstico y evaluación.** Qué es un buen pronóstico y cómo se verifica.
7. **Modelos de espacio de estados.** Estimar lo que no se observa.
8. **Tendencias y ciclos.** Brecha del producto y la incertidumbre que rara vez se reporta.
9. **Volatilidad cambiante.** Cuando el tamaño de los choques no es constante.
10. **Parámetros que cambian.** ¿Cambió la economía o cambiaron los choques?
11. **Aprendizaje automático.** Muchos predictores, pocas observaciones.

Los apéndices reúnen la notación del libro y un repaso bayesiano mínimo.

Cómo usarlo

Cada capítulo sigue el mismo recorrido: pregunta económica, modelo, supuestos, estimación, interpretación, implementación en Python, aplicación, diagnósticos, límites y ejercicios. El código aparece en dos versiones cuando conviene: una escrita desde cero con NumPy, que hace visible el algoritmo, y otra con MacroPy, que es la que se usa en la práctica.

Todas las figuras y todos los cuadros se regeneran ejecutando el libro. Para reproducirlo hace falta Quarto y uv:

```
git clone
↪ https://github.com/RenatoVassallo/MacroeconometricsBook.git
cd MacroeconometricsBook
uv sync                          # crea el entorno con Python 3.11
uv run quarto preview            # versión web
uv run quarto render --to pdf    # versión PDF
```

Quien prefiera `pip` puede usar `environment/requirements.txt`, que fija las mismas versiones.

Estado de esta versión

Primera edición **en preparación**, en español. Esta versión se publica mientras se escribe, de modo que los capítulos aparecen a medida que quedan listos y el texto puede cambiar de un día para otro.

Completos y revisados: los capítulos 1 (datos macroeconómicos), 3 (vectores autorregresivos), 4 (VAR estructurales) y 5 (VAR bayesianos), más el apéndice de notación.

En preparación: los capítulos 2 y 6 a 11. De cada uno se publica el esqueleto, con la pregunta que responde y el orden de la exposición, para que se vea el alcance del libro completo.

Los resultados numéricos se verifican contra fuentes publicadas siempre que es posible: el capítulo 4 replica Stock y Watson (2001) y el 5 parte de esa misma muestra. El texto, las cifras y el código están en revisión permanente. Si encuentras un error, el enlace de la barra lateral abre el repositorio y puedes reportarlo como *issue*.

Prefacio

A quién está dirigido

A quien ya tomó un curso de econometría y quiere usar series de tiempo macroeconómicas para responder preguntas de política: estudiantes de maestría y de doctorado, y economistas de bancos centrales, ministerios y organismos internacionales.

Se asume econometría de pregrado, álgebra matricial, probabilidad y algo de Python. No se asume experiencia previa con métodos bayesianos: el Apéndice B reúne lo necesario.

Por qué otro libro

Hay manuales excelentes de teoría y hay repositorios de código. Lo que escasea es el puente: el paso del supuesto de identificación al algoritmo, y del algoritmo a la figura que termina en un informe de política. Este libro ocupa ese espacio, con una deuda explícita con *Applied Bayesian Econometrics for Central Bankers* de Blake y Mumtaz (2017), que sigue siendo la referencia más cercana a este espíritu, aquí en Python, con notación uniforme y algo más de formalidad.

Cómo está escrito cada capítulo

1. Una pregunta económica.
2. La motivación: por qué el modelo anterior no alcanza.
3. El modelo y sus supuestos estadísticos.
4. La estimación y su algoritmo.
5. La interpretación económica.
6. La implementación en Python.
7. Una aplicación empírica.
8. Diagnósticos, límites y extensiones.
9. Ejercicios.

La progresión es siempre la misma: modelo simple, límite del modelo simple, modelo más rico.

Convenciones

Los vectores van en negrita ($\mathbf{y}_t$, $\mathbf{u}_t$) para no confundirlos con el PBI (y_t); las matrices en mayúscula (A_ℓ, Σ). Los bloques marcados **Supuesto** contienen restricciones de identificación: son el lugar donde el modelo deja de hablar de los datos y empieza a hablar de economía. El Apéndice A reúne la notación completa.

El código privilegia la transparencia sobre la eficiencia: primero el algoritmo escrito a mano, luego la versión de biblioteca.

Reproducibilidad

Todo lo que aparece en el libro se regenera con `quarto render`. Los datos crudos no se editan nunca a mano; lo que se descarga vive en `data/raw/` y lo construido en `data/processed/`. Las figuras usan una sola hoja de estilo, `styles/matplotlib.mplstyle`.

Licencia

CAPÍTULO 1

Datos macroeconómicos

LA PREGUNTA ECONÓMICA

¿Qué mide en realidad la serie que estamos a punto de modelar, y qué le hicimos antes de verla?

1. Descomponer una serie en tendencia, ciclo, estacionalidad e irregular, y saber qué parte de esa separación es un supuesto.
2. Desestacionalizar con X-13ARIMA-SEATS, entender qué hace el algoritmo y por qué los bancos centrales usan además TRAMO-SEATS.
3. Elegir la transformación con un criterio, y explicar por qué los economistas trabajan casi siempre en logaritmos.

Requisitos previos. Nociones de series de tiempo y manejo básico de `pandas`.

DATOS UTILIZADOS

Serie	Fuente	Frecuencia	Código
PBI real de EE.UU. (interpolado a mensual)	Uhlig (2005)	mensual	`y`
Deflactor del PBI de EE.UU. (índice)	Uhlig (2005)	mensual	`pi`
PBI del Perú (índice 2007 = 100, desestacionalizado)	BCRP	mensual	`PN01773AM`
IPC de Lima Metropolitana (sin desestacionalizar)	BCRP	mensual	`PN38705PM`

Los archivos crudos están en `data/raw/` y los CSV que leen los capítulos los construyen `code/python/build_us_data.py` y `code/python/build_peru_data.py`.

1.1. Qué mide la serie

Ningún dato macroeconómico se observa directamente. El PBI no se mide: se estima, combinando censos, encuestas, registros administrativos y supuestos de imputación, y se publica como un índice encadenado cuya unidad depende del año base. La

inflación no es el cambio de "los precios" sino el de una canasta fija durante un tiempo, con sustituciones y ajustes de calidad que son decisiones metodológicas. Entre el concepto teórico del modelo y la serie que entra al código hay una cadena de convenciones, y conviene conocerla antes de estimar algo.

Tres consecuencias prácticas atraviesan todo el libro.

La primera es que **el nivel casi nunca es comparable y la tasa casi siempre sí**. Un índice con base 2007 = 100 no dice nada por sí solo; su tasa de crecimiento, en cambio, se compara entre países y entre épocas. Por eso casi todo el trabajo empírico se hace con transformaciones y no con niveles crudos.

La segunda es que **los datos se revisan**. La primera estimación del PBI de un trimestre se publica con semanas de retraso y se corrige durante años. Un ejercicio de pronóstico evaluado con los datos finales de hoy sobreestima lo que un analista podría haber hecho en su momento, porque usa información que entonces no existía. Cuando el objetivo es evaluar pronósticos, hay que usar datos en tiempo real (Croushore y Stark 2001).

La tercera es que **las series traen ruido con estructura**: estacionalidad, efectos de calendario, valores atípicos. Ese ruido no es aleatorio ni pequeño, y si no se trata explícitamente termina interpretándose como economía. Buena parte de este capítulo trata de eso.

Vale un ejemplo concreto, porque aparece en los datos que este libro usa. En Estados Unidos las cuentas nacionales son trimestrales: el PBI mensual no existe como estadística oficial. La serie mensual de Uhlig (2005) que usa este capítulo es una interpolación del PBI trimestral con indicadores mensuales, de modo que su variación de un mes al siguiente es en parte producto del método de interpolación y no de la economía. Eso no la invalida (es la única forma de poner producto y tasa de interés en un mismo VAR mensual), pero obliga a no leer un movimiento mensual aislado como si fuera un dato. Los capítulos Capítulo 4 y Capítulo 5 evitan el problema trabajando con las muestras trimestrales de Stock y Watson (2001) y Blanchard y Quah (1989).

1.2. Descarga reproducible

La regla del libro es simple: `data/raw/` es inmutable y todo lo demás se reconstruye con un script. Ninguna serie se edita a mano, y ningún resultado depende de un archivo que sólo existe en la computadora de quien escribe.

`MacroPy` trae funciones de descarga para las dos fuentes que usa el libro, la API del BCRP y FRED:

```
from MacroPy import get_bcrp_data, get_fred_data

series = {"PN01773AM": "gdp_sa",  # PBI desestacionalizado
          "PN38705PM": "cpi",     # IPC de Lima, sin ajustar
          "PN01210PM": "fx",      # tipo de cambio bancario
          "PD04722MM": "mpr"}     # tasa de referencia
peru = get_bcrp_data(series, frequency="M",
                     start_period="1996-1")

usa = get_fred_data(["GDPC1", "GDPDEF"], ["gdp", "deflator"],
                    frequency="q", api_key=API_KEY,
                    start_period="1960-1")
```

Esa celda no se ejecuta al compilar el libro, a propósito. Los archivos que produce están guardados en `data/raw/`, de modo que el libro se compila sin conexión y los resultados no cambian cuando la fuente revisa el pasado. Para actualizar, se corre la descarga y se vuelve a compilar: la diferencia entre las dos versiones es exactamente la revisión de los datos.

```
import numpy as np
import pandas as pd

from macrobook import data_path

usa = pd.read_csv(data_path("uhlig2005.csv"), index_col="date")
usa.index = pd.PeriodIndex(usa.index, freq="M")
peru = pd.read_csv(data_path("peru_monthly.csv"),
                   index_col="date")
peru.index = pd.PeriodIndex(peru.index, freq="M")

gdp_us = np.exp(usa["y"])          # el archivo trae logaritmos
price_us = np.exp(usa["pi"])
gdp_pe = peru["gdp_sa"].dropna()
cpi_pe = peru["cpi"].dropna()

print("EE.UU.:", usa.index[0], "a", usa.index[-1])
print("Perú:  ", cpi_pe.index[0], "a", cpi_pe.index[-1])
```

```
EE.UU.: 1965-01 a 2003-12
Perú:   1996-01 a 2026-08
```

1.3. Cuatro series y sus hechos estilizados

```
import matplotlib.pyplot as plt

from macrobook import use_style, figsize, PALETTE, charts

use_style()

panels = [(gdp_us, "PBI real, EE.UU."),
          (price_us, "deflactor, EE.UU."),
          (gdp_pe, "PBI, Perú"),
          (cpi_pe, "IPC, Perú")]

size = figsize(0.62)
fig, axes = plt.subplots(2, 2, figsize=size)
for ax, (series, title) in zip(axes.ravel(), panels):
    ax.plot(series.index.to_timestamp(how="end"), series,
            color=PALETTE["ink"], lw=1.0)
    ax.set_yscale("log")
    charts.plain_log_ticks(ax)
    charts.year_ticks(ax, every=10)
    ax.set_title(title, fontsize=8.5)
plt.show()
```

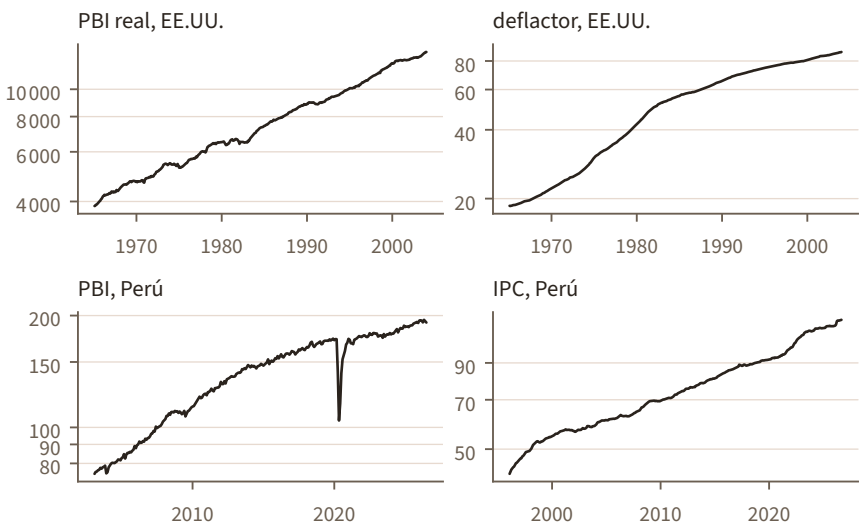


Figura 1.1. Las cuatro series del capítulo. Arriba, Estados Unidos entre 1965 y 2003; abajo, Perú entre 1996 y 2026. La escala vertical es logarítmica en las cuatro.

Las cuatro series comparten una propiedad que decide casi todo lo que sigue: crecen de forma multiplicativa, no aditiva. En escala logarítmica las cuatro se ven como rectas con quiebres, y esa es la forma en que conviene mirarlas.

```
def summary(series, freq=12):
    growth = (series / series.shift(freq) - 1) * 100
    return {"media": growth.mean(), "desv.": growth.std(),
            "autocorr.": growth.autocorr(1),
            "mín": growth.min(), "máx": growth.max()}

table = pd.DataFrame({
    "PBI EE.UU.": summary(gdp_us),
    "Precios EE.UU.": summary(price_us),
    "PBI Perú": summary(gdp_pe),
    "Precios Perú": summary(cpi_pe)}).T
digits = {"media": 1, "desv.": 1, "autocorr.": 2,
          "mín": 1, "máx": 1}
table.round(digits)
```

	media	desv.	autocorr.	mín	máx
PBI EE.UU.	3.3	2.3	0.95	-3.3	8.9
Precios EE.UU.	4.1	2.4	0.99	0.9	11.0
PBI Perú	4.4	6.9	0.81	-39.0	60.5
Precios Perú	3.4	2.2	0.98	-1.1	11.0

El cuadro resume las tasas interanuales y contiene cuatro hechos que reaparecen en todo el libro.

El crecimiento es persistente pero acotado. La autocorrelación de primer orden está entre 0.8 y 0.95. Parte de esa persistencia es mecánica: una tasa interanual solapa doce meses de información y sería autocorrelacionada aun si el crecimiento mensual fuera independiente. Lo que no es mecánico es que las series oscilen alrededor de una media positiva sin alejarse de ella de forma permanente. Un modelo razonable de la tasa de crecimiento es estacionario y persistente, no una caminata aleatoria.

La inflación es todavía más persistente. Con autocorrelaciones de 0.98 y 0.99, la inflación se comporta casi como si tuviera una raíz unitaria, y esa persistencia es el motivo de que los VAR del Capítulo 3 funcionen bien en niveles de la tasa.

La volatilidad no es comparable entre países, pero el cuadro exagera la diferencia. El PBI peruano tiene una desviación estándar tres veces mayor que el estadounidense, y sus extremos son de −39 y +60 por ciento. Esos dos números son el mismo episodio: el confinamiento de 2020 y el rebote de 2021. Si se quitan esos dos años, la desviación estándar peruana baja de 6.9 a 2.9, apenas por encima de la estadounidense. Un

momento muestral que se mueve así al sacar veinticuatro observaciones no describe la economía: describe un episodio. Volveremos a él en el Sección 1.7, porque un dato así rompe cualquier modelo lineal estimado sin cuidado.

Producto e inflación no se mueven juntos de manera estable. La correlación entre ambas tasas es −0.34 en Estados Unidos y −0.02 en Perú. No hay una curva de Phillips visible en el dato crudo, y por eso hace falta un modelo estructural como el del Capítulo 4 para decir algo sobre la relación.

1.4. La descomposición de una serie

La tradición que arranca con Burns y Mitchell (1946) mira una serie macroeconómica como la suma de componentes no observados con interpretación distinta.

DEFINICIÓN 1.1 · DESCOMPOSICIÓN POR COMPONENTES NO OBSERVADOS

En logaritmos, una serie se escribe como

$$\ln y_t = \tau_t + c_t + s_t + \eta_t, \tag{1.1}$$

donde τ_t es la tendencia, c_t el ciclo, s_t el componente estacional y η_t el irregular. En niveles, la misma relación es multiplicativa: $y_t = T_t \cdot C_t \cdot S_t \cdot I_t$.

Trabajar en logaritmos convierte la descomposición multiplicativa en aditiva, y esa es la primera razón práctica por la que los economistas usan logaritmos: la amplitud de la estacionalidad de una serie que crece suele ser proporcional al nivel, no fija en unidades.

CUIDADO

La Ecuación 1.1 no se puede estimar sin supuestos adicionales: cuatro componentes no observados y una sola serie observada. Todo método de descomposición añade restricciones, casi siempre sobre las frecuencias que ocupa cada componente o sobre la forma de su dinámica. Dos métodos razonables dan resultados distintos, y ninguno es "el" verdadero. Lo que se puede exigir es que el supuesto sea explícito y que el resultado se reporte junto con él.

La separación no es igual de difícil en todos los componentes. La estacionalidad ocupa frecuencias conocidas, se repite y es la parte mejor identificada del problema; separar tendencia de ciclo, en cambio, es una decisión casi puramente teórica, con una literatura entera en desacuerdo: el filtro de Hodrick y Prescott (1997), la crítica de Hamilton (2018), la descomposición de Beveridge y Nelson (1981) y los modelos de componentes no observados de Harvey (1989) producen ciclos distintos con los mismos datos. El Capítulo 8 trata ese problema. Este capítulo se ocupa del componente estacional, donde el acuerdo es mucho mayor.

Vale la pena ver el problema con una serie inventada, donde sí conocemos la respuesta.

```
rng = np.random.default_rng(7)
T = 240
index = pd.period_range("2006-01", periods=T, freq="M")

trend = np.log(100) + 0.0025 * np.arange(T)
cycle = np.zeros(T)
for t in range(2, T):
    cycle[t] = (1.5 * cycle[t - 1] - 0.6 * cycle[t - 2]
                + rng.normal(0, 0.004))
pattern = np.array([-1.2, -1.8, 2.8, 0.6, 0.2, -0.4,
                    0.9, 0.4, -0.3, 0.1, -0.8, -0.5]) / 100
pattern = pattern - pattern.mean()
season = pattern[np.arange(T) % 12]
irregular = rng.normal(0, 0.004, T)

log_y = trend + cycle + season + irregular
y_sim = pd.Series(np.exp(log_y), index=index.to_timestamp())

size = figsize(0.52)
fig, axes = plt.subplots(2, 2, sharex=True, figsize=size)
pieces = [(log_y, "serie (en logaritmos)"),
          (trend + cycle, "tendencia y ciclo"),
          (season, "estacional"),
          (irregular, "irregular")]
for ax, (values, title) in zip(axes.ravel(), pieces):
    ax.plot(y_sim.index, values, color=PALETTE["ink"], lw=0.9)
    ax.set_title(title, fontsize=8.5)
plt.show()
```

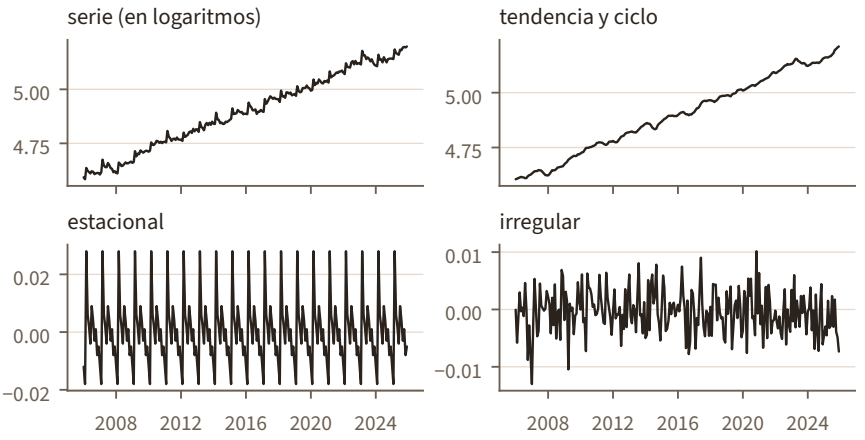


Figura 1.2. Una serie simulada y sus cuatro componentes. La tendencia y el ciclo son suaves; la estacionalidad se repite; el irregular no tiene estructura.

1.5. Desestacionalizar

La estacionalidad es un fenómeno económico real: en diciembre se produce y se compra más, en marzo suben las matrículas escolares, la pesca depende de las vedas. No es un error de medición. El problema es que domina la variación de corto plazo y esconde justamente lo que interesa: si la economía se está acelerando o desacelerando ahora.

Hay dos maneras de lidiar con eso, y sólo una funciona bien.

La primera es comparar contra el mismo mes del año anterior. Es gratis, es la que usan los titulares y tiene dos defectos serios: promedia los últimos doce meses, de modo que llega tarde a los puntos de quiebre, y arrastra lo que pasó hace un año, el llamado **efecto base**. La segunda es estimar el componente estacional y removerlo, que es lo que hacen las oficinas de estadística.

```
yoy = (gdp_pe / gdp_pe.shift(12) - 1) * 100
monthly = ((gdp_pe / gdp_pe.shift(1)) ** 12 - 1) * 100
window = slice("2019-01", "2022-12")
dates_pe = gdp_pe[window].index.to_timestamp(how="end")
mark = pd.Period("2021-04", freq="M").to_timestamp(how="end")

size = figsize(0.40)
fig, axes = plt.subplots(1, 2, figsize=size)
axes[0].plot(dates_pe, gdp_pe[window], color=PALETTE["ink"],
             lw=1.2)
axes[0].axhline(gdp_pe["2019"].mean(), color=PALETTE["muted"],
```

```
                lw=0.8, ls=(0, (2, 2)))
axes[0].plot([mark], [gdp_pe["2021-04"]], "o", ms=4,
             color=PALETTE["accent"])
axes[0].set_title("nivel (2007 = 100)", fontsize=8.5)
axes[1].plot(dates_pe, yoy[window], color=PALETTE["ink"],
             lw=1.2, label="interanual")
axes[1].plot(dates_pe, monthly[window], color=PALETTE["accent"],
             lw=1.2, ls=(0, (4, 2)),
             label="mensual anualizada")
axes[1].plot([mark], [yoy["2021-04"]], "o", ms=4,
             color=PALETTE["accent"])
axes[1].set_ylim(-60, 75)
charts.zero_line(axes[1])
for ax in axes:
    charts.year_ticks(ax, every=2)
axes[1].set_title("%", fontsize=8.5)
charts.legend_outside(axes[1], ncol=2)
plt.show()
```

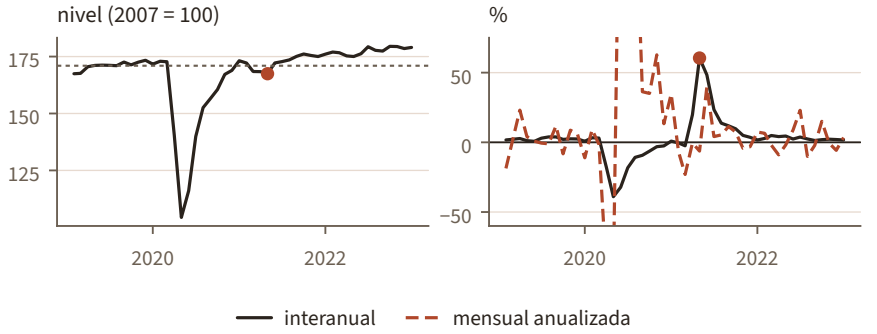


Figura 1.3. Perú: el nivel del PBI desestacionalizado (izquierda) y dos lecturas de su crecimiento (derecha). El punto marca abril de 2021, cuando la tasa interanual llegó a 60 por ciento con el nivel todavía por debajo del promedio de 2019 (línea punteada). El eje derecho está recortado: la tasa mensual anualizada baja a -97 y sube a +842 por ciento en 2020.

En abril de 2021 el PBI peruano creció 60.5 % interanual y el titular decía que la economía se expandía a un ritmo histórico. El punto rojo del panel izquierdo muestra el mismo mes en niveles: el PBI estaba 2.1 % por debajo del de abril de 2019 y algo por debajo del de marzo de 2021. La tasa mensual anualizada de la serie desestacionalizada era de −6.3 % ese mes, porque la segunda ola de contagios estaba frenando la actividad. Las dos cifras son correctas y miden cosas distintas: la interanual compara contra abril de 2020, el mes del confinamiento más estricto, y por lo tanto habla sobre todo del pasado. Eso es un efecto base.

1.5.1. X-13ARIMA-SEATS

El estándar internacional es el programa de la Oficina del Censo de Estados Unidos, heredero del X-11 de los años sesenta (Dagum 1980) y del X-12-ARIMA (Findley et al. 1998). El algoritmo tiene dos etapas.

En la **primera etapa**, llamada regARIMA, se estima un modelo de la serie con regresores deterministas y un ARIMA estacional para el residuo:

$$\ln y_t = \beta' x_t + z_t, \qquad \phi(L)\Phi(L^{12})(1-L)^d(1-L^{12})^D z_t = \theta(L)\Theta(L^{12})a_t. \tag{1.2}$$

Los regresores x_t recogen efectos de calendario y valores atípicos: días hábiles del mes, año bisiesto, Semana Santa, feriados nacionales, cambios de nivel, valores extremos. Con el modelo estimado, la serie se limpia de esos efectos y se extiende hacia atrás y hacia adelante con pronósticos, para que los promedios móviles de la etapa siguiente no se rompan en los extremos de la muestra.

En la **segunda etapa** se separan los componentes, y aquí el programa ofrece dos caminos. El filtro X-11 aplica una secuencia iterativa de promedios móviles: estima una tendencia, obtiene el componente estacional como promedio móvil de los residuos por mes, lo remueve, vuelve a estimar la tendencia y repite hasta converger. El camino SEATS, en cambio, es de extracción de señal: a partir del ARIMA estimado deriva los modelos de cada componente y calcula el filtro óptimo en el sentido de error cuadrático medio mínimo.

> INTUICIÓN
>
> X-11 es un filtro: un conjunto de promedios móviles afinados durante décadas de práctica. SEATS es un modelo: dado un ARIMA, deduce cuál es la mejor separación posible. El primero es robusto y transparente; el segundo entrega además medidas de incertidumbre de los componentes, que el primero no tiene.

`MacroPy` trae el binario `x13as` y lo expone en una sola función.

```
from MacroPy import seasonal_adjust

result = seasonal_adjust(y_sim, method="x13", outlier=True)
estimated = np.log(y_sim / result.seasadj) * 100

comparison = pd.DataFrame(
    {"simulado": pattern * 100,
     "X-13": estimated.groupby(estimated.index.month).mean()},
    index=range(1, 13))
comparison.round(2).T
```

	1	2	3	4	5	6	7	8	9	10	11	12
simulado	-1.20	-1.80	2.80	0.60	0.20	-0.40	0.90	0.40	-0.30	0.10	-0.80	-0.50
X-13	-1.24	-1.87	2.97	0.52	0.27	-0.26	0.87	0.52	-0.43	0.06	-0.71	-0.78

Sobre la serie simulada, donde conocemos la respuesta, X-13 recupera el patrón estacional con un error medio de una décima de punto porcentual y una correlación de 0.995 con el patrón verdadero. La serie desestacionalizada que devuelve correlaciona 0.9999 con la serie sin estacionalidad que se simuló. Este es el sentido en que la separación estacional es un problema bien planteado, a diferencia de la separación entre tendencia y ciclo.

```
adjusted = seasonal_adjust(cpi_pe.to_timestamp(how="end"),
                           method="x13", outlier=True)
factors = np.log(adjusted.observed / adjusted.seasadj) * 100
by_month = factors.groupby(factors.index.month).mean()

size = figsize(0.40)
fig, axes = plt.subplots(1, 2, figsize=size,
                         gridspec_kw={"width_ratios": [1.4, 1]})
for month in range(1, 13):
    line = factors[factors.index.month == month]
    axes[0].plot(line.index.year, line.values, lw=0.7,
                 color=PALETTE["accent"] if month == 3
                 else PALETTE["observed"][2],
                 zorder=3 if month == 3 else 1)
axes[0].annotate("marzo", (2006, 0.44), fontsize=8,
                 color=PALETTE["accent"])
charts.zero_line(axes[0])
axes[0].set_title("factor por mes, año a año (%)", fontsize=8.5)
axes[1].bar(by_month.index, by_month.values,
            color=PALETTE["steel"], width=0.7)
charts.zero_line(axes[1])
axes[1].set_xticks(range(1, 13, 2))
axes[1].set_title("promedio por mes (%)", fontsize=8.5)
plt.show()
```

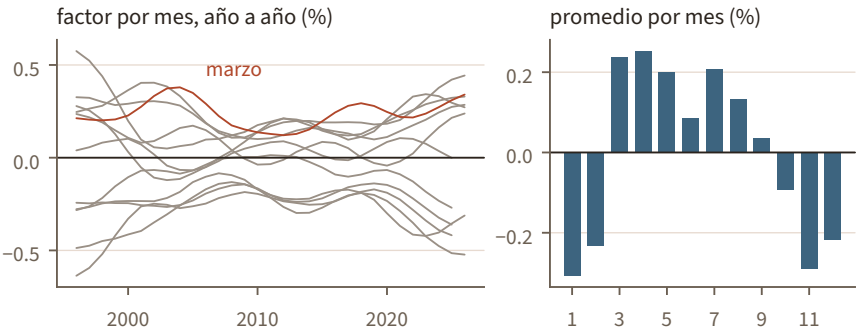


Figura 1.4. IPC del Perú, factor estacional estimado por X-13. A la izquierda, el factor de cada mes año por año, con marzo resaltado; a la derecha, el promedio por mes de toda la muestra. El patrón se repite, pero no es fijo.

El patrón estimado para el Perú tiene sentido económico: marzo, abril y mayo por encima de la tendencia, por la temporada escolar y el fin de la campaña agrícola, y noviembre, diciembre y enero por debajo. La amplitud es pequeña, algo más de medio punto entre el mes más alto y el más bajo, y aun así basta para que la inflación mensual de marzo parezca alta todos los años si no se corrige.

El panel izquierdo muestra la propiedad que separa a X-13 de una regresión con dummies: el factor de marzo no es una constante. Sube de 0.21 puntos en 1996 a 0.38 en 2004, cae a 0.12 en 2012 y vuelve a 0.34 al final de la muestra, un rango de tres a uno. El factor de agosto es 0.18 puntos más alto en la segunda mitad de la muestra que en la primera. Un patrón fijo estimado sobre treinta años no describe bien ninguna de las dos décadas.

1.5.2. Otros métodos

STL (Cleveland et al. 1990) descompone con regresiones locales ponderadas en lugar de promedios móviles. Es robusto a atípicos, no impone estacionalidad fija y está disponible en `statsmodels`; `MacroPy` lo expone con `method="stl"`. No maneja efectos de calendario, que es su principal limitación para datos oficiales.

Variables dummy estacionales en una regresión son la opción más simple y la peor: imponen un patrón fijo para toda la muestra, cuando la estacionalidad cambia lentamente con la estructura productiva.

Modelos de componentes no observados en forma de espacio de estados estiman la estacionalidad junto con la tendencia y el ciclo dentro de un mismo modelo, con el filtro de Kalman (Harvey 1989). Es la opción natural cuando la desestacionalización es parte del modelo y no un paso previo, y es lo que hace el Capítulo 7.

CUIDADO

Usar la variación interanual como si fuera una desestacionalización es un error frecuente. La diferencia de doce meses elimina el patrón estacional, sí, pero también introduce una media móvil que distorsiona la dinámica: crea autocorrelación artificial, desplaza los puntos de quiebre y arrastra los efectos base. Para describir la coyuntura, desestacionalizar; para el titular, la interanual.

1.5.3. Por qué los bancos centrales usan TRAMO-SEATS

TRAMO-SEATS es un software independiente desarrollado en el Banco de España por Gómez y Maravall (1996). TRAMO se ocupa de la primera etapa, que en su nombre original es *Time series Regression with ARIMA noise, Missing values and Outliers*, y SEATS de la extracción de señal.

Conviene despejar primero una confusión de nombres. X-13ARIMA-SEATS incorporó el motor SEATS, de modo que la elección entre los dos programas no es "filtro contra modelo": con X-13 se puede usar SEATS. Lo que distingue a TRAMO-SEATS es su etapa de preajuste y el ecosistema institucional construido a su alrededor. Con eso dicho, que el Banco Central Europeo, Eurostat y buena parte de los bancos centrales y ministerios lo sigan usando tiene tres razones concretas.

La primera es la **flexibilidad del calendario**. Los efectos de días hábiles y feriados son específicos de cada país: Semana Santa se mueve entre marzo y abril, el Perú tiene Fiestas Patrias en julio y feriados largos que cambian de fecha, y el efecto del año bisiesto depende de la estructura productiva. TRAMO permite definir regresores de calendario a medida, con identificación automática del ARIMA y de los atípicos, y contrastar su significancia. Eso es exactamente lo que hace falta cuando la serie es de un país que el programa no trae precalibrado, y es lo que las directrices europeas piden: regresores de calendario contrastados uno por uno, no un paquete fijo.

La segunda es que **entrega incertidumbre, no sólo una cifra**. Al derivar los componentes del ARIMA estimado, SEATS produce errores estándar de la serie desestacionalizada y diagnósticos formales sobre si la descomposición es admisible, es decir, si existen componentes con espectro no negativo compatibles con el modelo. Para una institución que publica una cifra oficial y tiene que defenderla, esa auditoría importa.

La tercera es **institucional**. Las directrices europeas de ajuste estacional (Eurostat 2024) no nombran programas: aceptan dos familias de métodos, la extracción de señal a partir de un modelo ARIMA y el método semiparamétrico de promedios móviles predefinidos, y recomiendan software libre publicado por institutos de estadística. Las dos familias son precisamente SEATS y X-11, y la plataforma JDemetra+ las implementa juntas. Una vez que un país documenta su metodología y forma a su equipo, el costo de cambiar es alto y el beneficio pequeño: en series bien comportadas las dos familias dan resultados muy parecidos.

1.6. Transformaciones

Elegir la transformación es elegir qué pregunta se puede responder después. Las opciones usuales son pocas.

Tabla 1.1. Transformaciones habituales y qué conserva cada una

Transformación	Fórmula	Qué preserva
Nivel	y_t	todo, incluida la tendencia
Logaritmo	$\ln y_t$	la tendencia, en escala proporcional
Diferencia logarítmica	$\Delta \ln y_t$	la dinámica de corto plazo
Tasa interanual	$y_t / y_{t-s} - 1$	el movimiento de baja frecuencia
Diferencia de orden s	$\Delta_s \ln y_t$	la dinámica sin estacionalidad

La primera decisión, niveles frente a diferencias, se discutió en el Sección 3.10: la respuesta depende de si interesan las relaciones de largo plazo y de si las series comparten tendencias estocásticas (Nelson y Plosser 1982). La segunda decisión, que es la de esta sección, es si trabajar con la variable o con su logaritmo.

1.6.1. Por qué logaritmos

Hay cuatro razones, y todas se ven en los datos.

Primera: linealiza el crecimiento exponencial. Una serie que crece a tasa constante es una exponencial en niveles y una recta en logaritmos. Los ojos juzgan mal las exponenciales y bien las rectas, de modo que un cambio de tendencia se ve en el gráfico logarítmico y se esconde en el de niveles.

```
steps = np.arange(len(gdp_us))
slope, intercept = np.polyfit(steps, np.log(gdp_us), 1)
fitted = intercept + slope * steps
dates_us = gdp_us.index.to_timestamp(how="end")

size = figsize(0.36)
fig, axes = plt.subplots(1, 2, figsize=size)
axes[0].plot(dates_us, gdp_us, color=PALETTE["ink"], lw=1.1)
axes[0].plot(dates_us, np.exp(fitted), color=PALETTE["accent"],
             lw=1.0, ls=(0, (4, 2)))
axes[0].set_title("nivel del índice", fontsize=8.5)
axes[1].plot(dates_us, np.log(gdp_us), color=PALETTE["ink"],
```

```
            lw=1.1)
axes[1].plot(dates_us, fitted, color=PALETTE["accent"], lw=1.0,
             ls=(0, (4, 2)))
axes[1].set_title("logaritmo", fontsize=8.5)
for ax in axes:
    charts.year_ticks(ax, every=10)
plt.show()
```

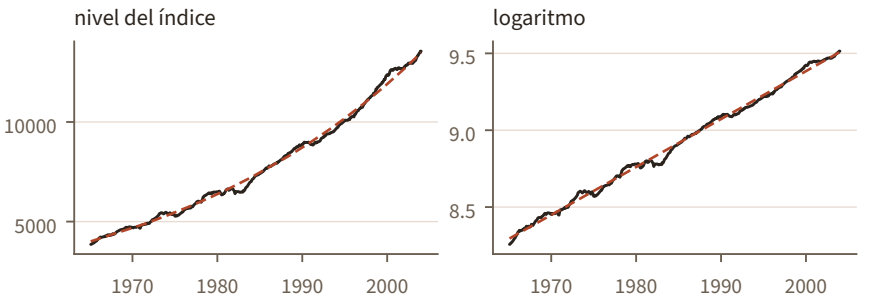


Figura 1.5. PBI real de Estados Unidos en niveles y en logaritmos. La recta punteada es la tendencia ajustada a los logaritmos; en niveles, la misma tendencia se curva.

Segunda: la diferencia logarítmica es casi la tasa de crecimiento, y es más manejable. Con $g_t = y_t / y_{t-1} - 1$,

$$\Delta \ln y_t = \ln(1 + g_t) \approx g_t - \frac{g_t^2}{2} + \frac{g_t^3}{3} - \cdots \tag{1.3}$$

La aproximación es excelente para tasas pequeñas y se deteriora rápido para tasas grandes.

```
grid = np.linspace(0.001, 0.6, 200)
error = (np.log1p(grid) - grid) * 100

size = figsize(0.38)
fig, ax = plt.subplots(figsize=size)
ax.plot(grid * 100, error, color=PALETTE["ink"], lw=1.2)
for point in (0.05, 0.10, 0.25):
    ax.plot(point * 100, (np.log1p(point) - point) * 100,
            marker="o", ms=4, color=PALETTE["accent"])
charts.zero_line(ax)
ax.set_xlabel("tasa de crecimiento (%)")
ax.set_ylabel("error en puntos porcentuales")
plt.show()
```

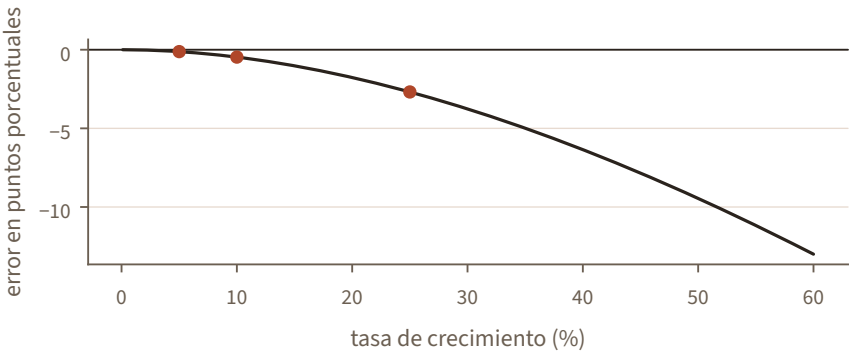


Figura 1.6. Error de aproximación de la diferencia logarítmica frente a la tasa de crecimiento. Los puntos marcan 5, 10 y 25 por ciento. Hasta 10 por ciento el error es menor a medio punto porcentual; en el rebote peruano de 2021, de 60 por ciento, el error llega a trece puntos.

Con crecimiento trimestral de 1 %, la diferencia entre las dos medidas es de medio punto básico y da igual cuál se use. Con el 60 % interanual del rebote peruano de 2021, la diferencia logarítmica da 47 % y la tasa ordinaria 60 %: ahí sí importa, y hay que decir cuál se está reportando.

Tercera: las diferencias logarítmicas se suman. El crecimiento acumulado de un año es la suma de los doce crecimientos logarítmicos mensuales, mientras que con tasas ordinarias hay que multiplicar factores. Lo mismo vale para descomponer un agregado o para anualizar: $(1 + g)^{12} - 1$ se vuelve $12\,\Delta \ln y_t$.

Cuarta: suele estabilizar la varianza y simetrizar la distribución. Cuando la amplitud de las fluctuaciones crece con el nivel, el logaritmo produce una serie de cambios con varianza aproximadamente constante. Es el caso $\lambda = 0$ de la familia de Box y Cox (1964), y es también lo que convierte una estacionalidad multiplicativa en aditiva.

El IPC peruano es el caso de manual. Entre 1996 y 2026 el índice pasa de 42 a 121, y la desviación estándar de su variación mensual medida en puntos del índice sube 76 % entre la primera y la segunda mitad de la muestra; medida en diferencias logarítmicas baja 7 %, es decir, es estable. La asimetría de los cambios cae de 2.3 a 1.0 al pasar a logaritmos.

Conviene no convertir eso en una ley. En el PBI mensual de Estados Unidos ocurre lo contrario: la desviación estándar de la primera diferencia en niveles es casi la misma en 1965-1984 y en 1985-2003, mientras que la de la diferencia logarítmica cae 41 %, porque la Gran Moderación redujo la volatilidad relativa más rápido de lo que creció el nivel. La estabilización de varianza es una hipótesis contrastable sobre cada serie, no una propiedad automática del logaritmo, y la forma de contrastarla es exactamente la de arriba: comparar la dispersión de los cambios entre submuestras en una y otra escala.

La contracara es que la transformación no es neutral para pronosticar: revertirla con la exponencial de la media no devuelve la media del nivel, sino la mediana, y la corrección importa cuando la varianza es grande (Lütkepohl y Xu 2012).

CUIDADO

Dos límites. El logaritmo exige variables positivas, y eso descarta más series de las que parece: saldos fiscales, cuenta corriente, brechas y cualquier tasa que pueda ser negativa. Forzar un logaritmo sumando una constante arbitraria cambia la interpretación de los coeficientes y no resuelve nada. Y la simetría se gana sobre el nivel, no sobre una tasa: aplicar el logaritmo a una variación interanual que ya es un cociente no la vuelve más simétrica. En el PBI peruano sólo cambia el signo de la asimetría, de +1.3 a −1.6.

1.6.2. Qué transformación usa cada quien

Conviene mirar qué hacen los artículos que el libro replica, porque la elección no es arbitraria.

Tabla 1.2. Transformaciones en los artículos que el libro replica

Artículo	Variables	Transformación
Sims (1980)	producto, precios, dinero, tasas	logaritmos en niveles
Blanchard y Quah (1989)	PBI, desempleo	tasa de crecimiento, desempleo sin tendencia
Stock y Watson (2001)	inflación, desempleo, tasa	$400\,\Delta \ln P_t$; tasas en niveles
Uhlig (2005)	PBI, deflactor, reservas, tasa	100 ln en niveles; tasa en nivel
Bańbura et al. (2010)	131 series	logaritmos en niveles, con prior de Minnesota

El patrón es claro. Cuando el objetivo es estructural, es decir estimar respuestas a choques, la práctica dominante es trabajar en logaritmos de los niveles y dejar que los rezagos capturen la persistencia, tal como argumentan Sims et al. (1990). Cuando el objetivo es pronosticar y las series tienen raíces cercanas a uno, se diferencia, o se encoge hacia una caminata aleatoria con el prior del Capítulo 5, que es una forma más suave de hacer lo mismo. Las tasas de interés y de desempleo se dejan en niveles siempre, porque ya están acotadas y su logaritmo no tiene interpretación útil.

INTUICIÓN

La regla operativa cabe en tres líneas. Si la variable es un índice o un valor que crece de forma multiplicativa, logaritmo. Si es una tasa o un porcentaje, nivel. Si el modelo va a leerse como elasticidades o como respuestas porcentuales a un choque, logaritmo otra vez, porque es lo que hace que los coeficientes tengan esa lectura.

1.7. Otros temas

1.7.1. Revisiones y datos en tiempo real

Las bases de datos de *vintages*, como ALFRED de la Reserva Federal de San Luis, guardan lo que se publicó en cada fecha. Un ejercicio de pronóstico honesto usa, en cada momento, sólo lo que estaba disponible entonces (Croushore y Stark 2001). La diferencia no es menor: las primeras estimaciones del PBI se revisan en varias décimas, y con series muy revisadas la evaluación con datos finales exagera la calidad del pronóstico.

1.7.2. Valores atípicos y la pandemia

Los dos extremos del PBI peruano del cuadro de momentos, −39 y +60 por ciento, son el mismo episodio. Un modelo lineal con errores gaussianos estimado sobre esa muestra interpreta esos meses como evidencia sobre la dinámica normal del sistema, y el resultado es una varianza sobreestimada y unos coeficientes deformados. Las opciones son tratarlos como atípicos con variables dummy, escalar la volatilidad de esos meses (Lenza y Primiceri 2022), excluir el tramo o modelar la volatilidad cambiante como en el Capítulo 9. Ninguna es gratis, y lo mínimo es decir cuál se usó.

1.7.3. Frecuencias y agregación

Pasar de mensual a trimestral parece inocuo y no lo es. Un promedio de tres meses no es lo mismo que el dato de fin de trimestre: el promedio suaviza y crea autocorrelación, el fin de periodo conserva el ruido. Para variables de flujo, como el PBI, corresponde promediar o sumar; para variables de saldo, como una tasa de interés o el tipo de cambio, el promedio y el fin de periodo responden preguntas distintas y la elección debe declararse. Los capítulos siguientes trabajan con series ya agregadas por la fuente y no re-agregan nada, precisamente para no añadir una decisión propia a las que la fuente ya tomó.

Ideas clave

1. Una serie macro es el resultado de una cadena de convenciones de medición. Conocerla es parte del trabajo empírico, no un prólogo.
2. La descomposición en tendencia, ciclo, estacionalidad e irregular necesita supuestos: la parte estacional está bien identificada, la separación entre tendencia y ciclo no.
3. Desestacionalizar no es un lujo: la variación interanual llega tarde a los quiebres y arrastra efectos base, como muestra el Perú de 2021.
4. X-13ARIMA-SEATS combina una etapa de regresión con ARIMA, que limpia calendario y atípicos, con una de extracción de componentes por filtros o por modelo.
5. El logaritmo se usa porque linealiza el crecimiento, aproxima la tasa y se suma; que además estabilice la varianza es una hipótesis que hay que contrastar serie por serie. Deja de usarse cuando la variable puede ser negativa o ya es una tasa.

Lecturas recomendadas

Ghysels y Osborn (2001) el tratamiento econométrico completo de la estacionalidad.

Findley et al. (1998) la descripción oficial del programa X-12-ARIMA, base del X-13 actual.

Gómez y Maravall (1996) el manual de TRAMO-SEATS, con la lógica de extracción de señal basada en modelos.

Eurostat (2024) las directrices europeas de ajuste estacional; el documento que siguen las oficinas de estadística.

Lütkepohl y Xu (2012) qué gana y qué pierde el pronóstico al tomar logaritmos.

Ejercicios

EJERCICIO 1.1

Reproduzca la simulación de la Figura 1.2 con una estacionalidad cuya amplitud crece 2 % por año. Desestacionalice con X-13 y con STL y compare cuál sigue mejor el cambio de patrón.

EJERCICIO 1.2

Tome la serie mensual del IPC peruano y calcule la inflación de tres maneras: interanual, mensual anualizada sobre la serie observada y mensual anualizada sobre la desestacionalizada. Grafique las tres desde 2020 y explique cuál habría sido más útil para decidir la tasa de política en cada momento.

EJERCICIO 1.3

Estime con `seasonal_adjust` el factor estacional del IPC en dos submuestras, 1996-2010 y 2011-2026. ¿Cambió el patrón? Relacione el resultado con el argumento contra las variables dummy estacionales.

EJERCICIO 1.4

Verifique numéricamente la Ecuación 1.3: calcule la diferencia entre $\Delta \ln y_t$ y g_t para el PBI peruano mes a mes, y encuentre el percentil de la distribución de tasas a partir del cual el error supera un punto porcentual.

EJERCICIO 1.5

Construya la serie trimestral del PBI peruano de dos maneras, promediando los meses y tomando el último mes de cada trimestre. Estime un AR(4) sobre cada una y compare la persistencia estimada. ¿Cuánto de lo que se lee como dinámica económica es consecuencia de la agregación?

CAPÍTULO 2

Modelos univariados

LA PREGUNTA ECONÓMICA

¿Cuánta memoria tiene una serie macroeconómica y cuánto de su futuro puede explicarse con su propio pasado?

Capítulo en preparación. El esqueleto de secciones fija el alcance y el orden de la exposición.

1. Cómo leer la función de autocorrelación de una serie macro y qué modelo sugiere.
2. Cómo estimar y diagnosticar un AR(p) y un ARMA(p,q) en Python.
3. Qué dice y qué no dice una prueba de raíz unitaria sobre la especificación.
4. Por qué la caminata aleatoria es el rival a vencer en cualquier ejercicio de pronóstico.

Requisitos previos. Álgebra matricial básica, estimación por máxima verosimilitud y Capítulo 1.

DATOS UTILIZADOS

Serie	Fuente	Frecuencia	Transformación
PBI real	BCRP	trimestral	var. % interanual
Inflación	BCRP	mensual	var. % interanual

Los datos crudos se descargan a `data/raw/` y las series construidas quedan en `data/processed/`.

2.1. Estacionariedad en sentido débil

Por escribir: qué exige y por qué casi ninguna serie macro en niveles la cumple.

2.2. Procesos AR, MA y ARMA

Por escribir: representación, condiciones de estabilidad e invertibilidad, autocorrelaciones teóricas.

2.3. Estimación

Por escribir: MCO condicional y máxima verosimilitud exacta; qué cambia en muestras cortas.

2.4. Selección de rezagos

Por escribir: AIC, BIC y validación fuera de muestra: tres criterios que rara vez coinciden.

2.5. Raíces unitarias

Por escribir: Dickey-Fuller aumentado, potencia baja y la decisión de diferenciar.

2.6. Pronóstico y evaluación

Por escribir: pronóstico a h pasos, intervalos y comparación contra la caminata aleatoria.

2.7. Diagnósticos

Por escribir: residuos, estabilidad de parámetros y qué hacer cuando fallan.

Ideas clave

1. Un AR(p) bien especificado ya es un pronosticador difícil de batir a horizontes cortos.
2. Las pruebas de raíz unitaria tienen poca potencia: la decisión de diferenciar es, en parte, una decisión económica.
3. Todo lo multivariado del libro es una extensión de este capítulo; conviene tener la intuición univariada firme.

Lecturas recomendadas

Hamilton (1994), caps. 3-5 el tratamiento de referencia de los procesos ARMA.

Box y Jenkins (1970) la metodología original de identificación, estimación y diagnóstico.

Dickey y Fuller (1979) la distribución del estimador bajo raíz unitaria.

Ejercicios

EJERCICIO 2.1

Estime un AR(4) para la inflación y compare su RECM fuera de muestra con la de una caminata aleatoria a horizontes 1, 4 y 8.

EJERCICIO 2.2

Simule 500 series de un AR(1) con coeficiente 0.95 y T = 80. ¿En qué fracción de los casos el ADF rechaza la raíz unitaria?

CAPÍTULO 3

Vectores autorregresivos

LA PREGUNTA ECONÓMICA

¿Cómo se mueven juntas varias series macroeconómicas cuando no queremos comprometernos con una teoría completa de la economía?

1. Escribir el mismo VAR de tres maneras (ecuación por ecuación, compacta y companion) y saber para qué sirve cada una.
2. Estimarlo por mínimos cuadrados, verificar su estabilidad y pronosticar con la forma companion.
3. Entender por qué los residuos de la forma reducida no son choques económicos.

Requisitos previos. Capítulo 2; álgebra matricial (inversa, valores propios, matrices particionadas); distribución normal multivariada.

DATOS UTILIZADOS

Serie	Fuente	Frecuencia	Transformación
Inflación (π_t)	BCRP	trimestral	var. % interanual del IPC
Tasa de referencia (r_t)	BCRP	trimestral	nivel, % anual

El archivo crudo es `data/raw/database_peru.xlsx` (hoja `domestic`) y el CSV que leen los capítulos lo construye `code/python/build_peru_data.py`. El VAR se estima entre 2003Q4 y 2017Q4, y los ocho trimestres siguientes (2018-2019) quedan fuera de la muestra para evaluar el pronóstico.

3.1. Qué hace un VAR

En su revisión de 2001, Stock y Watson describen el trabajo del macroeconometrista en cuatro tareas: describir y resumir series macroeconómicas, pronosticar, recuperar la estructura de la economía a partir de los datos y asesorar a quien decide la política. El VAR es la herramienta estadística que sostiene las cuatro.

Tomemos dos variables, la inflación (π_t) y la tasa de interés de política (r_t). Un VAR ayuda a responder:

- ¿Cómo es el comportamiento dinámico de estas variables y cómo interactúan entre sí?
- ¿Qué inflación es consistente con la trayectoria de tasas de los próximos trimestres?
- ¿Cuál es el efecto de un choque de política monetaria sobre la inflación?
- ¿Cuánto aportaron históricamente los choques de política a las fluctuaciones de la inflación?

Las dos primeras se responden con el modelo de forma reducida de este capítulo. Las dos últimas exigen una hipótesis de identificación y son el asunto del Capítulo 4. La Figura 3.1 ordena el recorrido.

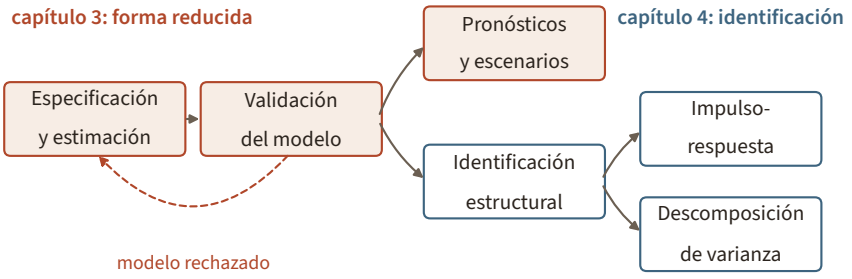


Figura 3.1. Del modelo a las preguntas. La forma reducida (este capítulo) alcanza para describir y pronosticar; las preguntas causales exigen identificar los choques.

INTUICIÓN

Un VAR es una regresión de cada variable sobre el pasado de todas. No hay teoría dentro: la teoría entra al elegir qué variables incluir, en qué transformación, con cuántos rezagos y, sobre todo, al identificar los choques.

3.2. El VAR(p) ecuación por ecuación

DEFINICIÓN 3.1 · VAR(P) DE FORMA REDUCIDA

Sea $\mathbf{y}_t$ un vector $n \times 1$ de variables observadas. El VAR de orden p es

$$\mathbf{y}_t = c + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \cdots + \Phi_p \mathbf{y}_{t-p} + \mathbf{u}_t, \qquad \mathbf{u}_t \sim N(0, \Sigma), \tag{3.1}$$

donde c es $n \times 1$, cada Φ_ℓ es $n \times n$, $\mathbb{E}[\mathbf{u}_t] = 0$ y $\mathbb{E}[\mathbf{u}_t \mathbf{u}_s'] = 0$ para $t \neq s$.

Con dos variables y dos rezagos, la Ecuación 3.1 es un sistema de dos ecuaciones:

$$\begin{bmatrix} \pi_t \\ r_t \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \phi^1_{11} & \phi^1_{12} \\ \phi^1_{21} & \phi^1_{22} \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ r_{t-1} \end{bmatrix} + \begin{bmatrix} \phi^2_{11} & \phi^2_{12} \\ \phi^2_{21} & \phi^2_{22} \end{bmatrix} \begin{bmatrix} \pi_{t-2} \\ r_{t-2} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}. \tag{3.2}$$

Dos detalles que conviene fijar desde el inicio. Primero, $\mathbf{u}_t$ no tiene autocorrelación pero sí correlación contemporánea: Σ no es diagonal, y por eso no se puede mover u_{2t} dejando fijo u_{1t}. A eso vuelve la Sección 3.8. Segundo, todas las ecuaciones tienen exactamente los mismos regresores, y de ahí saldrá el resultado de estimación de la Sección 3.7.

El conteo de parámetros crece con el cuadrado del número de variables. Cada ecuación tiene $k = np + 1$ coeficientes y el sistema tiene nk:

Tabla 3.1. Conteo de parámetros de un VAR sin restricciones

Variables (n)	Rezagos (p)	Coeficientes por ecuación	Total
2	2	5	10
3	4	13	39
5	4	21	105
7	4	29	203

La muestra trimestral de este capítulo tiene 57 observaciones para los diez coeficientes del caso más pequeño de la Tabla 3.1. Con cinco variables y cuatro rezagos harían falta 105, más de los datos disponibles en casi cualquier economía. Por eso existe el Capítulo 5.

3.3. La forma compacta

Para estimar conviene apilar las T observaciones. Con $p = 2$ y las dos variables de la Ecuación 3.2:

$$\underbrace{\begin{bmatrix} \pi_1 & r_1 \\ \pi_2 & r_2 \\ \vdots & \vdots \\ \pi_T & r_T \end{bmatrix}}_{Y\ (T\times n)} = \underbrace{\begin{bmatrix} 1 & \pi_0 & r_0 & \pi_{-1} & r_{-1} \\ 1 & \pi_1 & r_1 & \pi_0 & r_0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \pi_{T-1} & r_{T-1} & \pi_{T-2} & r_{T-2} \end{bmatrix}}_{X\ (T\times k)} \underbrace{\begin{bmatrix} c_1 & c_2 \\ \phi^1_{11} & \phi^1_{21} \\ \phi^1_{12} & \phi^1_{22} \\ \phi^2_{11} & \phi^2_{21} \\ \phi^2_{12} & \phi^2_{22} \end{bmatrix}}_{B\ (k\times n)} + \underbrace{\begin{bmatrix} u_{11} & u_{21} \\ u_{12} & u_{22} \\ \vdots & \vdots \\ u_{1T} & u_{2T} \end{bmatrix}}_{U\ (T\times n)}, \tag{3.3}$$

es decir

$$Y = XB + U. \tag{3.4}$$

La columna j de B contiene los coeficientes de la ecuación j, la primera fila es el intercepto y la matriz X es la misma para todas.

> **CUIDADO**
>
> El lugar de la constante es una convención, no un resultado. En el libro va primero, como en las pizarras y en Blake y Mumtaz; `MacroPy` la pone al final, después de todos los rezagos. Si se mezclan las dos convenciones, el prior del Capítulo 5 termina encogiendo el intercepto en lugar del primer rezago. Conviene imprimir una fila de B y verificar el orden antes de seguir.

3.4. Los datos: inflación y tasa de referencia

El resto del capítulo trabaja con dos series trimestrales del Perú: la inflación interanual y la tasa de referencia del BCRP. La muestra de estimación va de 2003Q4 a 2017Q4, y los dos años siguientes se reservan para evaluar el pronóstico.

Las dos fechas de corte son decisiones sustantivas, no detalles técnicos. El inicio viene impuesto por los datos: el BCRP adopta las metas explícitas de inflación en 2002 y publica la tasa de referencia recién desde septiembre de 2003, de modo que antes de esa fecha no hay un instrumento de política observado que poner en el sistema. El final deja fuera la pandemia y el episodio inflacionario posterior, dos movimientos de una magnitud que un VAR lineal con innovaciones gaussianas no describe bien: incluirlos haría que unos pocos trimestres determinaran los coeficientes. El costo de ambos recortes es una muestra de 57 observaciones para 10 parámetros, que es exactamente la tensión que motiva el Capítulo 5.

```
import numpy as np
import pandas as pd

from macrobook import data_path, var

csv = data_path("peru_domestic.csv")
data = pd.read_csv(csv, index_col="date")
data.index = pd.PeriodIndex(data.index, freq="Q")

series = ["inf", "mpr"]
sample = data.loc["2003Q4":"2017Q4", series]   # estimación
future = data.loc["2018Q1":"2019Q4", series]   # evaluación
dates = sample.index.to_timestamp(how="end")
y = sample.to_numpy()
n, p = y.shape[1], 2

span = f"{sample.index[0]} a {sample.index[-1]}"
```

```
print(f"muestra: {span}, {len(sample)} observaciones")
sample.describe().round(2).T
```

```
muestra: 2003Q4 a 2017Q4, 57 observaciones
```

	count	mean	std	min	25%	50%	75%	max
inf	57.0	3.01	1.34	0.41	2.17	3.01	3.60	6.65
mpr	57.0	3.83	1.12	1.15	3.13	4.16	4.37	6.56

```
import matplotlib.pyplot as plt

from macrobook import use_style, figsize, PALETTE, charts

use_style()

size = figsize(0.58)
fig, axes = plt.subplots(2, 1, sharex=True, figsize=size)
labels = ["inflación (%)", "tasa de referencia (%)"]
for ax, column, label in zip(axes, sample.columns, labels):
    ax.plot(dates, sample[column], color=PALETTE["ink"], lw=1.0)
    ax.set_ylabel(label)
charts.zero_line(axes[0])
axes[1].set_xlabel("")
plt.show()
```

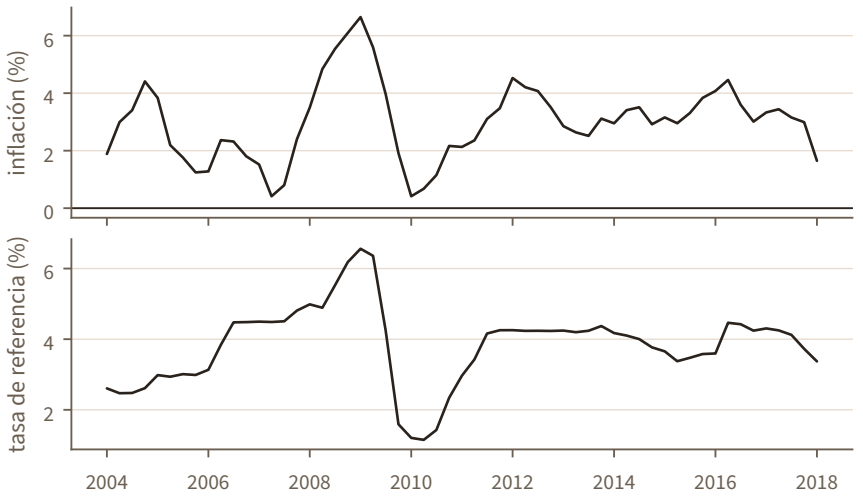


Figura 3.2. Inflación interanual y tasa de referencia del BCRP en la muestra de estimación, 2003Q4 a 2017Q4. Fuente: BCRP.

Las dos series se mueven juntas y con persistencia: los episodios de inflación alta (2008, 2011) vienen acompañados de subidas de la tasa, con rezago. Ese comovimiento es lo que el VAR va a describir, y también la razón por la que describirlo no basta para hablar de efectos de política.

3.5. La forma companion

Todo VAR(p) es un VAR(1) de mayor dimensión. Con el sistema de la Ecuación 3.2, basta apilar el vector de hoy con el de ayer:

$$\underbrace{\begin{bmatrix} \pi_t \\ r_t \\ \pi_{t-1} \\ r_{t-1} \end{bmatrix}}_{\tilde{\mathbf{y}}_t} = \underbrace{\begin{bmatrix} c_1 \\ c_2 \\ 0 \\ 0 \end{bmatrix}}_{\tilde{c}} + \underbrace{\begin{bmatrix} \phi_{11}^1 & \phi_{12}^1 & \phi_{11}^2 & \phi_{12}^2 \\ \phi_{21}^1 & \phi_{22}^1 & \phi_{21}^2 & \phi_{22}^2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}}_{F\ (4\times4)} \begin{bmatrix} \pi_{t-1} \\ r_{t-1} \\ \pi_{t-2} \\ r_{t-2} \end{bmatrix} + \underbrace{\begin{bmatrix} u_{1t} \\ u_{2t} \\ 0 \\ 0 \end{bmatrix}}_{\tilde{\mathbf{u}}_t}. \tag{3.5}$$

Las dos últimas filas son identidades contables: dicen que π_{t-1} de hoy es π_t de ayer. En general, con $\tilde{\mathbf{y}}_t = (\mathbf{y}_t', \mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p+1}')'$ de dimensión np,

$$\tilde{\mathbf{y}}_t = \tilde{c} + F\tilde{\mathbf{y}}_{t-1} + \tilde{\mathbf{u}}_t, \qquad F = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\ I_n & 0 & \cdots & 0 & 0 \\ 0 & I_n & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{bmatrix}, \tag{3.6}$$

y con el selector $J = [\, I_n \; 0 \; \cdots \; 0 \,]$ se recupera el vector original, $\mathbf{y}_t = J\tilde{\mathbf{y}}_t$.

> **INTUICIÓN**
>
> La forma companion cambia p matrices por una sola. Todo lo que quiera decirse sobre el futuro del sistema, o sobre su estabilidad, se vuelve una afirmación sobre potencias de F.

El código imprime, junto con F, el módulo del mayor de sus valores propios. Ese número decide si el sistema es estable, que es el asunto de la sección siguiente.

Desde cero (NumPy)

```
def companion(B, n, p):
    """Build F from B (k x n, constant in the first row)."""
    F = np.zeros((n * p, n * p))
    for lag in range(p):
```

```
        rows = slice(1 + lag * n, 1 + (lag + 1) * n)
        F[:n, lag * n : (lag + 1) * n] = B[rows, :].T
    if p > 1:
        F[n:, : n * (p - 1)] = np.eye(n * (p - 1))
    return F

fit = var.ols(y, p)
F = companion(fit["B"], n, p)
eigenvalues = np.linalg.eigvals(F)

print("máximo módulo:", np.abs(eigenvalues).max().round(3))
np.round(F, 3)
```

```
máximo módulo: 0.838
```

```
array([[ 1.297,  0.16 , -0.6  , -0.081],
       [ 0.145,  1.326, -0.217, -0.493],
       [ 1.   ,  0.   ,  0.   ,  0.   ],
       [ 0.   ,  1.   ,  0.   ,  0.   ]])
```

macrobook

```
F_package = var.companion(fit["B"], n, p)
np.allclose(F, F_package)
```

```
True
```

3.6. Estabilidad y estacionariedad

DEFINICIÓN 3.2 · ESTACIONARIEDAD EN COVARIANZA

El proceso $\{\mathbf{y}_t\}_{t\in\mathbb{Z}}$ es estacionario en covarianza, o débilmente estacionario, si sus dos primeros momentos existen y no dependen de la fecha:

- **Media constante:** $\mathbb{E}[\mathbf{y}_t] = \mu$ para todo t.
- **Autocovarianzas que sólo dependen del rezago:** $\mathbb{E}\left[(\mathbf{y}_t - \mu)(\mathbf{y}_{t-j} - \mu)'\right] = \Gamma_j$ para todo t y todo $j \geq 0$, con Γ_0 finita.

La Definición 3.2 es una propiedad del proceso, no de la muestra: ninguna prueba la verifica directamente. La condición que la garantiza en un VAR es una propiedad de F.

PROPOSICIÓN 3.1 · ESTABILIDAD

El VAR de la Ecuación 3.1 es estable si y sólo si todos los valores propios de la matriz companion F tienen módulo menor que uno,

$$\max_i |\lambda_i(F)| < 1 \iff \det\left(I_n - \Phi_1 z - \cdots - \Phi_p z^p\right) \neq 0 \text{ para } |z| \leq 1.$$

Si el VAR es estable, admite la representación de Wold

$$\mathbf{y}_t = \mu + \sum_{j=0}^{\infty} \Psi_j \mathbf{u}_{t-j}, \qquad \Psi_j = J F^j J', \tag{3.7}$$

con media incondicional $\mu = (I_n - \Phi_1 - \cdots - \Phi_p)^{-1} c$, y es estacionario en covarianza.

La idea detrás de la Ecuación 3.7 es una sustitución recursiva. Partiendo de $\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \mathbf{u}_t$ y reemplazando el rezago una y otra vez,

$$\mathbf{y}_t = \Phi_1(\Phi_1 \mathbf{y}_{t-2} + \mathbf{u}_{t-1}) + \mathbf{u}_t = \cdots = \Phi_1^t \mathbf{y}_0 + \sum_{j=0}^{t-1} \Phi_1^j \mathbf{u}_{t-j},$$

de modo que cada observación es una combinación de los choques presentes y pasados más una condición inicial. Si Φ_1^j no converge a cero, la condición inicial nunca se olvida y la varianza crece sin límite.

```
roots = var.roots(fit["B"], n, p)

ratios = {"width_ratios": [1, 1.4]}
fig, (ax_plane, ax_mod) = plt.subplots(
    1, 2, figsize=figsize(0.36), gridspec_kw=ratios)
charts.unit_circle(ax_plane, roots)

modulus = np.abs(roots)
position = np.arange(len(modulus))[::-1]
ax_mod.hlines(position, 0, modulus,
              color=PALETTE["rule"], lw=1.0)
ax_mod.plot(modulus, position, ls="none", marker="o", ms=4,
            color=PALETTE["accent"])
ax_mod.axvline(1.0, color=PALETTE["ink"], lw=0.8,
               ls=(0, (4, 2)))
n_roots = len(modulus)
tick_labels = [f"$\\lambda_{{{i+1}}}$" for i in range(n_roots)]
ax_mod.set_yticks(position, tick_labels)
ax_mod.set_xlim(0, 1.15)
ax_mod.set_xlabel("módulo")
ax_mod.grid(False, axis="y")
plt.show()
```

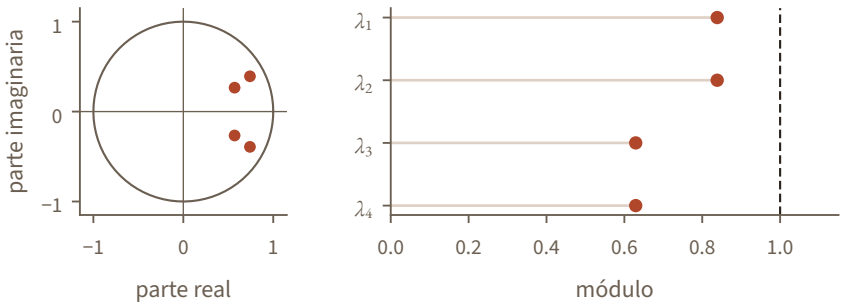


Figura 3.3. Valores propios de la matriz companion estimada: en el plano complejo (izquierda) y ordenados por módulo (derecha). El sistema es estable.

La estabilidad tiene consecuencias visibles: sin choques, el sistema regresa a su media incondicional desde cualquier punto de partida.

```
mu = var.unconditional_mean(fit["B"], n, p)
phis, c_hat = var.coefficients_by_lag(fit["B"], n, p)

def path_without_shocks(start, steps=24):
    history = [np.asarray(start, float)] * p
    out = []
    for _ in range(steps):
        terms = (Phi @ history[lag]
                 for lag, Phi in enumerate(phis))
        nxt = c_hat + sum(terms)
        out.append(nxt)
        history = [nxt] + history[:-1]
    return np.array(out)

fig, ax = plt.subplots(figsize=figsize(0.52))
for k, start in enumerate([[0.0, 2.0], [7.0, 3.0], [3.0, 7.0]]):
    path = path_without_shocks(start)
    ax.plot(path[:, 0], color=PALETTE["shocks"][k], lw=1.1,
            ls=PALETTE["shock_linestyle"][k],
            label=f"$\\pi_0 = {start[0]:.0f}$, "
                  f"$r_0 = {start[1]:.0f}$")
ax.axhline(mu[0], color=PALETTE["accent"], lw=1.0,
           ls=(0, (1, 2)))
ax.set_ylabel("inflación (%)")
ax.set_xlabel("trimestres")
charts.legend_outside(ax, ncol=3)
plt.show()
```

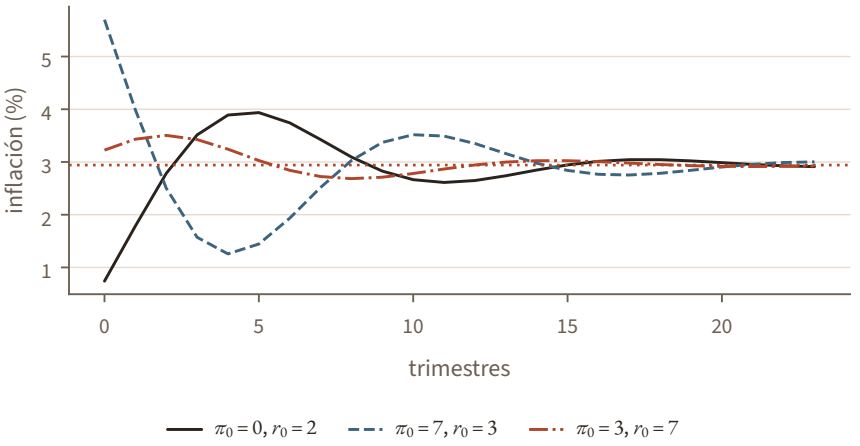


Figura 3.4. Sendas sin choques desde tres puntos de partida. Con el VAR estable, todas convergen a la media incondicional (línea punteada).

CUIDADO

La estabilidad hace falta para que existan los momentos incondicionales, para que las respuestas al impulso converjan a cero y para que la descomposición de varianza tenga sentido. No hace falta para estimar: mínimos cuadrados está bien definido aunque haya una raíz unitaria, y el pronóstico a horizonte finito también. La práctica estándar no es diferenciar por reflejo (Sims et al. 1990): qué variable entra al sistema y en qué transformación se decide antes de estimar, y la Sección 3.10 discute con qué criterio.

3.7. Estimación por mínimos cuadrados

PROPOSICIÓN 3.2 · MÍNIMOS CUADRADOS ECUACIÓN POR ECUACIÓN

En el VAR sin restricciones, donde todas las ecuaciones comparten los regresores X, el estimador de mínimos cuadrados ecuación por ecuación

$$\hat{B} = (X'X)^{-1}X'Y, \qquad \hat{\Sigma} = \frac{(Y - X\hat{B})'(Y - X\hat{B})}{T - k} \tag{3.8}$$

coincide con el estimador de mínimos cuadrados generalizados del sistema completo y, bajo normalidad, con el estimador de máxima verosimilitud condicionado a las p observaciones iniciales.

El resultado, que se demuestra en el Ejercicio 5.1 con el sistema escrito como una sola regresión, es el que permite tratar n ecuaciones como n regresiones separadas. Su contracara aparece en el Capítulo 5: en cuanto el prior enlaza las ecuaciones, o Σ entra en la restricción, la separabilidad se pierde.

Desde cero (NumPy)

```
def lag_matrix(y, p):
    """Y = y[p:] and X = [1, y_{t-1}, ..., y_{t-p}]."""
    T, n = y.shape
    columns = [y[p - lag : T - lag] for lag in range(1, p + 1)]
    X = np.column_stack([np.ones(T - p)] + columns)
    return y[p:], X

Y, X = lag_matrix(y, p)
B_ols = np.linalg.solve(X.T @ X, X.T @ Y)
U = Y - X @ B_ols
T_eff, k = X.shape
Sigma_ols = U.T @ U / (T_eff - k)

# Standard errors: equation i uses Sigma_ii * (X'X)^{-1}.
XtX_inv = np.linalg.inv(X.T @ X)
se = np.sqrt(np.outer(np.diag(XtX_inv), np.diag(Sigma_ols)))

rows = ["constante", "π(-1)", "r(-1)", "π(-2)", "r(-2)"]
cols = sample.columns
coef = pd.DataFrame(B_ols, index=rows, columns=cols).round(3)
stderr = pd.DataFrame(se, index=rows, columns=cols).round(3)
report = coef.astype(str) + " (" + stderr.astype(str) + ")"
report.columns = ["ecuación de π", "ecuación de r"]
report
```

	ecuación de π	ecuación de r
constante	0.583 (0.3)	0.867 (0.193)
π(-1)	1.297 (0.139)	0.145 (0.09)
r(-1)	0.16 (0.19)	1.326 (0.123)
π(-2)	-0.6 (0.139)	-0.217 (0.09)
r(-2)	-0.081 (0.19)	-0.493 (0.122)

macrobook

```
fit = var.ols(y, p)
(np.allclose(fit["B"], B_ols),
 np.allclose(fit["Sigma"], Sigma_ols))
```

```
(True, True)
```

Entre paréntesis van los errores estándar de mínimos cuadrados, $\sqrt{\hat{\sigma}_{ii}\,[(X'X)^{-1}]_{jj}}$ para el coeficiente j de la ecuación i. La lectura inmediata es que cada serie depende sobre todo de su propio pasado. En la ecuación de la inflación los coeficientes de la tasa son pequeños frente a su error estándar; en la ecuación de la tasa los de la inflación son mayores, pero con signos opuestos entre el primer y el segundo rezago. Leer coeficientes de a uno en un VAR rara vez informa: conviene mirar sumas que resuman el sistema.

```
phis, c_hat = var.coefficients_by_lag(np.asarray(B_ols), n, p)
persistence = [sum(Phi[i, i] for Phi in phis) for i in range(n)]
rate_on_inflation = sum(Phi[0, 1] for Phi in phis)  # r -> π
inflation_on_rate = sum(Phi[1, 0] for Phi in phis)  # π -> r

print(f"persistencia propia de π: {persistence[0]:.2f}")
print(f"persistencia propia de r: {persistence[1]:.2f}")
print(f"suma de r en ecuación de π: {rate_on_inflation:+.2f}")
print(f"suma de π en ecuación de r: {inflation_on_rate:+.2f}")
```

```
persistencia propia de π: 0.70
persistencia propia de r: 0.83
suma de r en ecuación de π: +0.08
suma de π en ecuación de r: -0.07
```

Cada cifra es la suma de los dos coeficientes correspondientes de la tabla anterior: la persistencia de la inflación es $1.297 - 0.600$, y la suma de los coeficientes de la tasa en esa misma ecuación es $0.160 - 0.081$. Las dos series son persistentes y las sumas cruzadas son casi cero, con la de la tasa sobre la inflación incluso ligeramente positiva. Sería un error leer esa cifra como "subir la tasa no baja la inflación, o la sube": lo que mide es la correlación parcial entre la tasa de ayer y la inflación de hoy en una muestra donde el banco central subía la tasa precisamente cuando esperaba más inflación. La Sección 3.8 desarrolla ese punto.

3.8. La matriz de covarianzas y el problema de los choques

Hasta aquí Σ ha sido un objeto secundario. Es, en realidad, donde se esconde todo lo que la forma reducida no puede responder. Por definición,

$$\Sigma = \mathbb{E}[\mathbf{u}_t \mathbf{u}_t'] = \begin{bmatrix} \mathbb{E}[u_{1t}^2] & \mathbb{E}[u_{1t}u_{2t}] \\ \mathbb{E}[u_{2t}u_{1t}] & \mathbb{E}[u_{2t}^2] \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}, \tag{3.9}$$

una matriz simétrica con $n(n+1)/2$ elementos distintos: tres en el caso bivariado. El elemento que importa es el de fuera de la diagonal.

```
Sigma_hat = fit["Sigma"]
sd = np.sqrt(np.diag(Sigma_hat))
rho = Sigma_hat[0, 1] / (sd[0] * sd[1])
beta = Sigma_hat[0, 1] / Sigma_hat[1, 1]      # E[u_1 | u_2 = 1]

print("Sigma estimada:\n", np.round(Sigma_hat, 3))
print(f"desviaciones estándar: π {sd[0]:.2f}, r {sd[1]:.2f}")
print(f"correlación entre innovaciones: {rho:.2f}")
print(f"E[u_π | u_r = 1] = {beta:.2f}")
```

```
Sigma estimada:
 [[0.363 0.101]
 [0.101 0.151]]
desviaciones estándar: π 0.60, r 0.39
correlación entre innovaciones: 0.43
E[u_π | u_r = 1] = 0.67
```

Los residuos de la forma reducida son **errores de pronóstico**, no choques económicos: u_{1t} es la parte de la inflación de hoy que el pasado del sistema no anticipaba, y lo mismo u_{2t} para la tasa. Nada garantiza que esas sorpresas tengan una interpretación económica, y de hecho están correlacionadas entre sí.

> **CUIDADO**
>
> Con $\rho \neq 0$, la pregunta "¿qué pasa con la inflación si la tasa sube una unidad de manera inesperada, manteniendo todo lo demás constante?" no tiene respuesta dentro de este modelo. En los datos, las veces en que $u_{2t} = 1$ vienen acompañadas en promedio por $u_{1t} = \sigma_{12}/\sigma_2^2 \neq 0$: mover una innovación dejando la otra fija describe una combinación de sorpresas que la muestra no contiene. Una supuesta impulso-respuesta calculada sobre $\mathbf{u}_t$ no es causal; es una correlación dinámica disfrazada.

Vale la pena leer lo mismo a nivel de coeficientes. En la ecuación de la inflación, la suma de los coeficientes de la tasa resultó prácticamente nula, y en la ecuación de la tasa la suma de los coeficientes de la inflación también. Ninguna de las dos cifras mide un efecto causal, porque ambas describen una muestra en la que el banco central movía la tasa **en respuesta a** lo que veía venir en los precios. Si la política fue sistemática y razonablemente exitosa, la tasa sube justo antes de los episodios inflacionarios y baja cuando ceden: la correlación parcial resultante puede tener cualquier signo, y de hecho la literatura empírica encontró durante años coeficientes con el signo "equivocado", el llamado *price puzzle*, precisamente por esta razón.

La salida es suponer que las sorpresas observadas son combinaciones lineales de choques no observados con significado económico,

$$\mathbf{u}_t = S\varepsilon_t, \qquad \mathbb{E}[\varepsilon_t \varepsilon_t'] = I_n, \qquad \text{de modo que} \quad \Sigma = SS'. \tag{3.10}$$

Aquí ε_t son, por ejemplo, un choque de demanda y un choque de política monetaria, y la matriz de impacto S dice cuánto mueve cada choque a cada variable en el momento cero. El problema es de conteo: Σ ofrece $n(n+1)/2 = 3$ ecuaciones y S tiene $n^2 = 4$ incógnitas, así que faltan $n(n-1)/2 = 1$ restricciones que los datos no pueden entregar.

INTUICIÓN

La estimación de la forma reducida es un problema estadístico y tiene solución única. La identificación es un problema económico: hay que aportar información que no está en los datos. Todo el Capítulo 4 trata de dónde sacar esas restricciones y de cuánto dependen los resultados de haberlas elegido así.

3.9. Pronóstico con la forma companion

El pronóstico óptimo bajo pérdida cuadrática es la esperanza condicional, y en el VAR se obtiene iterando la Ecuación 3.6 hacia adelante con los choques futuros puestos en su media.

Algoritmo 3.1. Pronóstico puntual e intervalos en un VAR estimado

1. Construir el estado inicial apilado con las últimas p observaciones:

$$\mathbf{y}_T^{\sim} = (\mathbf{y}_T', \mathbf{y}_{T-1}', \ldots, \mathbf{y}_{T-p+1}')'.$$

2. Iterar la recursión companion hasta el horizonte deseado y recuperar el vector original con el selector J:

$$\hat{\mathbf{y}}^{\sim}_{T+h|T} = \tilde{c} + F\,\hat{\mathbf{y}}^{\sim}_{T+h-1|T}, \qquad \hat{\mathbf{y}}_{T+h|T} = J\,\hat{\mathbf{y}}^{\sim}_{T+h|T}.$$

3. Acumular la matriz de error cuadrático medio con los pesos de Wold:

$$\mathrm{ECM}_h = \sum_{j=0}^{h-1} \Psi_j \Sigma \Psi_j', \qquad \Psi_j = JF^j J'.$$

4. Formar el intervalo al 95 % para cada variable i:

$$\hat{y}_{i,T+h|T} \pm 1.96 \sqrt{[\,\mathrm{ECM}_h\,]_{ii}}.$$

```
h = len(future)
paths, variances = var.forecast(y, fit["B"], fit["Sigma"], p, h)
sd_h = np.sqrt(variances)

hist = sample.iloc[-32:]
hist_dates = hist.index.to_timestamp(how="end")
fc_dates = future.index.to_timestamp(how="end")
edge_dates = np.r_[hist_dates[-1:], fc_dates]

size = figsize(0.66)
fig, axes = plt.subplots(2, 1, sharex=True, figsize=size)
labels = ["inflación (%)", "tasa de referencia (%)"]
for i, (ax, label) in enumerate(zip(axes, labels)):
    last = y[-1, i]
    band = [(np.r_[last, paths[:, i] - 1.96 * sd_h[:, i]],
             np.r_[last, paths[:, i] + 1.96 * sd_h[:, i]])]
    band_labels = ["IC 95%"] if i == 0 else None
    charts.fan_chart(ax, hist_dates, hist.iloc[:, i],
                     edge_dates, band, np.r_[last, paths[:, i]],
                     realized=future.iloc[:, i],
                     dates_realized=fc_dates,
                     band_labels=band_labels)
    ax.axvline(hist_dates[-1], color=PALETTE["muted"],
               lw=0.8, ls=(0, (3, 2)), zorder=1)
    ax.set_ylabel(label)
top = axes[0].get_ylim()[1]
axes[0].annotate("pronóstico", xy=(edge_dates[2], top),
                 xytext=(0, -10), textcoords="offset points",
                 fontsize=7.5, color=PALETTE["muted"])
charts.legend_outside(axes[0], ncol=4)
plt.show()
```

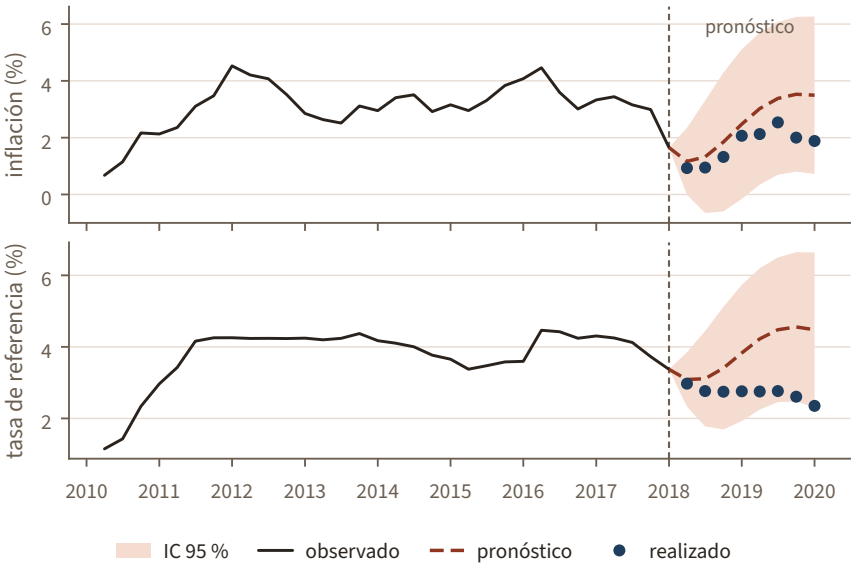


Figura 3.5. Pronóstico a ocho trimestres con intervalo al 95 %, y los datos observados de 2018 y 2019, que no entraron en la estimación. La línea vertical marca el final de la muestra.

Los dos años reservados permiten una evaluación mínima: qué tan lejos quedó el pronóstico puntual y con qué frecuencia el intervalo contuvo al dato.

```
actual = future.to_numpy()
error = actual - paths
rmse = np.sqrt((error ** 2).mean(axis=0))
inside = (np.abs(error) ⩽ 1.96 * sd_h).mean(axis=0)

pct = [f"{x:.0%}" for x in inside]

print(f"RMSE a 8 trimestres: π {rmse[0]:.2f}, r {rmse[1]:.2f}")
print(f"dentro del IC 95%: π {pct[0]}, r {pct[1]}")
```

```
RMSE a 8 trimestres: π 0.94, r 1.38
dentro del IC 95 %: π 100%, r 100%
```

El pronóstico no se equivoca al azar. Sobreestima las dos series, el error crece con el horizonte y la Figura 3.5 muestra por qué: con el sistema estable, la proyección converge a la media incondicional implícita, $\hat{\mu} = (2.94,\ 3.92)$, mientras que la economía de 2018 y 2019 se quedó en un régimen de inflación cerca de 2 % con la tasa de referencia alrededor de 2.75 %. A ocho trimestres el pronóstico de la tasa erra por más de dos puntos porcentuales, y los ocho errores tienen el mismo signo.

Los ocho datos caen dentro del intervalo al 95 %, y eso no es evidencia de que el intervalo esté bien calibrado: son ocho observaciones correlacionadas entre sí, con

las que no se distingue una cobertura de 95 % de una de 100 %. El intervalo tampoco es exigente. A ocho trimestres mide casi tres puntos porcentuales hacia cada lado para la inflación, un rango que la serie no iba a abandonar. Que sea ancho y aun así optimista es el punto del recuadro siguiente.

CUIDADO

El intervalo trata $\hat{B}$ y $\hat{\Sigma}$ como si fueran los valores verdaderos, de modo que subestima la incertidumbre. El sesgo crece cuando la muestra es corta o el sistema es grande, justo donde más importa. La versión honesta integra sobre la distribución de los parámetros, que es lo que hace el pronóstico por densidad del Capítulo 5.

Queda una decisión de reporte: el nivel. En pronóstico frecuentista lo habitual es 90 % o 95 %, y los abanicos de banca central muestran varios tramos hasta el 90 %. Las bandas al 68 %, tan frecuentes en la literatura de VAR estructurales, vienen de otra tradición: son intervalos de una desviación estándar, popularizados por Sims y Zha (1999), que argumentaron que en muestras cortas el 95 % queda tan ancho que no disciplina nada. La crítica moderna es que esas bandas ni son intervalos de confianza al 68 % en sentido estricto ni son simultáneas a lo largo del horizonte, de modo que la cobertura conjunta es bastante menor de lo que el número sugiere (Montiel Olea y Plagborg-Møller 2019). La regla del libro: declarar el nivel, decir si es puntual o simultáneo, y no cambiarlo de capítulo en capítulo para que la figura se vea mejor.

3.10. Otros temas del VAR clásico

Lo que sigue son cuestiones estándar de la práctica con VAR que este libro no desarrolla, porque están bien tratadas en los manuales y porque el hilo del texto va hacia la identificación y los métodos bayesianos. Van aquí los criterios que conviene tener presentes y la referencia donde estudiarlos.

3.10.1. Selección de rezagos

Los criterios de información penalizan el ajuste por el número de parámetros: AIC penaliza $2k$, el de Hannan y Quinn $2k \ln(\ln T)$, con el logaritmo aplicado dos veces, y el bayesiano (BIC o SIC) $k \ln T$. El BIC es consistente para el orden verdadero y el AIC no, pero el AIC suele elegir modelos que pronostican mejor en muestras cortas, así que la elección depende del objetivo. Dos reglas prácticas: comparar siempre sobre la misma muestra efectiva, porque cambiar p cambia el número de observaciones utilizables; y, si el fin es el análisis de impulso-respuesta con datos trimestrales, seguir a Ivanov y Kilian (2005), que recomiendan AIC en esa configuración y HQ en muestras largas o en frecuencia mensual. Con datos trimestrales, $p = 4$ es el punto

de partida habitual y $p = 2$ una elección defendible cuando la muestra es corta. Véase Lütkepohl (2005, cap. 4).

3.10.2. Diagnóstico de residuos

El supuesto que hay que vigilar es la ausencia de autocorrelación, porque de él depende que los rezagos incluidos basten para blanquear el error. Se contrasta con un estadístico Portmanteau (Ljung-Box multivariado) o con la prueba LM de Breusch y Godfrey. La normalidad se examina con una versión multivariada de Jarque-Bera, y conviene recordar que no es necesaria para la consistencia: sólo afecta a la inferencia en muestras pequeñas. La heterocedasticidad condicional se detecta con ARCH-LM y se modela en el Capítulo 9. Para la estabilidad de los coeficientes existen pruebas de quiebre y estadísticos CUSUM, y el tratamiento explícito está en el Capítulo 10. Buena práctica: ante autocorrelación residual, agregar rezagos antes que cambiar de modelo. Véase Lütkepohl (2005, cap. 4).

3.10.3. Niveles, diferencias y cointegración

Conviene separar dos decisiones que suelen confundirse. La primera es **qué variable económica interesa**: el nivel de precios o la inflación, el PIB o su tasa de crecimiento. Esa elección es sustantiva y viene antes que la econometría; en este capítulo interesan la inflación y la tasa de política, no el índice de precios. La segunda, ya fijadas las variables, es **si las series que entran al sistema tienen raíz unitaria** y qué hacer al respecto. La discusión que sigue es sobre la segunda.

- **Variables con tendencia estocástica, estimadas en niveles.** Mínimos cuadrados sigue siendo consistente y las respuestas de corto plazo no se distorsionan, pero la inferencia sobre funciones de largo plazo deja de ser estándar y algunos estadísticos ya no son asintóticamente normales (Sims et al. 1990). Es lo que hace la mayor parte de la literatura estructural desde Sims (1980), con un argumento claro: si las series comparten una relación de equilibrio, diferenciarlas la destruye.
- **Variables con tendencia estocástica, en diferencias o con cointegración explícita.** Si el objeto de interés es esa relación de largo plazo, el modelo correcto es el VECM, que separa el ajuste de corto plazo del equilibrio; diferenciar sin el término de corrección de error pierde el equilibrio, y dejarlo implícito en un VAR en niveles complica la inferencia sobre él. El número de vectores de cointegración se contrasta con el procedimiento de máxima verosimilitud de Johansen (1995), sobre la representación de Engle y Granger (1987).

Este capítulo no está en ninguno de los dos casos, y vale la pena ver por qué. La inflación interanual es una transformación del IPC, pero no es la variable original diferenciada para forzar estacionariedad: es la variable que el banco central controla y sobre la que anuncia su meta. Lo mismo la tasa de referencia, que es un nivel.

Ninguna de las dos es candidata natural a raíz unitaria bajo un régimen de metas, y el módulo máximo de las raíces estimadas (0.84) es compatible con ello, aunque mínimos cuadrados sesga las raíces hacia abajo en muestras de este tamaño. La discusión anterior se vuelve inevitable cuando al sistema entran el producto, el tipo de cambio o los precios en logaritmos. Para el tratamiento completo, Lütkepohl (2005, caps. 6-8).

3.10.4. Tamaño del sistema

La decisión que más determina los resultados no es el orden de rezagos sino qué variables entran. Un VAR pequeño y bien argumentado suele ser preferible a uno grande estimado sin restricciones, y cuando el sistema debe crecer la respuesta no es eliminar variables sino encogerlas: eso es el Capítulo 5.

Ideas clave

1. El VAR es un dispositivo de resumen de la dinámica conjunta: describe correlaciones dinámicas, no efectos causales.
2. Las tres representaciones son el mismo modelo: la compacta sirve para estimar y la companion para proyectar y para estudiar la estabilidad. El Capítulo 5 agrega una cuarta, la vectorizada, cuando hace falta enlazar las ecuaciones.
3. La estabilidad es una propiedad de los valores propios de la companion, y es lo que permite hablar de media incondicional, de representación de Wold y de respuestas que se disipan.
4. Con los mismos regresores en todas las ecuaciones, mínimos cuadrados ecuación por ecuación no deja nada sobre la mesa; con priors o restricciones cruzadas, sí.
5. Σ no es diagonal, y por eso los residuos de la forma reducida no se pueden mover de a uno: sin una hipótesis de identificación no hay impulso-respuesta con interpretación causal.

Lecturas recomendadas

Sims (1980) El artículo que inició el programa; conviene leer la crítica a las restricciones increíbles de los modelos de ecuaciones simultáneas.

Stock y Watson (2001) Revisión breve y honesta de lo que un VAR puede y no puede hacer.

Lütkepohl (2005), caps. 2 y 3 El tratamiento completo de la forma companion, la estabilidad y la inferencia asintótica.

Kilian y Lütkepohl (2017), cap. 2 La misma materia con el énfasis puesto en lo que hará falta para el capítulo siguiente.

Sims y Zha (1999) El origen de las bandas al 68 % y el argumento a su favor.

Ejercicios

EJERCICIO 3.1

Escriba a mano la matriz companion de un VAR(3) con dos variables y verifique en Python que iterarla reproduce el pronóstico obtenido por recursión directa sobre la Ecuación 3.1.

EJERCICIO 3.2

Estime el VAR con $p = 1, 2, 3, 4$ sobre la misma muestra efectiva y compare AIC, BIC y el módulo máximo de las raíces. ¿Coinciden los criterios? ¿Cambia alguna conclusión sobre la persistencia del sistema?

EJERCICIO 3.3

Reestime el modelo extendiendo la muestra hasta 2019Q4, es decir agregando las ocho observaciones que aquí se reservaron. Compare los coeficientes, la correlación entre innovaciones y el ancho del intervalo a ocho trimestres. ¿Cuánto mueven ocho observaciones a un sistema estimado con 57?

EJERCICIO 3.4

Calcule $\hat{\mu} = (I - \hat{\Phi}_1 - \hat{\Phi}_2)^{-1}\hat{c}$ y compárela con el promedio muestral de cada serie. ¿Por qué la media implícita se estima mejor que el intercepto? Compare también $\hat{\mu}_\pi$ con la meta de inflación del BCRP.

EJERCICIO 3.5

Simule con `var.simulate` un sistema con los coeficientes estimados pero con $\sigma_{12} = 0$, y vuelva a estimar. ¿Cambian los coeficientes? ¿Cambia lo que se puede afirmar sobre el efecto de un choque de tasa? Use el resultado para explicar por qué la identificación es un problema aparte de la estimación.

EJERCICIO 3.6

Tome $\hat{\Phi}_1$ y multiplíquelo por un escalar creciente hasta que el módulo máximo de los valores propios pase de uno. Simule el sistema justo antes y justo después del umbral y grafique las trayectorias. ¿Qué cambia en la escala del eje vertical?

CAPÍTULO 4

VAR estructurales

LA PREGUNTA ECONÓMICA

¿Qué parte del movimiento conjunto de las series podemos atribuir a un choque económico concreto?

1. Plantear el problema de identificación con precisión y contar cuántas restricciones hacen falta.
2. Identificar los choques por orden recursivo, por efectos de largo plazo o por signos, y leer las respuestas, la varianza y la historia que resultan.
3. Reconocer qué parte de la incertidumbre reportan las bandas y cuál no.

Requisitos previos. Capítulo 3; descomposición de Cholesky; nociones de identificación en modelos de ecuaciones simultáneas.

DATOS UTILIZADOS

Serie	Fuente	Frecuencia	Transformación
Inflación (π_t)	BEA	trimestral	$400 \ln(P_t/P_{t-1})$, deflactor del PBI
Desempleo (u_t)	BLS	trimestral	tasa civil, promedio del trimestre
Tasa de fondos federales (R_t)	Reserva Federal	trimestral	promedio del trimestre

Los datos de Estados Unidos entre 1960Q1 y 2000Q4 son los del material de replicación de Stock y Watson (2001): `data/raw/SW2001_Data.xlsx`. El Sección 4.8 usa además la muestra de Blanchard y Quah (1989) (`data/raw/BQ1989_Data.xlsx`). Ambos CSV los construye `code/python/build_us_data.py`.

4.1. De los residuos a los choques

El Capítulo 3 terminó en un callejón deliberado. La forma reducida describe bien la dinámica conjunta, pronostica de manera razonable y no permite responder la

pregunta que interesa a quien decide la política: qué pasa con la inflación y el producto si la tasa sube por una decisión propia del banco central y no como respuesta a lo que el banco central observa. El obstáculo es que los residuos $\mathbf{u}_t$ son errores de pronóstico correlacionados entre sí, no experimentos.

La salida es suponer que esas sorpresas observadas son combinaciones lineales de un número igual de choques no observados con significado económico.

SUPUESTO 4.1 · CHOQUES ESTRUCTURALES

Existe una matriz S de $n \times n$, invertible, y un vector de choques estructurales ε_t tal que

$$\mathbf{u}_t = S\varepsilon_t, \qquad \mathbb{E}[\varepsilon_t] = 0, \qquad \mathbb{E}[\varepsilon_t \varepsilon_t'] = I_n, \qquad \mathbb{E}[\varepsilon_t \varepsilon_s'] = 0 \ \text{ para } t \neq s.$$

Los choques son ortogonales entre sí y están normalizados a varianza unitaria, de modo que mover uno dejando los demás en cero es una operación bien definida. La columna k de S recoge el efecto contemporáneo del choque k sobre cada variable.

La normalización a varianza unitaria no tiene contenido: reescalar ε_{kt} y dividir la columna k de S por el mismo número deja el modelo igual. Lo que sí tiene contenido es la ortogonalidad, porque es la que permite hablar de un choque a la vez, y es también la que hace falta interpretar: dos choques ortogonales en el modelo son dos fuerzas que en la economía no se mueven juntas por construcción, y eso hay que argumentarlo.

Sustituyendo el supuesto en la forma reducida,

$$\mathbf{y}_t = c + \Phi_1 \mathbf{y}_{t-1} + \cdots + \Phi_p \mathbf{y}_{t-p} + S\varepsilon_t. \tag{4.1}$$

La Ecuación 4.1 es el VAR estructural. Tiene dos piezas con estatus distinto. Las matrices Φ_ℓ son la **dinámica**, y se estiman por mínimos cuadrados como en el Sección 3.7. La matriz S es el **impacto**, no se estima con los datos solos y es donde entra la economía.

INTUICIÓN

El SVAR puede leerse como una aproximación lineal a la solución de un modelo estructural: la dinámica resume cómo se propagan las perturbaciones y la matriz de impacto dice quién empuja a quién en el momento cero. La diferencia con un DSGE es de dónde vienen las restricciones: allá, de la forma completa del modelo; aquí, de unos pocos supuestos que se declaran uno por uno.

4.2. El problema de identificación

Los datos informan sobre Σ, no sobre S. La relación entre ambos sale del supuesto:

$$\Sigma = \mathbb{E}[\mathbf{u}_t \mathbf{u}_t'] = \mathbb{E}[S\varepsilon_t \varepsilon_t' S'] = S\,\mathbb{E}[\varepsilon_t \varepsilon_t']\,S' = SS'. \tag{4.2}$$

Identificar es resolver la Ecuación 4.2 para S. El problema es de conteo.

PROPOSICIÓN 4.1 · GRADOS DE LIBERTAD DE LA IDENTIFICACIÓN

Σ es simétrica, de modo que aporta $n(n+1)/2$ ecuaciones distintas, mientras que S tiene n^2 incógnitas. Hacen falta

$$n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2}$$

restricciones adicionales para que la solución sea única, y esas restricciones no pueden venir de los datos: cualquier $\tilde{S} = SQ$ con Q ortonormal reproduce exactamente la misma Σ y, por lo tanto, la misma verosimilitud.

Con dos variables falta una restricción; con tres, como en el sistema de este capítulo, faltan tres; con siete variables, veintiuna. El último punto de la Proposición 4.1 es el que conviene grabar: el conjunto de matrices de impacto compatibles con los datos no es pequeño ni raro, es una variedad continua del mismo tamaño que el grupo ortogonal, y todos sus elementos ajustan los datos igual de bien. Ninguna prueba estadística puede escoger entre ellos.

> **CUIDADO**
>
> Que dos modelos ajusten igual no significa que digan lo mismo. Dos matrices de impacto distintas compatibles con la misma Σ pueden implicar que la política monetaria contrae el producto o que lo expande. La elección entre ellas es un supuesto económico que el investigador declara y defiende, no un resultado que los datos entreguen. Por eso un SVAR se reporta siempre junto con su esquema de identificación, y por eso la sensibilidad a ese esquema es parte del resultado.

Las secciones siguientes recorren los tres esquemas más usados: restricciones de cero contemporáneas, restricciones de cero de largo plazo y restricciones de signo. Cada uno aporta la información que falta de una manera distinta, y cada uno cuesta algo.

4.2.1. El sistema de trabajo

El Capítulo 3 trabajó con datos peruanos para mostrar cómo se estima y se pronostica un VAR. Para el análisis estructural conviene un sistema con historia publicada, donde los resultados se puedan contrastar con los de otros, así que este capítulo

replica el ejemplo de Stock y Watson (2001): inflación, desempleo y tasa de fondos federales de Estados Unidos, datos trimestrales de 1960Q1 a 2000Q4, cuatro rezagos. Los ejercicios del final vuelven sobre los datos peruanos.

```
import numpy as np
import pandas as pd

from macrobook import data_path, var, svar

csv = data_path("sw2001.csv")
data = pd.read_csv(csv, index_col="date")
data.index = pd.PeriodIndex(data.index, freq="Q")

order = ["infl", "unemp", "ff"]
sample = data[order]
dates = sample.index.to_timestamp(how="end")
y = sample.to_numpy()
n, p = len(order), 4

fit = var.ols(y, p)
max_root = np.abs(var.roots(fit["B"], n, p)).max()
print("muestra:", sample.index[0], "a", sample.index[-1])
print("observaciones:", len(sample))
print("módulo máximo de las raíces:", max_root.round(3))
```

```
muestra: 1960Q1 a 2000Q4
observaciones: 164
módulo máximo de las raíces: 0.969
```

```
import matplotlib.pyplot as plt

from macrobook import use_style, figsize, PALETTE, charts

use_style()

labels = ["inflación (%)", "desempleo (%)",
          "fondos federales (%)"]
names = ["inflación", "desempleo", "tasa"]

size = figsize(0.62)
fig, axes = plt.subplots(3, 1, sharex=True, figsize=size)
for ax, column, label in zip(axes, order, labels):
    ax.plot(dates, sample[column], color=PALETTE["ink"], lw=1.0)
    ax.set_ylabel(label)
```

```
plt.show()
```

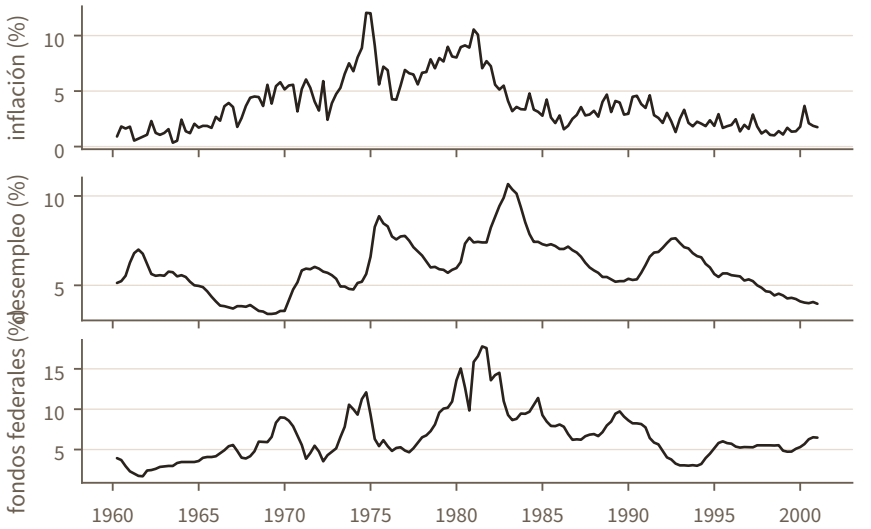


Figura 4.1. Las tres series del sistema de Stock y Watson (2001), 1960Q1 a 2000Q4.

La muestra cubre la Gran Inflación de los años setenta, la desinflación de Volcker y la larga expansión de los noventa. El módulo máximo de las raíces, 0.97, avisa de entrada que el sistema es muy persistente: las respuestas van a decaer despacio y las bandas a horizontes largos serán anchas.

La matriz de covarianzas de los residuos vuelve a ser el objeto central. Conviene mirarla en correlaciones, que se leen más rápido.

```
corr = np.corrcoef(fit["residuals"].T)
pd.DataFrame(corr.round(2), index=order, columns=order)
```

	infl	unemp	ff
infl	1.00	-0.06	0.12
unemp	-0.06	1.00	-0.45
ff	0.12	-0.45	1.00

La correlación relevante es la de −0.45 entre la sorpresa de desempleo y la sorpresa de tasa: cuando el desempleo sorprende al alza, la tasa sorprende a la baja. Eso es exactamente lo que se espera de una regla de política que reacciona al ciclo, y es también la razón por la que esa correlación no se puede leer como el efecto de la política sobre el desempleo.

4.3. Identificación recursiva

El esquema más antiguo y todavía el más usado impone ceros en la matriz de impacto (Sims 1980; Christiano et al. 1999). La idea es ordenar las variables de la más lenta a la más rápida en reaccionar dentro del período, y suponer que cada variable no responde contemporáneamente a los choques de las que vienen después.

Con el orden (π_t, u_t, R_t) de Stock y Watson (2001) la matriz de impacto es triangular inferior:

$$\begin{bmatrix} u_{\pi,t} \\ u_{u,t} \\ u_{R,t} \end{bmatrix} = \begin{bmatrix} s_{11} & 0 & 0 \\ s_{21} & s_{22} & 0 \\ s_{31} & s_{32} & s_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon^{\text{pol}}_{3,t} \end{bmatrix}. \tag{4.3}$$

Los tres ceros son exactamente las $n(n-1)/2 = 3$ restricciones que faltaban, y dicen dos cosas distintas. Hacia abajo: la tasa de fondos federales responde dentro del trimestre a la inflación y al desempleo, que es lo que hace un banco central que mira los datos al decidir. Hacia arriba: el choque de política no mueve la inflación ni el desempleo dentro del mismo trimestre, porque los precios y las decisiones de gasto y contratación tardan en reaccionar. El supuesto es fuerte en frecuencia mensual y bastante cómodo en frecuencia trimestral.

PROPOSICIÓN 4.2 · IDENTIFICACIÓN RECURSIVA

Si S es triangular inferior con diagonal positiva, entonces S es el único factor de Cholesky de Σ, esto es, la única matriz triangular inferior con diagonal positiva que cumple $\Sigma = SS'$.

EJEMPLO 4.1 · CHOLESKY A MANO

Con $\hat{\Sigma} = \begin{bmatrix} 4 & 2 \\ 2 & 10 \end{bmatrix}$, igualar $\hat{\Sigma} = SS'$ con S triangular inferior da tres ecuaciones:

$$s_{11}^2 = 4, \qquad s_{11}s_{21} = 2, \qquad s_{21}^2 + s_{22}^2 = 10.$$

Resolviendo en cadena, $s_{11} = 2$, $s_{21} = 1$ y $s_{22} = 3$. Tres ecuaciones y tres incógnitas: la triangularidad convirtió un problema indeterminado en uno exactamente identificado.

Desde cero (NumPy)

```
S = np.linalg.cholesky(fit["Sigma"])
shocks = ["inflación", "desempleo", "política"]
pd.DataFrame(S.round(3), index=order, columns=shocks)
```

	inflación	desempleo	política
infl	0.985	0.000	0.000
unemp	-0.013	0.226	0.000
ff	0.109	-0.395	0.784

macrobook

```
S_package = svar.cholesky_impact(fit["Sigma"])
np.allclose(S, S_package)
```

```
True
```

La tercera columna es el choque de política: eleva la tasa de fondos federales en 0.78 puntos porcentuales en el trimestre del choque y, por construcción, no toca a la inflación ni al desempleo en ese momento. El elemento $s_{32} = -0.40$ es el que da contenido al orden: ante un choque de desempleo, la tasa baja cuatro décimas dentro del mismo trimestre. Los dos primeros choques llevan el nombre de la variable que encabezan, no el de una interpretación económica, y conviene que sea así: el orden elegido identifica con un argumento defendible el choque de política, y deja a los otros dos como las combinaciones que quedan. Ponerles nombre de choque de oferta o de demanda exigiría un argumento adicional que este esquema no aporta.

> **CUIDADO**
>
> El orden de las variables es una restricción económica, no una convención de programación. Cambiarlo cambia los choques y puede cambiar las conclusiones. La práctica mínima es reportar el orden elegido, justificarlo y mostrar qué pasa con otro orden razonable. Cuando el resultado depende del orden, lo que hay que reportar es esa dependencia, no el orden más favorable.

4.4. Funciones de impulso-respuesta

Identificada la matriz de impacto, el resto es aritmética de la forma companion.

DEFINICIÓN 4.1 · FUNCIÓN DE IMPULSO-RESPUESTA

La respuesta de las variables al choque estructural k al cabo de h períodos es

$$\Theta_h = \frac{\partial \mathbf{y}_{t+h}}{\partial \varepsilon_t'} = \Psi_h S, \qquad \Psi_h = J F^h J', \tag{4.4}$$

donde F es la matriz companion y J el selector del Sección 3.5. El elemento (i, k) de Θ_h es la respuesta de la variable i al choque k en el horizonte h, manteniendo los demás choques en cero.

La derivación es la sustitución recursiva de siempre. Partiendo de $\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + S\varepsilon_t$ e iterando hacia adelante, el término que sobrevive del choque de hoy en $t + h$ es $\Phi_1^h S\varepsilon_t$; con $p > 1$, la potencia de la companion hace el mismo trabajo.

EJEMPLO 4.2 · UNA RESPUESTA CALCULADA A MANO

Con $\hat{\Phi} = \begin{bmatrix} 0.5 & 0.1 \\ 0.2 & 0.4 \end{bmatrix}$ y $\hat{S} = \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix}$, un choque unitario al primer elemento da

$$\Theta_0 \varepsilon = \hat{S} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \qquad \Theta_1 \varepsilon = \hat{\Phi} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.1 \\ 0.8 \end{bmatrix}, \qquad \Theta_2 \varepsilon = \hat{\Phi} \begin{bmatrix} 1.1 \\ 0.8 \end{bmatrix} = \begin{bmatrix} 0.63 \\ 0.54 \end{bmatrix}.$$

La respuesta decae porque el sistema es estable. En un sistema inestable esta misma recursión explota, que es la razón práctica por la que la estabilidad se verifica antes de graficar nada.

La escala del choque es una convención más. Un choque de una desviación estándar es el tamaño natural en los datos; un choque unitario, que se obtiene dividiendo la columna k de S por s_{kk}, hace que el choque mueva a su propia variable exactamente un punto porcentual y es más fácil de comunicar. Stock y Watson (2001) usan el segundo y aquí se hace lo mismo.

Desde cero (NumPy)

```
def responses(B, S, n, p, h):
    """Theta_j = Psi_j S para j = 0, ..., h."""
    return var.ma_weights(B, n, p, h) @ S

S_unit = S / np.diag(S)     # un choque de un punto porcentual
theta = responses(fit["B"], S_unit, n, p, 24)
pd.DataFrame(theta[:5, :, 2].round(3), columns=order)
```

	infl	unemp	ff
0	0.000	0.000	1.000
1	0.158	0.005	0.946
2	0.109	0.062	0.531
3	-0.006	0.109	0.444
4	-0.012	0.140	0.496

macrobook

```
theta_package = svar.responses(fit["B"], svar.unit_scale(S),
                               n, p, 24)
np.allclose(theta, theta_package)
```

```
True
```

4.5. Bandas de confianza

Las respuestas son funciones no lineales de los coeficientes estimados, así que su distribución muestral no tiene forma cerrada útil. La práctica estándar es el bootstrap (Kilian 1998; Kilian y Lütkepohl 2017, cap. 12).

Algoritmo 4.1. Bandas bootstrap para las respuestas estructurales

1. Estimar el VAR y guardar $\hat{B}$, $\hat{\Sigma}$ y los residuos $\hat{\mathbf{u}}_t$.
2. Remuestrear con reemplazo los residuos para obtener $\hat{\mathbf{u}}^*_t$, y construir una muestra artificial **recursivamente**, con las mismas p observaciones iniciales:

$$\mathbf{y}^*_t = \hat{c} + \hat{\Phi}_1 \mathbf{y}^*_{t-1} + \cdots + \hat{\Phi}_p \mathbf{y}^*_{t-p} + \hat{\mathbf{u}}^*_t.$$

3. Reestimar el VAR sobre $\mathbf{y}^*$, volver a identificar y calcular Θ^*_b. Repetir N veces.
4. Reportar los percentiles 2.5 y 97.5 de los Θ^*_b en cada horizonte.

El paso 2 es el que suele salir mal. La muestra artificial tiene que generarse iterando el modelo, no sumando residuos remuestreados a los valores ajustados: si los regresores se dejan en sus valores originales mientras la variable dependiente se construye con ruido nuevo, los rezagos y la variable dependiente dejan de corresponderse y el estimador bootstrap queda sesgado hacia cero en la persistencia.

```
lower, upper = svar.bootstrap_bands(
    y, p, h=24, draws=2000, level=0.95, unit=True, seed=7)
print("respuesta del desempleo en h=8:")
print("  punto  %.2f" % theta[8, 1, 2])
print("  banda [%.2f, %.2f]" % (lower[8, 1, 2], upper[8, 1, 2]))
```

```
respuesta del desempleo en h=8:
  punto  0.20
  banda [0.11, 0.30]
```

```
resp = [[theta[:, i, k] for k in range(n)] for i in range(n)]
bands = [[[(lower[:, i, k], upper[:, i, k])] for k in range(n)]
         for i in range(n)]
fig, axes = charts.irf_grid(resp, labels, shocks, bands=bands)
axes[-1][1].set_xlabel("trimestres")
plt.show()
```

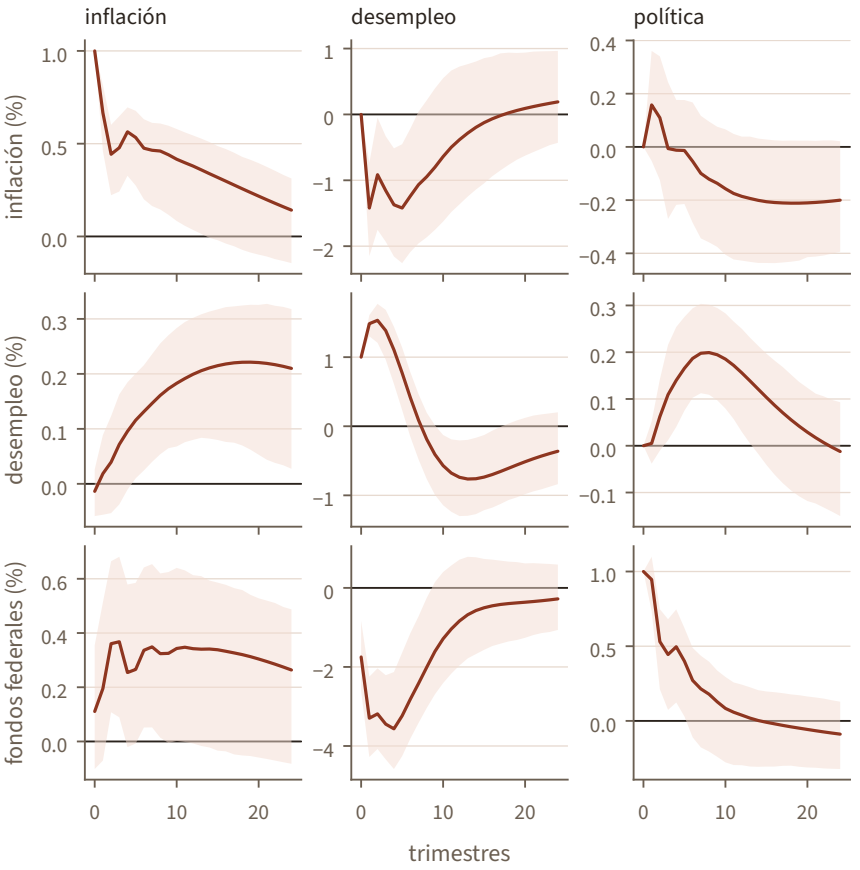


Figura 4.2. Respuestas a un choque unitario bajo identificación recursiva, con bandas bootstrap al 95 %. Las filas son variables y las columnas, choques. Replica la Figura 1 de Stock y Watson (2001).

La columna de la derecha es el choque de política monetaria y reproduce el resultado central de Stock y Watson (2001). Un aumento de un punto porcentual en la tasa de fondos federales eleva el desempleo, que alcanza un máximo de 0.20 puntos a los ocho trimestres con una banda de 0.11 a 0.30, y luego vuelve a cero hacia el sexto año. El signo, la forma de joroba y el rezago de dos años son los hechos estilizados que la literatura de política monetaria reporta desde entonces.

La respuesta de la inflación es el punto incómodo: sube un par de décimas durante el primer año antes de volverse negativa, y su banda contiene al cero durante todo el recorrido. Es el *price puzzle* del Sección 3.8, que aparece aquí a pesar de haber identificado. La lectura estándar es que el banco central reacciona a información sobre precios futuros que el sistema no contiene, de modo que parte de lo que el esquema llama choque de política es en realidad la respuesta a una presión inflacionaria que el VAR no ve (Christiano et al. 1999). El remedio usual es ampliar el sistema con un índice de precios de materias primas, que actúa como indicador adelantado de la inflación.

> **CUIDADO**
>
> La banda bootstrap recoge la incertidumbre de estimación de $\hat{B}$ y $\hat{\Sigma}$, y nada más. La incertidumbre sobre el esquema de identificación, que suele ser mayor, queda fuera por construcción: la banda se calcula manteniendo fijo el supuesto de triangularidad. Una banda estrecha alrededor de una respuesta mal identificada es precisión sobre el objeto equivocado.

4.6. Descomposición de varianza

La impulso-respuesta dice cómo se mueve el sistema ante un choque. La descomposición de varianza dice cuánto importa ese choque en promedio.

La descomposición de varianza no depende de cómo se escale el choque, así que aquí se vuelve a las respuestas de una desviación estándar.

DEFINICIÓN 4.2 · DESCOMPOSICIÓN DE VARIANZA DEL ERROR DE PRONÓSTICO

El error de pronóstico a h períodos es $\mathbf{y}_{t+h} - \mathbb{E}_t[\mathbf{y}_{t+h}] = \sum_{j=0}^{h} \Theta_j \varepsilon_{t+h-j}$. Como los choques son ortogonales y de varianza unitaria, la varianza del error de la variable i se parte sin residuo entre los choques:

$$\mathrm{FEVD}_{i,k}(h) = \frac{\sum_{j=0}^{h} \theta_{ik,j}^2}{\sum_{m=1}^{n} \sum_{j=0}^{h} \theta_{im,j}^2}, \qquad \sum_{k=1}^{n} \mathrm{FEVD}_{i,k}(h) = 1. \tag{4.5}$$

Desde cero (NumPy)

```
theta_sd = responses(fit["B"], S, n, p, 24)
contributions = np.cumsum(theta_sd ** 2, axis=0)
shares = contributions / contributions.sum(axis=2,
                                           keepdims=True)

quarters = [1, 4, 8, 12]
```

```
rows = pd.MultiIndex.from_product([names, quarters],
                                  names=["variable", "h"])
table = pd.DataFrame(
    [shares[q - 1, i, :] * 100 for i in range(n)
     for q in quarters], index=rows, columns=shocks)
table.round(0)
```

variable	h	inflación	desempleo	política
inflación	1	100.0	0.0	0.0
	4	88.0	10.0	1.0
	8	83.0	16.0	1.0
	12	83.0	15.0	2.0
desempleo	1	0.0	100.0	0.0
	4	2.0	96.0	2.0
	8	10.0	76.0	13.0
	12	21.0	60.0	19.0
tasa	1	2.0	20.0	79.0
	4	9.0	51.0	41.0
	8	11.0	60.0	29.0
	12	15.0	59.0	26.0

macrobook

```
shares_package = svar.fevd(fit["B"], S, n, p, 24)
np.allclose(shares, shares_package)
```

```
True
```

MacroPy

```
from MacroPy import ClassicVAR

frame = sample.copy()
frame.index = dates
model = ClassicVAR(frame, lags=p, hor=25, bstrap=False)
fevd_macropy = model.compute_fevd(plot_fevd=False)
np.allclose(shares[11] * 100, fevd_macropy[11].T)
```

```
True
```

```
size = figsize(0.40)
fig, axes = plt.subplots(1, 3, sharey=True, figsize=size)
horizons = np.arange(shares.shape[0])
for i, (ax, name) in enumerate(zip(axes, names)):
    ax.stackplot(horizons, shares[:, i, :].T * 100,
                 colors=PALETTE["fills"][:n], labels=shocks,
                 edgecolor="white", lw=0.4)
    ax.set_title(name, fontsize=8.5)
    ax.set_xlabel("trimestres")
    ax.set_ylim(0, 100)
    ax.grid(False, axis="x")
axes[0].set_ylabel("% de la varianza")
charts.legend_outside(axes[1], ncol=3)
plt.show()
```

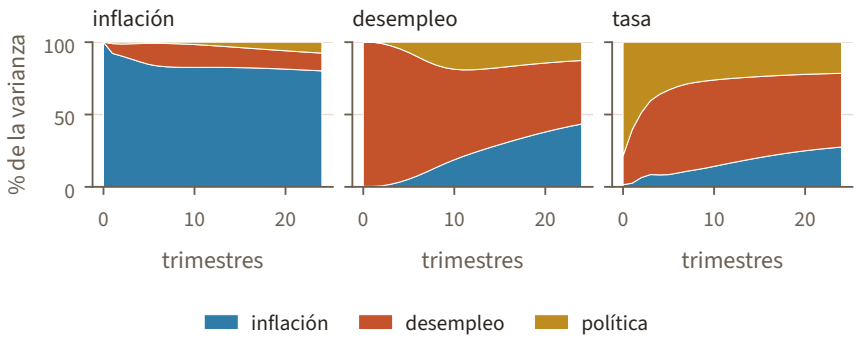


Figura 4.3. Descomposición de la varianza del error de pronóstico por horizonte. Cada panel reparte el 100 % de la varianza de una variable entre los tres choques.

El cuadro es el de Stock y Watson (2001) y sirve de verificación de la replicación. A doce trimestres el artículo reporta 82, 16 y 2 por ciento para la inflación y 16, 59 y 25 para la tasa de fondos federales, cifras que coinciden con las de arriba; para el desempleo reporta 16, 66 y 18 frente a 21, 60 y 19 aquí, una diferencia de unos pocos puntos que probablemente venga de la versión de la serie de desempleo, revisada desde 2001. Vale la pena registrar que una replicación puede ser correcta y aun así no coincidir en el último dígito.

La lectura económica es la misma en ambos casos. El choque de política explica alrededor de una quinta parte de la varianza del desempleo a tres años y casi nada de la inflación. Es poco, y es un resultado repetido en la literatura: los choques de política monetaria, entendidos como desviaciones respecto de la regla sistemática, explican una fracción pequeña de las fluctuaciones (Christiano et al. 1999; Ramey 2016). La afirmación no es que la política monetaria sea irrelevante, sino que casi todo lo que hace es sistemático, y lo sistemático no es un choque.

4.7. Descomposición histórica

La descomposición de varianza es un promedio sobre toda la muestra. La descomposición histórica baja al detalle de cada trimestre: qué choques explican que la inflación de 1980 haya estado donde estuvo.

DEFINICIÓN 4.3 · DESCOMPOSICIÓN HISTÓRICA

Usando la representación de Wold estructural, cada observación se escribe como la suma de las contribuciones de los choques pasados más un término determinista y la condición inicial:

$$\mathbf{y}_t = \underbrace{\text{base}_t}_{\text{media e inicio}} + \sum_{k=1}^{n} \underbrace{\sum_{j=0}^{t-1} \theta_{\cdot k,j}\, \varepsilon_{k,t-j}}_{\text{aporte del choque } k} . \tag{4.6}$$

Los choques se recuperan invirtiendo la matriz de impacto sobre los residuos estimados, $\hat{\varepsilon}_t = S^{-1}\hat{\mathbf{u}}_t$.

Desde cero (NumPy)

```
eps = np.linalg.solve(S, fit["residuals"].T).T
T_eff = eps.shape[0]
weights = responses(fit["B"], S, n, p, T_eff - 1)

contributions = np.zeros((T_eff, n, n))
for t in range(T_eff):
    for lag in range(t + 1):
        contributions[t] += weights[lag] * eps[t - lag]

baseline = fit["Y"] - contributions.sum(axis=2)
np.allclose(baseline + contributions.sum(axis=2), fit["Y"])
```

```
True
```

macrobook

```
contrib_pkg, base_pkg = svar.historical_decomposition(
    y, fit["B"], S, n, p)
np.allclose(contributions, contrib_pkg)
```

```
True
```

```
hd_dates = dates[p:]
pieces = {name: contributions[:, 0, k]
          for k, name in enumerate(shocks)}
observed = fit["Y"][:, 0] - baseline[:, 0]

size = figsize(0.42)
fig, ax = plt.subplots(figsize=size)
charts.stacked(ax, hd_dates, pieces, series=observed,
               series_label="desviación observada")
ax.set_ylabel("inflación (%)")
charts.legend_outside(ax, ncol=4)
plt.show()
```

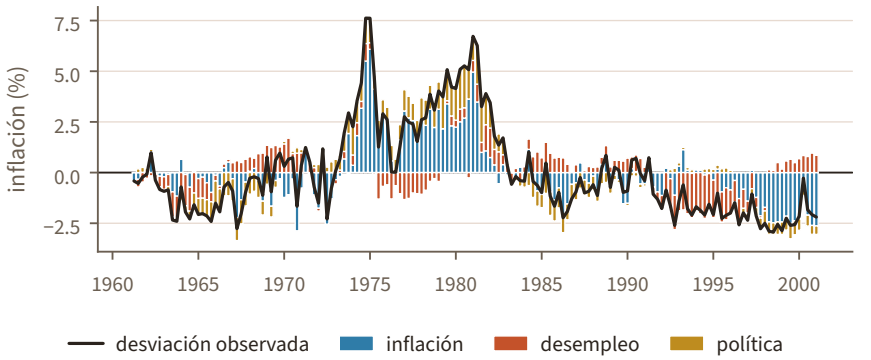


Figura 4.4. Descomposición histórica de la inflación: aporte de cada choque a la desviación respecto de la base del modelo. La línea es la desviación observada.

La Gran Inflación se lee con claridad. Entre 1979 y 1981 la inflación estuvo cinco puntos por encima de la base del modelo, y el gráfico reparte esa desviación entre el choque propio de los precios, que aporta entre dos y tres puntos y medio, y el choque de política, que aporta cerca de dos. Ese segundo término es una afirmación fuerte y hay que leerla con cuidado: el modelo dice que las desviaciones acumuladas respecto de la regla estimada, es decir una política sistemáticamente más laxa de lo que la propia regla habría indicado, sostuvieron alrededor de dos puntos de inflación. Hacia 1983 esa contribución desaparece, que es como se ve la desinflación de Volcker desde este modelo.

Conviene mirar también las décadas completas: el aporte medio del choque de política a la inflación es de +0.8 puntos en los setenta y de −0.2 en los noventa. Con los mismos coeficientes, el mismo choque cambia de papel según la época.

INTUICIÓN

Las tres lecturas de un SVAR responden preguntas distintas sobre el mismo objeto. La impulso-respuesta es un experimento: qué pasaría si. La descomposición de varianza es un promedio: cuánto importa este choque en general. La descomposición histórica es un relato: qué pasó en este trimestre concreto. Las tres se calculan con las mismas matrices Θ_j.

4.8. Restricciones de largo plazo

La identificación recursiva restringe el momento cero. Hay supuestos económicos que no tienen forma de restricción contemporánea pero sí de restricción sobre el efecto acumulado: la neutralidad del dinero en el largo plazo es el ejemplo canónico (Blanchard y Quah 1989; Galí 1999).

Si el VAR es estable, la suma de todas las respuestas de un choque converge. Sumando la Ecuación 4.4 sobre todos los horizontes,

$$\Theta_\infty = \sum_{h=0}^{\infty} \Psi_h S = \left(I_n - \Phi_1 - \cdots - \Phi_p\right)^{-1} S = A^{-1} S, \tag{4.7}$$

donde $A = I_n - \Phi_1 - \cdots - \Phi_p$ es conocida una vez estimada la dinámica. El elemento (i, k) de Θ_∞ es el efecto acumulado del choque k sobre la variable i, y sobre él se imponen los ceros.

PROPOSICIÓN 4.3 · IDENTIFICACIÓN POR RESTRICCIONES DE LARGO PLAZO

Si $\Theta_\infty = A^{-1}S$ es triangular inferior con diagonal positiva, entonces es el factor de Cholesky de

$$\Omega = \Theta_\infty \Theta_\infty' = A^{-1}\Sigma\,(A^{-1})',$$

que es observable. La matriz de impacto queda identificada como $S = A \cdot \mathrm{chol}(\Omega)$.

La matriz S que resulta no es triangular: los efectos contemporáneos quedan libres y las restricciones actúan sobre el acumulado. Ese es justamente el atractivo del esquema y también su costo, porque el efecto de largo plazo es la parte del modelo peor estimada.

El ejercicio clásico es el de Blanchard y Quah (1989), y aquí se replica con su propia muestra: crecimiento trimestral del PBI y tasa de desempleo de Estados Unidos entre 1948Q2 y 1987Q4, ambos ya sin tendencia como en el artículo, con ocho rezagos. La única restricción es que el segundo choque no tenga efecto permanente sobre el nivel del producto. El primero queda sin restringir y se interpreta como choque de oferta; el segundo, como choque de demanda.

```
bq = pd.read_csv(data_path("bq1989.csv"), index_col="date")
bq.index = pd.PeriodIndex(bq.index, freq="Q")
bq_order = ["growth", "unemp"]
y_bq = bq[bq_order].to_numpy()
p_bq = 8

fit_bq = var.ols(y_bq, p_bq)
S_lr = svar.long_run_impact(fit_bq["Sigma"], fit_bq["B"],
                            2, p_bq)
S_lr = S_lr * np.sign(S_lr[0])   # ambos elevan el producto

bq_shocks = ["oferta", "demanda"]
pd.DataFrame(S_lr.round(3), index=bq_order,
             columns=bq_shocks)
```

	oferta	demanda
growth	0.075	0.930
unemp	0.220	-0.208

La normalización de la última línea fija el signo de cada choque, que la identificación no determina: se eligen los dos choques que elevan el producto al impacto. Hecho eso, el resto lo dicen los datos, y lo que dicen es el resultado que hizo famoso al artículo. El choque de demanda eleva el producto y baja el desempleo; el choque de oferta eleva el producto pero **también** el desempleo durante los primeros trimestres. Ese segundo hecho es el llamado puzzle de la productividad, y reaparece en la literatura posterior (Galí 1999).

```
theta_lr = svar.responses(fit_bq["B"], S_lr, 2, p_bq, 40)
cumulated = np.cumsum(theta_lr, axis=0)
steps_bq = np.arange(theta_lr.shape[0])

size = figsize(0.40)
fig, axes = plt.subplots(1, 2, figsize=size)
for k, name in enumerate(bq_shocks):
    axes[0].plot(steps_bq, cumulated[:, 0, k], lw=1.2,
                 label=name, color=PALETTE["shocks"][k],
                 ls=PALETTE["shock_linestyle"][k])
    axes[1].plot(steps_bq, theta_lr[:, 1, k], lw=1.2,
                 label=name, color=PALETTE["shocks"][k],
                 ls=PALETTE["shock_linestyle"][k])
for ax, title in zip(axes, ["producto (nivel)", "desempleo"]):
    charts.zero_line(ax)
```

```
    ax.set_title(title, fontsize=8.5)
    ax.set_xlabel("trimestres")
charts.legend_outside(axes[0], ncol=2)
plt.show()
```

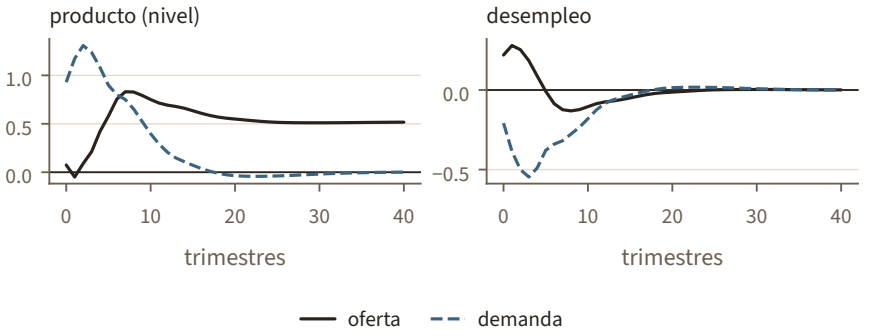


Figura 4.5. Réplica de Blanchard y Quah (1989): respuesta acumulada del nivel del producto y respuesta del desempleo. El choque de demanda no tiene efecto permanente sobre el producto por construcción.

El panel izquierdo reproduce la figura del artículo. El choque de oferta lleva el nivel del producto a un máximo cercano a 0.8 % y lo deja permanentemente medio punto por encima; el de demanda produce una joroba que se agota y vuelve a cero, que es la restricción impuesta. El panel derecho muestra el puzzle: el desempleo sube dos décimas tras un choque de oferta favorable antes de empezar a caer.

La crítica seria a este esquema no es la aritmética sino la estadística: A^{-1} depende de la suma de todos los coeficientes, que en muestras cortas se estima con poca precisión, y pequeñas diferencias en esa suma mueven mucho la matriz de impacto. Las restricciones de largo plazo son las más exigentes con los datos de los tres esquemas de este capítulo.

4.9. Restricciones de signo

Los dos esquemas anteriores imponen ceros, que son afirmaciones fuertes. Las restricciones de signo imponen algo más débil y a menudo más creíble: la dirección del efecto (Faust 1998; Canova y De Nicoló 2002; Uhlig 2005). Volviendo al sistema de Stock y Watson (2001), un choque contractivo de política monetaria sube la tasa de fondos federales y baja la inflación; sobre el desempleo no se impone nada, porque es precisamente lo que se quiere averiguar. Es la estrategia agnóstica de Uhlig (2005), que en su artículo deja libre la respuesta del producto.

La mecánica se apoya en el resultado de la Proposición 4.1. Si P es el factor de Cholesky de Σ y Q es ortonormal, entonces

$$\tilde{S} = PQ \quad \Longrightarrow \quad \tilde{S}\tilde{S}' = PQQ'P' = PP' = \Sigma,$$

de modo que toda $\tilde{S}$ así construida es un candidato válido. El conjunto de candidatos se recorre muestreando Q y quedándose con los que cumplen los signos.

Algoritmo 4.2. Identificación por restricciones de signo

1. Extraer una matriz ortonormal Q de dimensión $n \times n$ de manera uniforme, por ejemplo con la descomposición QR de una matriz de normales estándar.
2. Formar el candidato $\tilde{S} = PQ$ y calcular sus respuestas $\Theta_h = \Psi_h \tilde{S}$ en los horizontes donde se imponen los signos.
3. Verificar el patrón de signos, permitiendo multiplicar una columna por -1, porque ε y $-\varepsilon$ son el mismo choque con el nombre cambiado. Si se cumple, guardar $\tilde{S}$; si no, descartarla.
4. Repetir hasta reunir N matrices aceptadas y reportar el conjunto completo de respuestas.

```
signs = np.zeros((n, n))
signs[0, 2] = -1   # la inflación baja
signs[2, 2] = +1   # la tasa sube

accepted = svar.sign_restricted_impacts(
    fit["Sigma"], fit["B"], n, p, signs,
    horizons=(0, 1), draws=1000, seed=5)

set_irf = np.array([svar.responses(fit["B"], Si, n, p, 24)
                    for Si in accepted])
share = 100 * (set_irf[:, 8, 1, 2] > 0).mean()
print("matrices aceptadas:", len(accepted))
print("desempleo en h=8: sube en%.0f%% de los modelos"
      % share)
```

```
matrices aceptadas: 1000
desempleo en h=8: sube en 46% de los modelos
```

```
median = np.median(set_irf, axis=0)
q05 = np.quantile(set_irf, 0.05, axis=0)
q95 = np.quantile(set_irf, 0.95, axis=0)

size = figsize(0.40)
```

```
fig, axes = plt.subplots(1, 3, figsize=size)
steps = np.arange(set_irf.shape[1])
for i, (ax, name) in enumerate(zip(axes, names)):
    ax.fill_between(steps, q05[:, i, 2], q95[:, i, 2],
                    color=PALETTE["bands"][1], lw=0, alpha=0.5,
                    label="conjunto (5 a 95 %)")
    ax.plot(steps, median[:, i, 2], lw=1.2, label="mediana",
            color=PALETTE["accent_dark"])
    ax.plot(steps, theta_sd[:, i, 2], lw=1.0, ls=(0, (3, 2)),
            color=PALETTE["steel"], label="recursiva")
    charts.zero_line(ax)
    ax.set_title(name, fontsize=8.5)
    ax.set_xlabel("trimestres")
charts.legend_outside(axes[1], ncol=3)
plt.show()
```

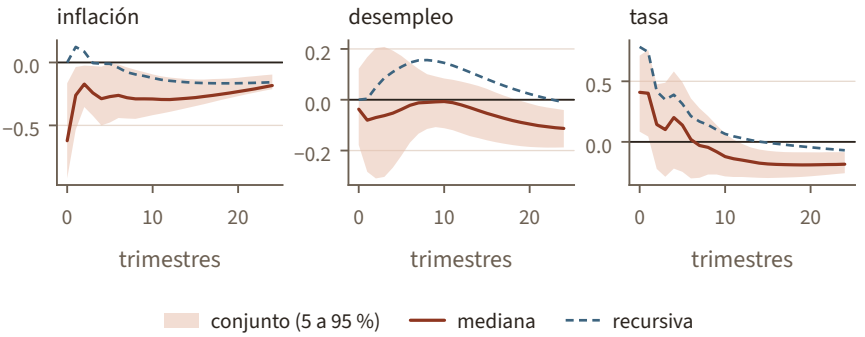


Figura 4.6. Respuestas al choque contractivo de política bajo restricciones de signo: mediana del conjunto identificado y rango del 5 al 95 %. La línea punteada es la respuesta recursiva. Las dos están en unidades de una desviación estándar del choque.

El resultado reproduce el hallazgo de Uhlig (2005) con los datos de Stock y Watson (2001): una vez que se deja libre la respuesta de la variable real, el conjunto de modelos admisibles no se pone de acuerdo sobre su signo. A dos años, sólo el 46 % de los modelos aceptados implica un aumento del desempleo, y el rango va de −0.12 a +0.10 puntos. La identificación recursiva daba un aumento nítido de dos décimas. La diferencia entre ambos resultados no está en los datos, que son los mismos, sino en cuánta información extra se estuvo dispuesto a imponer.

Conviene mirar también el panel izquierdo. La inflación baja durante los dos trimestres en que se le impuso bajar y se recupera apenas se suelta la restricción. Las restricciones de signo no arreglan el *price puzzle*: lo desplazan al primer horizonte no restringido. Imponerlas por más períodos lo empujaría más lejos, a costa de un supuesto cada vez más fuerte.

Una advertencia de escala: el tamaño del choque no es comparable entre esquemas. Cada matriz aceptada implica un impacto distinto sobre la tasa, de modo que la figura usa desviaciones estándar y no puntos porcentuales. Lo que sí es comparable, y lo que importa, es el signo y la fracción del conjunto que lo sostiene.

> **CUIDADO**
>
> Un conjunto de modelos no tiene mediana en ningún sentido útil. La mediana punto a punto de las respuestas aceptadas no corresponde en general a ninguna matriz de impacto del conjunto, así que reportarla como si fuera "el" resultado es un error frecuente (Fry y Pagan 2011). Además, muestrear Q uniformemente induce un prior sobre las respuestas que no es plano y que puede dominar al dato cuando las restricciones son pocas (Baumeister y Hamilton 2015). Lo honesto es reportar el rango y decir qué fracción del conjunto sostiene cada conclusión.

4.10. Proyecciones locales

Hay una manera distinta de estimar la misma respuesta: en lugar de iterar un modelo, correr una regresión por horizonte (Jordà 2005).

DEFINICIÓN 4.4 · PROYECCIÓN LOCAL

Dado un choque identificado $\hat{\varepsilon}_{kt}$, la respuesta de la variable i en el horizonte h es el coeficiente $\beta_{i,h}$ de

$$y_{i,t+h} = \alpha_{i,h} + \beta_{i,h}\,\hat{\varepsilon}_{kt} + \gamma_{i,h}'\mathbf{w}_t + \xi_{i,t+h}, \tag{4.8}$$

con $\mathbf{w}_t$ un vector de controles, típicamente los rezagos del sistema. Se estima una regresión por cada horizonte y los errores estándar se corrigen por autocorrelación.

Las dos estrategias estiman el mismo objeto poblacional: con los mismos controles y sin restricciones, VAR y proyecciones locales tienen el mismo límite (Plagborg-Møller y Wolf 2021). Difieren en muestras finitas, y la diferencia es un intercambio conocido. El VAR impone la estructura dinámica en todos los horizontes, lo que lo hace eficiente si el modelo está bien especificado y sesgado si no lo está; la proyección local no extrapola, lo que la hace robusta a la mala especificación y bastante más ruidosa a horizontes largos.

```
policy = eps[:, 2] * S[2, 2]   # en puntos porcentuales
shock_series = np.r_[np.full(p, np.nan), policy]
grid = np.arange(0, 17)
lp = np.array([svar.local_projection(y, shock_series, h,
                                     lags=p)
               for h in grid], dtype=object)
```

```
point = np.array([row[0][1] for row in lp])
low = np.array([row[1][1] for row in lp])
high = np.array([row[2][1] for row in lp])

size = figsize(0.42)
fig, ax = plt.subplots(figsize=size)
ax.fill_between(grid, low, high, color=PALETTE["bands"][1],
                lw=0, alpha=0.5,
                label="IC 95% (proyección local)")
ax.plot(grid, point, lw=1.2, color=PALETTE["accent_dark"],
        label="proyección local")
ax.plot(grid, theta[grid, 1, 2], lw=1.2, ls=(0, (4, 2)),
        color=PALETTE["ink"], label="VAR estructural")
charts.zero_line(ax)
ax.set_xlabel("trimestres")
ax.set_ylabel("desempleo (%)")
charts.legend_outside(ax, ncol=3)
plt.show()
```

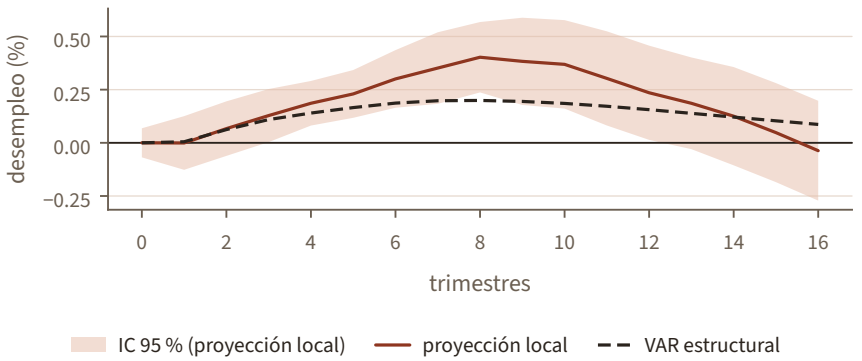


Figura 4.7. Respuesta del desempleo a un choque de política de un punto porcentual: VAR estructural y proyección local con intervalo al 95 %.

Las dos estimaciones coinciden en el signo y en el perfil: el desempleo no se mueve en el trimestre del choque, sube a lo largo de los dos años siguientes y después vuelve. Difieren en la magnitud del pico, que la proyección local sitúa en 0.40 puntos frente a 0.20 del VAR, y en la suavidad, porque la forma companion obliga a la respuesta del VAR a decaer de manera regular mientras que la proyección local no impone nada entre horizontes. Cuál de las dos creer depende de qué se esté dispuesto a suponer: la proyección local no extrapola y paga el precio en varianza, el VAR extrapola y cobra el precio en sesgo si la estructura que impone es falsa.

4.11. Otros esquemas y temas

Los tres esquemas anteriores cubren la mayor parte de la práctica, pero no agotan la literatura. Lo que sigue es un mapa mínimo, con la referencia donde estudiarlos.

4.11.1. Instrumentos externos

Si existe una variable observable correlacionada con el choque de interés y no con los demás, sirve de instrumento y evita restringir toda la matriz S: basta con identificar una columna. Los ejemplos típicos son las sorpresas de alta frecuencia alrededor de los anuncios de política monetaria y las series narrativas de cambios impositivos (Mertens y Ravn 2013; Stock y Watson 2018; Gertler y Karadi 2015). Es el enfoque dominante hoy cuando hay un instrumento creíble disponible.

4.11.2. Restricciones de signo con ceros y restricciones narrativas

Se pueden combinar signos con ceros exactos (Arias et al. 2018) o exigir que los choques recuperados sean consistentes con episodios históricos conocidos, por ejemplo que el choque de política de un trimestre determinado haya sido contractivo (Antolín-Díaz y Rubio-Ramírez 2019). Ambas cosas reducen el conjunto identificado con información que el investigador tiene y el modelo no.

4.11.3. Identificación por heterocedasticidad

Si la varianza de los choques cambia entre regímenes conocidos y la matriz de impacto no, los distintos regímenes aportan ecuaciones adicionales y la identificación puede lograrse sin restricciones de signo ni de cero (Rigobon 2003). El supuesto que se paga es la estabilidad de S entre regímenes.

4.11.4. Invertibilidad

Todo lo anterior supone que los choques estructurales se pueden recuperar de los residuos del VAR, es decir, que la información del econometrista incluye la de los agentes. Cuando no es así, por ejemplo con noticias sobre el futuro que los agentes ven y el sistema no, ningún esquema de identificación recupera el choque verdadero. El problema se conoce como no invertibilidad o no fundamentalidad, y es una razón adicional para preferir instrumentos externos cuando existen (Ramey 2016).

4.11.5. Tamaño del sistema y frecuencia

Los supuestos contemporáneos son más creíbles cuanto más corto el período: una restricción de que la política no afecta a los precios dentro del trimestre es cómoda, dentro del mes es más fuerte. Y los sistemas grandes, que permitirían incluir los controles que evitarían el *price puzzle*, chocan con el conteo de parámetros del Capítulo 3. Esa tensión es la que resuelve el Capítulo 5.

Ideas clave

1. Los datos determinan Σ y nada más: toda la variedad de matrices S con $SS' = \Sigma$ ajusta igual de bien, así que la identificación es un supuesto económico que se declara, se defiende y se somete a sensibilidad.
2. Faltan exactamente $n(n-1)/2$ restricciones. Los ceros contemporáneos las aportan con un orden, los ceros de largo plazo con la neutralidad de un choque y los signos con la dirección de un efecto.
3. Las tres lecturas del SVAR (respuestas, varianza, historia) se construyen con las mismas matrices $\Theta_j = \Psi_j S$ y responden preguntas distintas.
4. Las bandas bootstrap cubren la incertidumbre de estimación, no la de identificación. Las restricciones de signo hacen lo contrario: describen la incertidumbre de identificación con los parámetros fijos.
5. Proyecciones locales y VAR estiman la misma respuesta poblacional; la elección entre ambos es un intercambio entre eficiencia y robustez a la mala especificación.

Lecturas recomendadas

Kilian y Lütkepohl (2017) el tratamiento moderno y completo del análisis SVAR, con la inferencia hecha en serio.

Christiano, Eichenbaum y Evans (1999) la referencia sobre choques de política monetaria identificados recursivamente y sobre el *price puzzle*.

Blanchard y Quah (1989) el artículo que introdujo las restricciones de largo plazo.

Uhlig (2005) el enfoque agnóstico y la incomodidad de sus resultados.

Ramey (2016) el estado del arte sobre identificación de choques macroeconómicos, incluidos los instrumentos externos.

Plagborg-Møller y Wolf (2021) proyecciones locales y VAR estiman la misma respuesta poblacional.

Ejercicios

EJERCICIO 4.1

Reestime el SVAR recursivo con los seis órdenes posibles de las tres variables y grafique las seis respuestas del desempleo al choque que eleva la tasa. ¿Qué conclusiones sobreviven a todos los órdenes y cuáles dependen de uno en particular?

EJERCICIO 4.2

Verifique numéricamente la Proposición 4.1: genere cien matrices ortonormales Q, calcule $\tilde{S} = PQ$ en cada caso y confirme que $\tilde{S}\tilde{S}'$ reproduce $\hat{\Sigma}$ hasta el error de redondeo. Grafique las cien respuestas de la inflación al tercer choque.

EJERCICIO 4.3

La banda bootstrap del Algoritmo 4.1 usa el diseño recursivo. Implemente la variante equivocada que remuestrea los residuos sobre los valores ajustados sin iterar el modelo, y compare la persistencia de las respuestas medianas con la del diseño correcto. ¿En qué dirección se sesga?

EJERCICIO 4.4

Agregue un índice de precios de materias primas al sistema, ordenado en primer lugar, y repita la identificación recursiva. ¿Desaparece el *price puzzle*? Discuta si el resultado justifica el supuesto de que ese índice no responde contemporáneamente a nada.

EJERCICIO 4.5

Repita el capítulo con los datos peruanos del Capítulo 3, tomando actividad, inflación y tasa de referencia entre 2003Q4 y 2017Q4 con dos rezagos. Compare la respuesta del producto al choque de política con la del desempleo estadounidense y discuta qué diferencias vienen de los datos y cuáles del tamaño de la muestra.

EJERCICIO 4.6

En el ejercicio de restricciones de signo, agregue la exigencia de que el desempleo suba en los dos primeros trimestres. ¿Qué fracción de las matrices candidatas sobrevive ahora? Compare el conjunto resultante con la respuesta recursiva y explique por qué la comparación no es una prueba de nada.

EJERCICIO 4.7

Calcule la descomposición histórica del desempleo y localice el trimestre en que el choque de política hizo su mayor aporte. ¿Coincide con la recesión de 1981-1982?

CAPÍTULO 5

VAR bayesianos

LA PREGUNTA ECONÓMICA

¿Cómo se estima un sistema que tiene más parámetros de los que la muestra puede fijar con precisión?

1. Ver el encogimiento como lo que es: un prior que promedia información previa con datos, ponderando por precisión.
2. Escribir el prior de Minnesota, muestrear la posterior con Gibbs y leer lo que el encogimiento le hace a los coeficientes.
3. Producir pronósticos por densidad y evaluarlos fuera de muestra contra mínimos cuadrados y contra la caminata aleatoria.

Requisitos previos. Capítulo 3 y Capítulo 4; distribución normal multivariada e inversa-Wishart; producto de Kronecker y el operador vec; el repaso bayesiano del Apéndice B.

DATOS UTILIZADOS

El mismo sistema del Capítulo 4: inflación, desempleo y tasa de fondos federales de Estados Unidos, trimestrales de 1960Q1 a 2000Q4, cuatro rezagos, en `data/processed/sw2001.csv`. Mantener los datos fijos entre capítulos permite comparar lo que cambia por el método y no por la muestra.

5.1. Por qué encoger

El sistema de los dos capítulos anteriores tiene $k = np + 1 = 13$ coeficientes por ecuación y 39 en total, contra 160 observaciones utilizables. No es un caso extremo, y aun así basta para que mínimos cuadrados pronostique peor que una regla trivial.

```
import numpy as np
import pandas as pd

from macrobook import data_path, var, svar, bvar
```

```
csv = data_path("sw2001.csv")
data = pd.read_csv(csv, index_col="date")
data.index = pd.PeriodIndex(data.index, freq="Q")

order = ["infl", "unemp", "ff"]
sample = data[order]
dates = sample.index.to_timestamp(how="end")
y = sample.to_numpy()
n, p = len(order), 4
k = n * p + 1

print("coeficientes por ecuación:", k)
print("coeficientes del sistema:", n * k)
print("observaciones utilizables:", len(y) - p)
```

```
coeficientes por ecuación: 13
coeficientes del sistema: 39
observaciones utilizables: 160
```

```
horizon = 8
start = sample.index.get_loc(pd.Period("1974Q4")) + 1
errors = {"MCO": [], "caminata": []}

for t in range(start, len(y) - horizon + 1):
    train, actual = y[:t], y[t:t + horizon]
    fit_t = var.ols(train, p)
    path, _ = var.forecast(train, fit_t["B"], fit_t["Sigma"],
                           p, horizon)
    errors["MCO"].append(actual - path)
    errors["caminata"].append(actual - train[-1])

rmse = {key: np.sqrt((np.array(err) ** 2).mean(axis=0))
        for key, err in errors.items()}
ratio = rmse["MCO"] / rmse["caminata"]
pd.DataFrame(ratio[[0, 3, 7]].round(2), index=[1, 4, 8],
             columns=order)
```

	infl	unemp	ff
1	1.19	0.88	1.16
4	1.24	0.97	1.34
8	1.44	0.84	1.35

El cuadro compara la raíz del error cuadrático medio de un VAR estimado por mínimos cuadrados con la de una caminata aleatoria, en 97 ventanas que empiezan en 1975 y se extienden un trimestre cada vez. Un número mayor que uno significa que el VAR pierde. Pierde en casi todas las casillas, y pierde más cuanto más largo el horizonte. El modelo con más información es el que pronostica peor, porque estima demasiadas cosas con demasiado poco.

La salida no es dejar de modelar sino restringir. El enfoque bayesiano escribe la restricción como una distribución a priori: en lugar de fijar coeficientes en cero, los concentra alrededor de un valor con una dispersión que el investigador elige. El resultado es un estimador sesgado y mucho más preciso, que es un buen negocio cuando la varianza es el problema (Litterman 1986; Bańbura et al. 2010).

5.2. Prior, verosimilitud y posterior

El teorema de Bayes dice que la información se combina multiplicando:

$$\underbrace{p(b, \Sigma \mid Y)}_{\text{posterior}} \;\propto\; \underbrace{p(Y \mid b, \Sigma)}_{\text{verosimilitud}} \times \underbrace{p(b, \Sigma)}_{\text{prior}}. \tag{5.1}$$

Con distribuciones normales esa multiplicación tiene una forma reconocible. Si el prior de un escalar es $\theta \sim N(\theta_0, \tau^2)$ y la muestra aporta $\hat{\theta} \sim N(\theta, s^2)$, la posterior es normal con media

$$\bar{\theta} = \frac{\tau^{-2}\,\theta_0 + s^{-2}\,\hat{\theta}}{\tau^{-2} + s^{-2}}, \qquad \operatorname{Var}(\theta \mid Y) = \left(\tau^{-2} + s^{-2}\right)^{-1}. \tag{5.2}$$

La media posterior es un promedio ponderado por **precisión**, es decir por el inverso de la varianza. Nadie fija ese peso a mano: si la muestra es informativa, s^{-2} es grande y la posterior se pega al dato; si es ruidosa, manda el prior. Y un prior difuso, $\tau^2 \to \infty$, devuelve exactamente el estimador clásico.

```
import matplotlib.pyplot as plt

from macrobook import use_style, figsize, PALETTE, charts

use_style()

def normal(x, mu, sd):
    z = (x - mu) / sd
    return np.exp(-0.5 * z ** 2) / (sd * np.sqrt(2 * np.pi))

grid = np.linspace(-2.5, 3.5, 400)
```

```
prior = (0.0, 0.6)
cases = [("muestra informativa", (1.6, 0.35)),
         ("muestra ruidosa", (1.6, 1.2))]

size = figsize(0.34)
fig, axes = plt.subplots(1, 2, sharey=True, figsize=size)
for ax, (title, like) in zip(axes, cases):
    precision = 1 / prior[1] ** 2 + 1 / like[1] ** 2
    mean = (prior[0] / prior[1] ** 2
            + like[0] / like[1] ** 2) / precision
    post = (mean, precision ** -0.5)
    for (mu, sd), label, color in zip(
            [prior, like, post],
            ["prior", "verosimilitud", "posterior"],
            [PALETTE["muted"], PALETTE["steel"],
             PALETTE["accent"]]):
        ax.plot(grid, normal(grid, mu, sd), color=color,
                lw=1.3, label=label)
    ax.set_title(title, fontsize=8.5)
    ax.set_yticks([])
charts.legend_outside(axes[0], ncol=3)
plt.show()
```

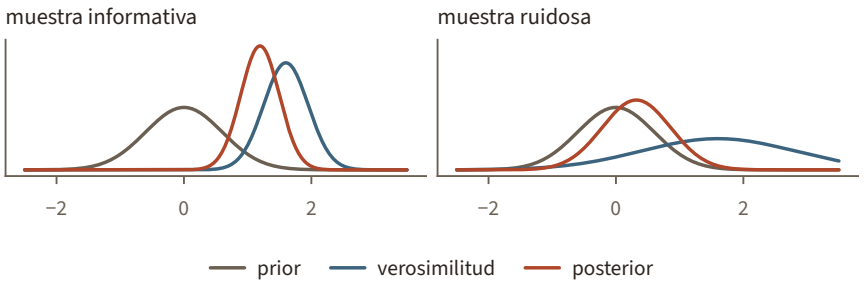


Figura 5.1. El mismo prior con dos muestras. A la izquierda la verosimilitud es angosta y la posterior se pega a ella; a la derecha es ancha y la posterior se queda cerca del prior.

INTUICIÓN

El encogimiento no es un truco de estimación: es la respuesta correcta cuando se tiene información previa y datos, y ninguno de los dos es perfecto. La discusión no es si usar información previa, porque elegir las variables y los rezagos ya lo es, sino si escribirla de manera explícita y auditable.

5.3. La forma vectorizada

El prior no se escribe sobre la matriz B sino sobre el vector de todos sus coeficientes, así que conviene apilar también las ecuaciones. Escribimos en minúsculas los objetos vectorizados,

$$y = \text{vec}(Y)\ (nT \times 1), \qquad b = \text{vec}(B)\ (nk \times 1), \qquad u = \text{vec}(U)\ (nT \times 1),$$

donde vec apila las columnas, es decir primero la ecuación 1 y después la 2. Aplicando vec a $Y = XB + U$ y usando $\text{vec}(XB) = (I_n \otimes X)\,\text{vec}(B)$,

$$y = (I_n \otimes X)\, b + u, \qquad \mathbb{E}[uu'] = \Sigma \otimes I_T. \tag{5.3}$$

Con dos variables, T observaciones y $k = np + 1$ coeficientes por ecuación, la Ecuación 5.3 se ve así:

$$\underbrace{\begin{bmatrix} y_{1,1:T} \\ y_{2,1:T} \end{bmatrix}}_{y\ (2T\times 1)} = \underbrace{\begin{bmatrix} X & 0_{T\times k} \\ 0_{T\times k} & X \end{bmatrix}}_{(I_2\otimes X)\ (2T\times 2k)} \underbrace{\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}}_{b\ (2k\times 1)} + \underbrace{\begin{bmatrix} u_{1,1:T} \\ u_{2,1:T} \end{bmatrix}}_{u\ (2T\times 1)}, \qquad \begin{aligned} b_1 &= (c_1,\ \phi^1_{11},\ \phi^1_{12},\ \ldots)' \\ b_2 &= (c_2,\ \phi^1_{21},\ \phi^1_{22},\ \ldots)' \end{aligned} \tag{5.4}$$

y la covarianza de u se escribe por bloques,

$$\mathbb{E}[uu'] = \Sigma \otimes I_T = \begin{bmatrix} \sigma_{11} I_T & \sigma_{12} I_T \\ \sigma_{12} I_T & \sigma_{22} I_T \end{bmatrix}, \tag{5.5}$$

que dice dos cosas a la vez: dentro de cada ecuación los errores no están correlacionados en el tiempo, y entre ecuaciones sí lo están en el mismo periodo.

Sobre este vector b actúa todo lo que viene: el prior le pone una normal con media b_0 y varianza H, y la posterior de la Sección 5.5 es una normal en ese mismo espacio.

5.4. El prior de Minnesota

Hace falta elegir b_0 y H, es decir escribir qué se cree antes de mirar los datos. La propuesta que se impuso (Litterman 1986) parte de una observación simple: una serie macroeconómica se parece mucho más a su propio pasado que a un sistema donde todas las variables importan con todos sus rezagos.

DEFINICIÓN 5.1 · PRIOR DE MINNESOTA

La media a priori del coeficiente del primer rezago propio de la variable i es δ_i,

y la de todos los demás coeficientes de rezago es cero. La varianza a priori del coeficiente del rezago ℓ de la variable j en la ecuación i es

$$\operatorname{Var}\left[(\Phi_\ell)_{ij}\right] = \begin{cases} \left(\dfrac{\lambda_1}{\ell^{\lambda_3}}\right)^2, & i = j, \\ \left(\dfrac{\lambda_1 \lambda_2 \, \sigma_i}{\ell^{\lambda_3} \, \sigma_j}\right)^2, & i \neq j, \end{cases} \tag{5.6}$$

donde σ_i es la desviación estándar del residuo de una regresión AR(p) univariada de la variable i. La constante recibe una varianza difusa, $(\sigma_i \lambda_4)^2$ con λ_4 grande.

Los cuatro hiperparámetros tienen lectura directa. λ_1 fija la intensidad general del encogimiento: con $\lambda_1 \to 0$ el prior manda y el sistema se vuelve n caminatas aleatorias independientes; con $\lambda_1 \to \infty$ el prior desaparece y queda mínimos cuadrados. $\lambda_2 \leq 1$ castiga más a los rezagos de las otras variables que a los propios. λ_3 apaga los rezagos lejanos. El cociente σ_i / σ_j sólo corrige unidades entre ecuaciones.

> **CUIDADO**
>
> $\delta_i = 1$ dice que la mejor descripción a priori de la serie es una caminata aleatoria, y es razonable para variables muy persistentes como las tres de este capítulo. Con tasas de crecimiento, que revierten rápido, corresponde $\delta_i = 0$ o un valor intermedio: fijar $\delta_i = 1$ impone una persistencia que la serie no tiene y arrastra los pronósticos de largo plazo.

Desde cero (NumPy)

```
sigma = bvar.ar_residual_sd(y, p)
lam = (0.2, 0.5, 1.0, 1e5)
l1, l2, l3, l4 = lam

B0 = np.zeros((k, n))
V = np.zeros((k, n))
for i in range(n):                    # ecuación i
    V[0, i] = (sigma[i] * l4) ** 2    # constante difusa
    for lag in range(1, p + 1):
        for j in range(n):            # variable j
            row = 1 + (lag - 1) * n + j
            if i == j:
                if lag == 1:
                    B0[row, i] = 1.0
                V[row, i] = (l1 / lag ** l3) ** 2
            else:
```

```
                V[row, i] = (l1 * l2 * sigma[i]
                             / (lag ** l3 * sigma[j])) ** 2

b0 = B0.flatten(order="F")
H = np.diag(V.flatten(order="F"))
np.sqrt(V[:5, 2]).round(3)   # desvíos a priori, ecuación de ff
```

```
array([9.7592521e+04, 9.3000000e-02, 3.9200000e-01, 2.0000000e-01,
       4.6000000e-02])
```

macrobook

```
b0_pkg, H_pkg = bvar.minnesota_prior(y, p, lam=lam, delta=1.0)
np.allclose(b0, b0_pkg), np.allclose(H, H_pkg)
```

```
(True, True)
```

La última línea imprime la desviación estándar a priori de los primeros coeficientes de la ecuación de la tasa: enorme para la constante, 0.2 para su propio primer rezago y alrededor de 0.1 para los rezagos de las otras variables. El prior no decreta que un coeficiente sea cero: dice que, a falta de evidencia en contra, se parecerá a cero.

5.5. De la prior a la posterior

Falta un prior para Σ. La elección natural es una inversa-Wishart, $\Sigma \sim \mathcal{IW}(S_0, \nu_0)$, que es la familia conjugada para una covarianza. Con un prior normal para b **independiente** del de Σ, la posterior conjunta no tiene forma cerrada, pero las dos condicionales sí, y eso basta para muestrear.

El paso 2 es la Ecuación 5.2 escrita con matrices: H^{-1} es la precisión del prior, $\Sigma^{-1} \otimes X'X$ la del dato, y la media posterior es el promedio de ambas fuentes ponderado por esas precisiones. El paso 3 suma a la escala del prior la suma de cuadrados de los residuos del sorteo anterior. Ninguno de los dos pasos requiere aproximación: ambas condicionales se muestrean de forma exacta.

Desde cero (NumPy)

```
from scipy.stats import invwishart

Y, X = var.lag_matrix(y, p)
T_eff = len(Y)
```

Algoritmo 5.1. Muestreador de Gibbs para el prior normal e inversa-Wishart independiente

1. Empezar con $\Sigma^{(0)} = \hat{\Sigma}_{\text{MCO}}$.
2. Extraer los coeficientes dada la covarianza:

$$b \mid \Sigma, Y \sim N(\bar{b}, \bar{V}), \qquad \bar{V} = \left[H^{-1} + \Sigma^{-1} \otimes X'X\right]^{-1},$$
$$\bar{b} = \bar{V}\left[H^{-1}b_0 + (\Sigma^{-1} \otimes X')\,y\right].$$

3. Extraer la covarianza dados los coeficientes:

$$\Sigma \mid b, Y \sim \mathcal{IW}\left(S_0 + (Y - XB)'(Y - XB),\ \nu_0 + T\right).$$

4. Repetir M veces, descartar las primeras extracciones (el *burn-in*) y quedarse con el resto.

```
S0, nu0 = np.eye(n), n + 1
H_inv = np.linalg.inv(H)
XtX = X.T @ X
vec_Y = Y.flatten(order="F")

rng = np.random.default_rng(11)
Sigma = var.ols(y, p)["Sigma"]
kept_b, kept_S = [], []
for step in range(6000):
    Sigma_inv = np.linalg.inv(Sigma)
    V = np.linalg.inv(H_inv + np.kron(Sigma_inv, XtX))
    mean = V @ (H_inv @ b0
                + np.kron(Sigma_inv, X.T) @ vec_Y)
    b = rng.multivariate_normal(mean, (V + V.T) / 2)
    B = b.reshape((k, n), order="F")
    if not var.is_stable(B, n, p):
        continue
    E = Y - X @ B
    Sigma = invwishart.rvs(df=nu0 + T_eff, scale=S0 + E.T @ E,
                           random_state=rng)
    if step ⩾ 2000:
        kept_b.append(b)
        kept_S.append(Sigma)

draws_b = np.array(kept_b)
draws_S = np.array(kept_S)
```

```
print("extracciones conservadas:", len(draws_b))
```

```
extracciones conservadas: 3967
```

macrobook

```
draws_b2, draws_S2 = bvar.gibbs(y, p, b0, H, draws=6000,
                                burn=2000, seed=11)
draws_b2.shape == draws_b.shape
```

```
True
```

El descarte de las extracciones explosivas merece un comentario. Cada sorteo de b define un sistema que puede ser estable o no; los inestables no tienen media incondicional ni impulso-respuesta que converja, así que se descartan. La práctica es estándar y equivale a un prior que pone probabilidad cero a la región explosiva (Blake y Mumtaz 2017).

```
labels = ["inflación", "desempleo", "tasa"]
positions = {"constante": 2 * k,
             "π(-1)": 2 * k + 1,
             "u(-1)": 2 * k + 2}

size = figsize(0.34)
fig, axes = plt.subplots(1, 3, figsize=size)
for ax, (name, pos) in zip(axes, positions.items()):
    charts.trace(ax, draws_b[:, pos])
    ax.set_title(name, fontsize=8.5)
    ax.set_xlabel("extracción")
plt.show()
```

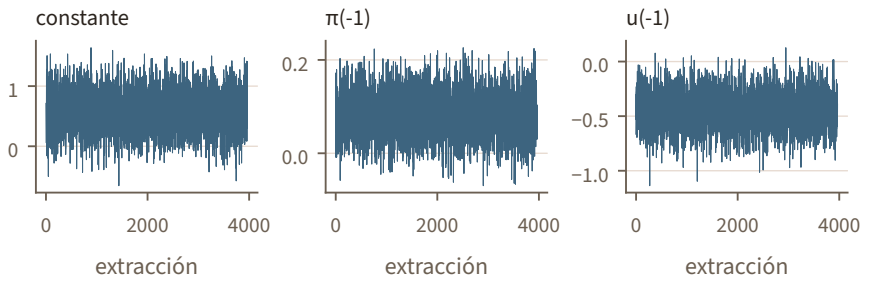


Figura 5.2. Trazas de tres coeficientes de la ecuación de la tasa. La cadena se mueve alrededor de un nivel estable, que es lo que se espera después del burn-in.

5.6. Qué hace el encogimiento

La comparación que importa es entre la media posterior y mínimos cuadrados.

```
B_post = draws_b.mean(axis=0).reshape((k, n), order="F")
fit = var.ols(y, p)

rows = ["constante"] + [f"{v}(-{lag})"
                        for lag in range(1, p + 1)
                        for v in ["π", "u", "R"]]
comparison = pd.DataFrame(
    {"MCO": fit["B"][:, 2], "posterior": B_post[:, 2]},
    index=rows).round(3)
comparison.head(7)
```

	MCO	posterior
constante	0.541	0.570
π(-1)	0.068	0.085
u(-1)	-1.643	-0.455
R(-1)	0.946	0.937
π(-2)	0.226	0.046
u(-2)	1.824	0.166
R(-2)	-0.367	-0.124

En la ecuación de la tasa de fondos federales, mínimos cuadrados le asigna al primer rezago del desempleo un coeficiente de -1.64, que en una regresión con trece coeficientes y ciento sesenta observaciones es en buena parte ruido. La posterior lo lleva a -0.45. El primer rezago propio de la tasa, en cambio, casi no se mueve: pasa de 0.95 a 0.94, porque ahí los datos son informativos y el prior no tiene nada que corregir. Eso es exactamente lo que debe hacer un prior bien puesto: encoger donde no hay información y apartarse donde la hay.

El encogimiento también se ve en las respuestas al impulso. Con las mismas extracciones de la posterior y la identificación recursiva del Sección 4.3, las bandas ya no son bootstrap sino cuantiles de la distribución posterior.

```
h_irf = 16
irf_draws = np.empty((len(draws_b), h_irf + 1, n, n))
for s, (b, S_draw) in enumerate(zip(draws_b, draws_S)):
    B_draw = b.reshape((k, n), order="F")
    impact = svar.unit_scale(np.linalg.cholesky(S_draw))
    irf_draws[s] = svar.responses(B_draw, impact, n, p, h_irf)
```

```
levels = [2.5, 16, 50, 84, 97.5]
q = np.percentile(irf_draws, levels, axis=0)
S_ols = svar.unit_scale(np.linalg.cholesky(fit["Sigma"]))
irf_ols = svar.responses(fit["B"], S_ols, n, p, h_irf)

steps = np.arange(h_irf + 1)
size = figsize(0.34)
fig, axes = plt.subplots(1, 3, figsize=size)
for i, (ax, name) in enumerate(zip(axes, labels)):
    ax.fill_between(steps, q[0, :, i, 2], q[4, :, i, 2],
                    color=PALETTE["bands"][0], lw=0, alpha=0.6,
                    label="95 %")
    ax.fill_between(steps, q[1, :, i, 2], q[3, :, i, 2],
                    color=PALETTE["bands"][2], lw=0, alpha=0.5,
                    label="68 %")
    ax.plot(steps, q[2, :, i, 2], lw=1.2, label="mediana",
            color=PALETTE["accent_dark"])
    ax.plot(steps, irf_ols[:, i, 2], lw=1.0, ls=(0, (3, 2)),
            color=PALETTE["ink"], label="MCO")
    charts.zero_line(ax)
    ax.set_title(name, fontsize=8.5)
    ax.set_xlabel("trimestres")
charts.legend_outside(axes[1], ncol=4)
plt.show()
```

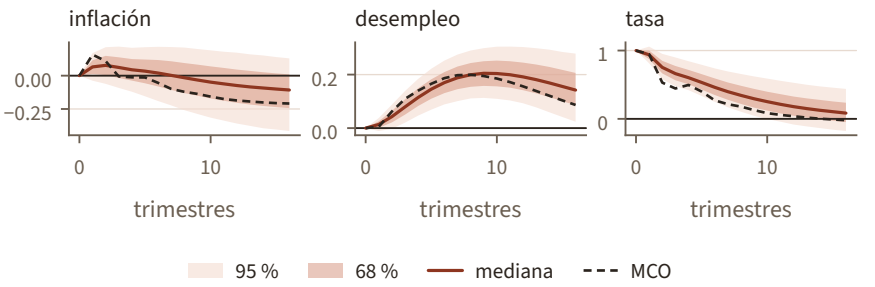


Figura 5.3. Respuesta a un choque de política de un punto porcentual: mediana posterior con bandas al 68 % y 95 %, y la respuesta por mínimos cuadrados.

La mediana posterior de la respuesta del desempleo alcanza 0.20 puntos a los dos años, prácticamente la misma cifra que mínimos cuadrados, con un intervalo al 68 % de 0.15 a 0.24 y uno al 95 % de 0.11 a 0.29. El encogimiento no movió el resultado central, y eso es informativo: donde los datos hablan, el prior no estorba. Lo que sí cambia es el resto del recorrido, que la posterior devuelve a cero de manera más suave, y el ancho de las bandas, más angostas que las bootstrap del Sección 4.5 porque el prior aporta información propia.

CUIDADO

Una banda posterior no es un intervalo de confianza. Dice que, dados el modelo, el prior y los datos, el parámetro está en ese rango con probabilidad 0.68 o 0.95. Es una afirmación más fuerte que la frecuentista y depende del prior: si el prior es informativo y equivocado, la banda será angosta y estará en el lugar equivocado. Por eso se reporta siempre qué prior se usó y se muestra la sensibilidad a sus hiperparámetros.

5.7. Pronóstico por densidad

El pronóstico bayesiano no es un número con una banda pegada después: es una distribución, la **densidad predictiva**, que se obtiene simulando hacia adelante con cada extracción de la posterior y con choques nuevos en cada trimestre.

Algoritmo 5.2. Densidad predictiva de un VAR bayesiano

Para cada extracción $(b^{(m)}, \Sigma^{(m)})$ de la posterior:

1. Poner las últimas p observaciones como estado inicial.
2. Para $j = 1, \dots, h$, extraer $\mathbf{u}_{T+j}^{(m)} \sim N(0, \Sigma^{(m)})$ e iterar el VAR con los coeficientes $b^{(m)}$.

Al final se tienen M trayectorias completas. Los percentiles de esas trayectorias en cada horizonte son las bandas del pronóstico.

La diferencia con el intervalo del Sección 3.9 es que aquí la incertidumbre de los parámetros está dentro: cada trayectoria usa un modelo distinto, extraído de la posterior. El intervalo frecuentista trataba $\hat{B}$ y $\hat{\Sigma}$ como los valores verdaderos, y por eso subestimaba la incertidumbre.

```
paths = bvar.predictive_draws(draws_b[::4], draws_S[::4],
                              y, p, h=12, seed=3)
qf = np.percentile(paths[:, :, 0], levels, axis=0)

hist = sample.iloc[-40:, 0]
hist_dates = hist.index.to_timestamp(how="end")
future = pd.period_range(sample.index[-1] + 1, periods=12,
                         freq="Q").to_timestamp(how="end")
edges = np.r_[hist_dates[-1:], future]
last = y[-1, 0]

size = figsize(0.42)
fig, ax = plt.subplots(figsize=size)
bands = [(np.r_[last, qf[0]], np.r_[last, qf[4]]),
```

```
        (np.r_[last, qf[1]], np.r_[last, qf[3]])]
charts.fan_chart(ax, hist_dates, hist, edges, bands,
                 np.r_[last, qf[2]],
                 band_labels=["95 %", "68 %"])
ax.axvline(hist_dates[-1], color=PALETTE["muted"],
           lw=0.8, ls=(0, (3, 2)), zorder=1)
ax.set_ylabel("inflación (%)")
charts.legend_outside(ax, ncol=4)
plt.show()
```

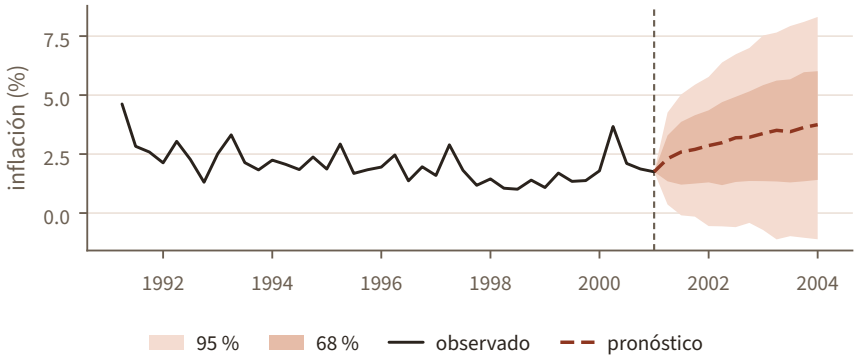


Figura 5.4. Densidad predictiva de la inflación a doce trimestres. Bandas al 68 % y 95 % de la distribución predictiva, que incorpora la incertidumbre sobre los parámetros.

La mediana sube de 1.7 % a cerca de 3.5 % en tres años. No es una predicción del modelo sobre la economía de 2001: es el resultado de que, con $\delta_i = 1$ y una muestra cuya inflación media es 3.9 %, la posterior sigue creyendo en un proceso persistente que vuelve despacio hacia la media histórica. Cambiar el prior cambia esa senda, y la sección siguiente muestra cuánto.

5.8. Elegir los hiperparámetros

Queda la pregunta incómoda: de dónde sale $\lambda_1 = 0.2$. Hay tres respuestas en uso, y conviene conocer las tres.

La primera es la tradición: Litterman probó valores en los años ochenta, funcionaron, y quedaron. La segunda es la evaluación fuera de muestra: elegir el valor que mejor pronostica en una parte de la muestra reservada para eso. La tercera, hoy la más defendible, es tratar los hiperparámetros como parámetros y ponerles su propio prior, de modo que los datos ayuden a elegirlos maximizando la verosimilitud marginal (Giannone et al. 2015).

La segunda es fácil de mostrar y explica la lógica de las otras dos.

```
lambdas = [0.05, 0.1, 0.2, 0.3, 0.5, 1.0, 5.0]
errors_lam = {value: [] for value in lambdas}

for t in range(start, len(y) - horizon + 1):
    train, actual = y[:t], y[t:t + horizon]
    for value in lambdas:
        prior = bvar.minnesota_prior(
            train, p, lam=(value, 0.5, 1.0, 1e5), delta=1.0)
        path = bvar.posterior_mean_forecast(
            train, p, prior[0], prior[1], horizon)
        errors_lam[value].append(actual - path)

relative = {value: (np.sqrt((np.array(err) ** 2).mean(axis=0))
                    / rmse["MCO"]).mean(axis=1)
            for value, err in errors_lam.items()}

size = figsize(0.42)
fig, ax = plt.subplots(figsize=size)
for j, step in enumerate([0, 3, 7]):
    ax.plot(lambdas, [relative[v][step] for v in lambdas],
            lw=1.2, marker=PALETTE["shock_marker"][j], ms=3.5,
            color=PALETTE["shocks"][j],
            ls=PALETTE["shock_linestyle"][j],
            label=f"h = {step + 1}")
ax.axhline(1.0, color=PALETTE["ink"], lw=0.8)
ax.set_xscale("log")
ax.set_xlabel("$\\lambda_1$ (escala logarítmica)")
ax.set_ylabel("RECM relativa a MCO")
charts.legend_outside(ax, ncol=3)
plt.show()
```

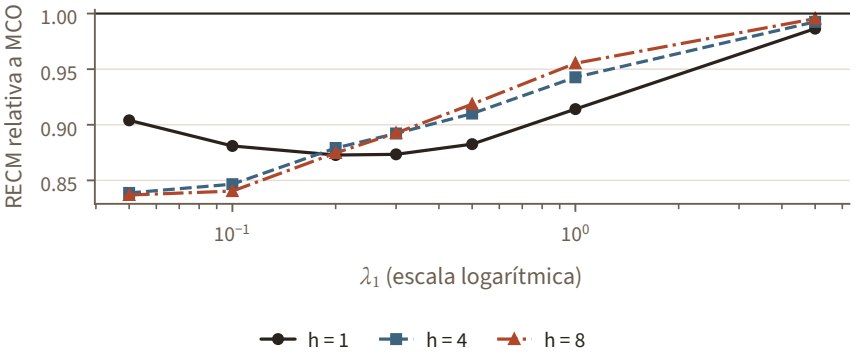


Figura 5.5. Error de pronóstico fuera de muestra según la intensidad del encogimiento, relativo a mínimos cuadrados. Cada línea es un horizonte; el promedio es sobre las tres variables y 97 ventanas.

La curva cuenta toda la historia del capítulo. Con $\lambda_1 = 5$ el prior es irrelevante y el BVAR pronostica como mínimos cuadrados, que es el punto de la derecha pegado a uno. A medida que el prior se ajusta, el error cae hasta un 12 o 16 % por debajo de mínimos cuadrados. A un trimestre el mínimo está alrededor de $\lambda_1 = 0.2$ y a horizontes largos el error sigue bajando con priors más ajustados, porque a esa distancia el mejor pronóstico se parece cada vez más a la caminata aleatoria que el prior describe.

> **CUIDADO**
>
> Elegir λ_1 mirando el error de pronóstico de toda la muestra y reportar después ese mismo error como evidencia del método es hacer trampa dos veces con los mismos datos. Si los hiperparámetros se eligen fuera de muestra, la evaluación debe hacerse sobre un tramo que no se usó para elegirlos, o con una ventana móvil que reelija en cada paso con la información disponible hasta ese momento.

5.9. Otros temas

5.9.1. Priors conjugados y observaciones ficticias

Si el prior de b se escribe con la misma Σ que la verosimilitud, es decir $b \mid \Sigma \sim N(b_0, \Sigma \otimes \Omega_0)$, el par normal-Wishart es conjugado y la posterior tiene forma cerrada: no hace falta Gibbs y todo se calcula con una regresión aumentada. El truco práctico es escribir el prior como **observaciones ficticias** que se apilan debajo de los datos (Bańbura et al. 2010). La misma técnica permite añadir priors de suma de coeficientes y de cointegración ficticia, que disciplinan el comportamiento de largo plazo del sistema.

5.9.2. VAR grandes

Con veinte o cien variables el conteo de parámetros se vuelve absurdo y el encogimiento pasa de conveniente a indispensable. La receta es escalar la intensidad del prior con el tamaño del sistema, de modo que el ajuste dentro de muestra se mantenga constante al agregar variables (Bańbura et al. 2010; Giannone et al. 2015). Un BVAR grande bien encogido pronostica tan bien como los factores dinámicos y permite mirar todas las variables a la vez.

5.9.3. Análisis estructural bayesiano

Todo el Capítulo 4 se traslada sin cambios: para cada extracción de la posterior se identifica y se calcula la respuesta, y el resultado es una distribución posterior de respuestas en lugar de una banda bootstrap. Con restricciones de signo, la combinación es natural, porque el muestreo de Q y el de la posterior se hacen en el mismo bucle (Uhlig 2005; Arias et al. 2018).

5.9.4. Datos atípicos y la pandemia

Los modelos lineales con errores gaussianos tratan una caída de treinta puntos como evidencia sobre la dinámica normal del sistema, y por eso 2020 rompió la mayoría de los VAR estimados hasta entonces. Las soluciones van desde variables de volatilidad estocástica hasta el tratamiento explícito de los trimestres de pandemia como atípicos (Lenza y Primiceri 2022), y el Capítulo 9 retoma el punto.

Ideas clave

1. El problema del VAR sin restricciones no es el sesgo sino la varianza: con pocos datos por parámetro, mínimos cuadrados pronostica peor que una caminata aleatoria.
2. La posterior es un promedio ponderado por precisión entre prior y datos. Con prior difuso reproduce mínimos cuadrados, así que el enfoque clásico es un caso particular.
3. El prior de Minnesota escribe una idea económica concreta, que cada serie se parece a su propio pasado, con cuatro hiperparámetros de lectura clara.
4. El muestreador de Gibbs alterna dos condicionales exactas; el precio de la flexibilidad es simulación en lugar de fórmulas.
5. El pronóstico bayesiano es una densidad que incorpora la incertidumbre de los parámetros, y su ventaja sobre mínimos cuadrados se verifica fuera de muestra.

Lecturas recomendadas

Blake y Mumtaz (2017) el manual del CCBS con los algoritmos escritos paso a paso; la referencia más cercana al espíritu de este capítulo.

Litterman (1986) el registro original de cinco años pronosticando con un VAR bayesiano, y el origen del nombre del prior.

Bańbura, Giannone y Reichlin (2010) cómo escalar el encogimiento con el tamaño del sistema.

Giannone, Lenza y Primiceri (2015) los hiperparámetros como parámetros: prior jerárquico y verosimilitud marginal.

Karlsson (2013) la revisión enciclopédica de priors y algoritmos para VAR bayesianos.

Ejercicios

EJERCICIO 5.1

Muestre que, cuando todas las ecuaciones comparten los mismos regresores, mínimos cuadrados ecuación por ecuación coincide con mínimos cuadrados generalizados sobre el sistema apilado.

SOLUCIÓN

Apilando por columnas, $\operatorname{vec}(Y) = (I_n \otimes X)\operatorname{vec}(B) + \operatorname{vec}(U)$ con $\operatorname{Var}[\operatorname{vec}(U)] = \Sigma \otimes I_T$. El estimador de MCG es

$$\hat{b}_{MCG} = \left[(I_n \otimes X)'(\Sigma^{-1} \otimes I_T)(I_n \otimes X)\right]^{-1}(I_n \otimes X)'(\Sigma^{-1} \otimes I_T)\operatorname{vec}(Y).$$

Usando $(A \otimes B)(C \otimes D) = AC \otimes BD$, el primer factor es $\Sigma^{-1} \otimes X'X$ y el segundo $(\Sigma^{-1} \otimes X')\operatorname{vec}(Y)$. Entonces

$$\hat{b}_{MCG} = (\Sigma \otimes (X'X)^{-1})(\Sigma^{-1} \otimes X')\operatorname{vec}(Y) = (I_n \otimes (X'X)^{-1}X')\operatorname{vec}(Y),$$

que es exactamente $\operatorname{vec}\left((X'X)^{-1}X'Y\right)$: MCO ecuación por ecuación. Σ desaparece.

EJERCICIO 5.2

Verifique numéricamente que un prior difuso devuelve mínimos cuadrados: multiplique H por 10^6, vuelva a muestrear y compare la media posterior con $\hat{B}$. ¿Cuánto hay que aflojar el prior para que la diferencia baje del uno por ciento?

EJERCICIO 5.3

Repita el Algoritmo 5.1 con $\delta_i = 0$ en lugar de $\delta_i = 1$ y compare la densidad predictiva de la inflación. ¿Hacia dónde revierte cada una y por qué?

EJERCICIO 5.4

Estudie la convergencia de la cadena: grafique la media acumulada de tres coeficientes y calcule el tamaño de muestra efectivo con la autocorrelación de las extracciones. ¿Cuántas extracciones hacen falta para estabilizar la segunda cifra decimal?

EJERCICIO 5.5

Añada al sistema el crecimiento del PBI real y el agregado monetario M2, de modo que $n = 5$ y el sistema tenga 105 coeficientes. Compare la RECM fuera de muestra de mínimos cuadrados y del BVAR. ¿Crece la ventaja del encogimiento con el tamaño del sistema?

EJERCICIO 5.6

Implemente el prior conjugado con observaciones ficticias: apile bajo Y y X las filas que reproducen la media y la varianza de la Definición 5.1, estime por mínimos cuadrados sobre la muestra aumentada y compare con la media posterior del muestreador de Gibbs.

CAPÍTULO 6

Pronóstico y evaluación

LA PREGUNTA ECONÓMICA

¿Cómo se construye un pronóstico defendible y cómo se sabe, después, si sirvió?

Capítulo en preparación. El esqueleto de secciones fija el alcance y el orden de la exposición.

1. La diferencia entre pronóstico puntual y pronóstico por densidad, y por qué un banco central necesita el segundo.
2. Cómo montar una evaluación pseudo fuera de muestra con ventanas recursivas o móviles sin mirar el futuro.
3. Qué mide la RECM relativa y qué añade la prueba de Diebold y Mariano.
4. Cómo se construyen y se comunican los escenarios condicionales.

Requisitos previos. Capítulo 3 y Capítulo 5.

DATOS UTILIZADOS

Serie	Fuente	Frecuencia	Transformación
Panel macrofiscal del libro	BCRP	trimestral	var. % interanual

Los datos crudos se descargan a `data/raw/` y las series construidas quedan en `data/processed/`.

6.1. Qué es un buen pronóstico

Por escribir: función de pérdida, horizonte y el pronóstico óptimo bajo error cuadrático.

6.2. Pronóstico por densidad

Por escribir: de la distribución predictiva al gráfico de abanico.

6.3. Diseño de la evaluación

Por escribir: orígenes, ventanas recursivas y móviles, y el error de mirar el futuro.

6.4. Métricas

Por escribir: RECM y EAM relativos, pérdida logarítmica y CRPS.

6.5. Comparación formal

Por escribir: Diebold y Mariano, y sus límites con modelos anidados.

6.6. Calibración

Por escribir: PIT y qué hacer cuando las bandas son sistemáticamente estrechas.

6.7. Combinación de pronósticos

Por escribir: por qué el promedio suele ganar y cuándo no.

6.8. Escenarios condicionales

Por escribir: Waggoner y Zha: imponer una trayectoria a una variable y leer el resto.

Ideas clave

1. Sin un modelo de referencia, un número de RECM no dice nada.
2. La calibración de la densidad importa tanto como el error del punto central.
3. Un escenario condicional no es un pronóstico: es una respuesta a una pregunta de política, con su propio supuesto de identificación.

Lecturas recomendadas

Diebold y Mariano (1995) la prueba estándar de precisión predictiva comparada.
Waggoner y Zha (1999) pronósticos condicionales en modelos multivariados dinámicos.
Karlsson (2013) revisión del pronóstico con VAR bayesianos.

Ejercicios

EJERCICIO 6.1

Monte una evaluación pseudo fuera de muestra con 20 orígenes y compare BVAR, VAR por MCO y caminata aleatoria a horizontes 1 a 8.

EJERCICIO 6.2

Calcule el PIT de la densidad predictiva del BVAR y grafique su histograma. ¿Está bien calibrada?

CAPÍTULO 7

Modelos de espacio de estados

LA PREGUNTA ECONÓMICA

¿Cómo estimamos algo que importa para la política pero que nadie observa?

Capítulo en preparación. El esqueleto de secciones fija el alcance y el orden de la exposición.

1. Escribir un modelo macro en forma de estado y medida.
2. Cómo funciona el filtro de Kalman y qué produce en cada paso.
3. Cómo obtener la verosimilitud por descomposición del error de predicción y estimar hiperparámetros.
4. Cómo extraer trayectorias del estado con el suavizador y con el simulador de Carter y Kohn.

Requisitos previos. Capítulo 3, distribución normal multivariada y nociones bayesianas del Apéndice B.

DATOS UTILIZADOS

Serie	Fuente	Frecuencia	Transformación
PBI real	BCRP	trimestral	logaritmo
Series mensuales para nowcasting	BCRP	mensual	varias

Los datos crudos se descargan a `data/raw/` y las series construidas quedan en `data/processed/`.

7.1. La forma de estado y medida

Por escribir: la estructura común a casi todo lo que sigue en el libro.

7.2. El filtro de Kalman

Por escribir: predicción, actualización y la ganancia como ponderador entre señal y ruido.

7.3. Verosimilitud y estimación

Por escribir: descomposición del error de predicción; máxima verosimilitud e inferencia bayesiana.

7.4. Suavizado

Por escribir: estimar el estado con toda la muestra y por qué difiere del filtrado.

7.5. Simulación del estado

Por escribir: Carter y Kohn dentro de un muestreador de Gibbs.

7.6. Datos faltantes y frecuencias mixtas

Por escribir: la ventaja natural del espacio de estados y el nowcasting.

7.7. Modelos de factores dinámicos

Por escribir: muchas series, pocos factores comunes.

Ideas clave

1. El filtro de Kalman no es un método de estimación sino una recursión para la distribución del estado.
2. Filtrar y suavizar responden preguntas distintas: en tiempo real sólo se dispone del primero.
3. Frecuencias mixtas y datos faltantes dejan de ser un problema cuando el modelo está en forma de estado.

Lecturas recomendadas

Durbin y Koopman (2012) el tratamiento de referencia del espacio de estados.

Kim y Nelson (1999) filtros, suavizadores y cambio de régimen con aplicaciones macro.

Carter y Kohn (1994) el simulador del estado que usan los capítulos 9 y 10.

Ejercicios

EJERCICIO 7.1

Escriba un AR(2) en forma de estado y verifique que la verosimilitud del filtro coincide con la verosimilitud exacta.

EJERCICIO 7.2

Estime un modelo de componentes no observados para el PBI y compare el estado filtrado con el suavizado en los últimos ocho trimestres.

CAPÍTULO 8

Tendencias y ciclos

LA PREGUNTA ECONÓMICA

¿Cuánto del producto observado es capacidad y cuánto es ciclo, si la separación no se observa?

Capítulo en preparación. El esqueleto de secciones fija el alcance y el orden de la exposición.

1. Qué supuesto sobre la tendencia hace cada filtro, aunque no lo declare.
2. Por qué el filtro de Hodrick y Prescott genera dinámica espuria y qué propone Hamilton en su lugar.
3. Cómo estimar una brecha del producto con componentes no observados y cómo reportar su incertidumbre.
4. Por qué la brecha estimada en tiempo real se revisa tanto y qué significa eso para la política.

Requisitos previos. Capítulo 7.

DATOS UTILIZADOS

Serie	Fuente	Frecuencia	Transformación
PBI real	BCRP / FRED	trimestral	logaritmo × 100

Los datos crudos se descargan a `data/raw/` y las series construidas quedan en `data/processed/`.

8.1. Qué es una tendencia

Por escribir: definiciones operativas y por qué no hay una sola correcta.

8.2. Filtros mecánicos

Por escribir: Hodrick y Prescott, Baxter y King: qué imponen sin decirlo.

8.3. La crítica de Hamilton

Por escribir: la regresión a h pasos como alternativa y sus propios supuestos.

8.4. Descomposición de Beveridge y Nelson

Por escribir: tendencia estocástica y ciclo a partir del pronóstico de largo plazo.

8.5. Componentes no observados

Por escribir: el modelo UC univariado y su versión bivariada con la curva de Phillips.

8.6. Incertidumbre y revisiones

Por escribir: bandas del ciclo y el problema del punto final.

8.7. Lectura de política

Por escribir: qué se puede afirmar con una brecha estimada y qué no.

Ideas clave

1. Todo filtro es un modelo con supuestos implícitos: conviene hacerlos explícitos y compararlos.
2. La brecha del producto es una cantidad latente con bandas anchas; presentarla como un número es engañoso.
3. El problema del punto final es justamente el que importa para la política.

Lecturas recomendadas

Hamilton (2018) por qué no usar el filtro HP y qué hacer en su lugar.
Hodrick y Prescott (1997) la referencia original, para leer con la crítica al lado.
Beveridge y Nelson (1981) la descomposición basada en el pronóstico de largo plazo.

Ejercicios

EJERCICIO 8.1

Compare las brechas obtenidas con HP, Hamilton y un UC univariado. ¿Coinciden en fechar las recesiones?

EJERCICIO 8.2

Estime la brecha con datos hasta cada trimestre entre 2015 y 2019 y grafique la revisión posterior de cada estimación.

CAPÍTULO 9

Volatilidad cambiante

LA PREGUNTA ECONÓMICA

¿Qué hacemos cuando el tamaño típico de los choques cambia con el tiempo?

Capítulo en preparación. El esqueleto de secciones fija el alcance y el orden de la exposición.

1. Reconocer heterocedasticidad condicional en series macro y financieras.
2. Estimar modelos ARCH y GARCH y entender qué dinámica imponen.
3. Estimar volatilidad estocástica con el muestreador de mezcla de Kim, Shephard y Chib.
4. Incorporar volatilidad estocástica en un VAR y qué cambia en las bandas de pronóstico.

Requisitos previos. Capítulo 7 y Capítulo 5.

DATOS UTILIZADOS

Serie	Fuente	Frecuencia	Transformación
Inflación, tipo de cambio	BCRP	mensual	var. % y logaritmos

Los datos crudos se descargan a `data/raw/` y las series construidas quedan en `data/processed/`.

9.1. Evidencia de volatilidad cambiante

Por escribir: qué se ve en los residuos y qué prueba lo confirma.

9.2. ARCH y GARCH

Por escribir: especificación, estimación y límites en frecuencia macro.

9.3. Volatilidad estocástica

Por escribir: la volatilidad como estado latente en escala logarítmica.

9.4. El muestreador de mezcla

Por escribir: la aproximación de Kim, Shephard y Chib y su implementación.

9.5. VAR con volatilidad estocástica

Por escribir: residuos heterocedásticos en un sistema y su efecto sobre la densidad predictiva.

9.6. La pandemia otra vez

Por escribir: volatilidad estocástica frente a dummies: dos formas de tratar 2020.

Ideas clave

1. En frecuencia macro, la volatilidad estocástica suele ajustar mejor que GARCH y se integra naturalmente con el resto del libro.
2. Ignorar la volatilidad cambiante no sesga tanto el punto central como las bandas: el costo aparece en la densidad.
3. La Gran Moderación y 2020 son los dos episodios que ningún modelo homocedástico explica.

Lecturas recomendadas

Kim, Shephard y Chib (1998) el muestreador que hace tratable la volatilidad estocástica.

Clark (2011) densidades predictivas de BVAR con volatilidad estocástica.

Engle (1982) el artículo fundacional de ARCH.

Ejercicios

EJERCICIO 9.1

Estime un GARCH(1,1) y un modelo de volatilidad estocástica para la misma serie y compare las trayectorias de volatilidad.

EJERCICIO 9.2

Repita la evaluación de densidad del Capítulo 6 con y sin volatilidad estocástica.

CAPÍTULO 10

Parámetros que cambian en el tiempo

LA PREGUNTA ECONÓMICA

¿Cambió la economía o cambió sólo el tamaño de los choques?

Capítulo en preparación. El esqueleto de secciones fija el alcance y el orden de la exposición.

1. Especificar un VAR con coeficientes que evolucionan como caminata aleatoria.
2. Estimarlo combinando Carter y Kohn con el muestreador de volatilidad del capítulo anterior.
3. Distinguir cambios en los coeficientes de cambios en la varianza de los choques.
4. Cuándo un modelo de cambio de régimen responde mejor que la variación suave.

Requisitos previos. Capítulo 7 y Capítulo 9.

DATOS UTILIZADOS

Serie	Fuente	Frecuencia	Transformación
PBI, inflación, tasa de política	FRED	trimestral	logaritmos y tasas

Los datos crudos se descargan a `data/raw/` y las series construidas quedan en `data/processed/`.

10.1. Por qué permitir variación

Por escribir: cambios de régimen de política, aprendizaje y quiebres estructurales.

10.2. El TVP-VAR

Por escribir: coeficientes en caminata aleatoria y la forma de estado del sistema.

10.3. Estimación

Por escribir: el muestreador de Primiceri, con la corrección de Del Negro y Primiceri.

10.4. Coeficientes frente a volatilidad

Por escribir: el debate sobre qué cambió en la Gran Moderación.

10.5. Sobreajuste y encogimiento

Por escribir: cuántos estados libres puede sostener una muestra macro.

10.6. Cambio de régimen

Por escribir: modelos de Markov como alternativa discreta a la variación suave.

Ideas clave

1. Un TVP-VAR puede acomodar casi cualquier cosa: la disciplina viene de los priors sobre la varianza de los estados.
2. Buena parte de lo que parece cambio en la transmisión es cambio en el tamaño de los choques.
3. Variación suave y cambio de régimen son supuestos distintos sobre la misma pregunta, y se eligen por argumento económico.

Lecturas recomendadas

Primiceri (2005) el TVP-VAR con volatilidad estocástica aplicado a política monetaria.

Del Negro y Primiceri (2015) la corrección al orden del muestreador; imprescindible antes de programarlo.

Cogley y Sargent (2005) derivas y volatilidades en la posguerra de Estados Unidos.

Ejercicios

EJERCICIO 10.1

Estime un TVP-VAR y compare la respuesta al choque de política en dos fechas distantes. ¿La diferencia sobrevive a las bandas?

EJERCICIO 10.2

Fije la volatilidad y reestime. ¿Cuánto de la variación atribuida a los coeficientes desaparece?

CAPÍTULO 11

Aprendizaje automático para macroeconomía

LA PREGUNTA ECONÓMICA

¿Qué aportan los métodos de aprendizaje automático cuando hay cientos de predictores y ochenta observaciones?

Capítulo en preparación. El esqueleto de secciones fija el alcance y el orden de la exposición.

1. Ver ridge, lasso y elastic net como priors y conectar el encogimiento con el Capítulo 5.
2. Usar factores y métodos de reducción de dimensión con datos macro.
3. Validar con datos dependientes en el tiempo sin contaminar la muestra.
4. Interpretar un modelo no lineal sin atribuirle causalidad.

Requisitos previos. Capítulo 5 y Capítulo 6.

DATOS UTILIZADOS

Serie	Fuente	Frecuencia	Transformación
Panel amplio tipo FRED-MD	FRED	mensual	transformaciones estándar del panel

Los datos crudos se descargan a `data/raw/` y las series construidas quedan en `data/processed/`.

11.1. El problema de dimensión, otra vez

Por escribir: muchos predictores, pocas observaciones y mucha correlación.

11.2. Regularización como prior

Por escribir: ridge y la Minnesota; lasso y la doble exponencial.

11.3. Factores

Por escribir: componentes principales, modelos de factores y FAVAR.

11.4. Métodos no lineales

Por escribir: bosques aleatorios y boosting: qué ganan y qué pierden en macro.

11.5. Validación temporal

Por escribir: por qué la validación cruzada aleatoria no sirve aquí.

11.6. Interpretación

Por escribir: importancia de variables y dependencia parcial: descripción, no identificación.

11.7. Qué gana y qué no gana el aprendizaje automático

Por escribir: la evidencia comparada en pronóstico macro.

Ideas clave

1. Casi todo método de aprendizaje automático útil en macro es una forma de encogimiento, y el encogimiento ya estaba en el capítulo 5.
2. Las ganancias aparecen sobre todo en horizontes largos y en variables reales, y son modestas.
3. La interpretabilidad no sustituye la identificación: un modelo predictivo no responde preguntas de política.

Lecturas recomendadas

Goulet Coulombe et al. (2022) qué funciona y qué no al pronosticar macro con aprendizaje automático.

Medeiros et al. (2021) pronóstico de inflación en un entorno rico en datos.

Stock y Watson (2016) factores dinámicos y su lugar en la macroeconometría.

Ejercicios

EJERCICIO 11.1

Compare lasso, ridge y un BVAR en el mismo ejercicio de pronóstico. ¿Cuál gana y a qué horizonte?

EJERCICIO 11.2

Extraiga cuatro factores de un panel amplio y agréguelos a un VAR. ¿Mejora la respuesta al choque de política?

Referencias

Antolín-Díaz, Juan, y Juan F. Rubio-Ramírez. 2019. «Narrative Sign Restrictions for SVARs». *The American Economic Review* 109 (7): 2802-29.

Arias, Jonas E., Juan F. Rubio-Ramírez, y Daniel F. Waggoner. 2018. «Inference Based on Structural Vector Autoregressions Identified with Sign and Zero Restrictions: Theory and Applications». *Econometrica* 86 (2): 685-720.

Bańbura, Marta, Domenico Giannone, y Lucrezia Reichlin. 2010. «Large Bayesian Vector Auto Regressions». *Journal of Applied Econometrics* 25 (1): 71-92.

Baumeister, Christiane, y James D. Hamilton. 2015. «Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information». *Econometrica* 83 (5): 1963-99.

Beveridge, Stephen, y Charles R. Nelson. 1981. «A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components». *Journal of Monetary Economics* 7 (2): 151-74.

Blake, Andrew P., y Haroon Mumtaz. 2017. *Applied Bayesian Econometrics for Central Bankers*. Technical Handbook. Centre for Central Banking Studies, Bank of England.

Blanchard, Olivier Jean, y Danny Quah. 1989. «The Dynamic Effects of Aggregate Demand and Supply Disturbances». *The American Economic Review* 79 (4): 655-73.

Box, George E. P., y David R. Cox. 1964. «An Analysis of Transformations». *Journal of the Royal Statistical Society, Series B* 26 (2): 211-52.

Burns, Arthur F., y Wesley C. Mitchell. 1946. *Measuring Business Cycles*. National Bureau of Economic Research.

Canova, Fabio, y Gianni De Nicoló. 2002. «Monetary Disturbances Matter for Business Fluctuations in the G-7». *Journal of Monetary Economics* 49 (6): 1131-59.

Christiano, Lawrence J., Martin Eichenbaum, y Charles L. Evans. 1999. «Monetary Policy Shocks: What Have We Learned and to What End?» En *Handbook of Macroeconomics*, editado por John B. Taylor y Michael Woodford, vol. 1. Elsevier.

Cleveland, Robert B., William S. Cleveland, Jean E. McRae, y Irma Terpenning. 1990. «STL: A Seasonal-Trend Decomposition Procedure Based on Loess». *Journal of Official Statistics* 6 (1): 3-73.

Croushore, Dean, y Tom Stark. 2001. «A Real-Time Data Set for Macroeconomists». *Journal of Econometrics* 105 (1): 111-30.

Dagum, Estela Bee. 1980. *The X-11-ARIMA Seasonal Adjustment Method*. Nos. 12-564E. Statistics Canada.

Engle, Robert F., y C. W. J. Granger. 1987. «Co-Integration and Error Correction: Representation, Estimation, and Testing». *Econometrica* 55 (2): 251-76.

Eurostat. 2024. *ESS Guidelines on Seasonal Adjustment*. Manuals and guidelines. Publications Office of the European Union.

Faust, Jon. 1998. «The Robustness of Identified VAR Conclusions about Money». *Carnegie-Rochester Conference Series on Public Policy* 49: 207-44.

Findley, David F., Brian C. Monsell, William R. Bell, Mark C. Otto, y Bor-Chung Chen. 1998. «New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program». *Journal of Business & Economic Statistics* 16 (2): 127-52.

Fry, Renée, y Adrian Pagan. 2011. «Sign Restrictions in Structural Vector Autoregressions: A Critical Review». *Journal of Economic Literature* 49 (4): 938-60.

Galí, Jordi. 1999. «Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?» *The American Economic Review* 89 (1): 249-71.

Gertler, Mark, y Peter Karadi. 2015. «Monetary Policy Surprises, Credit Costs, and Economic Activity». *American Economic Journal: Macroeconomics* 7 (1): 44-76.

Giannone, Domenico, Michele Lenza, y Giorgio E. Primiceri. 2015. «Prior Selection for Vector Autoregressions». *The Review of Economics and Statistics* 97 (2): 436-51.

Gómez, Víctor, y Agustín Maravall. 1996. *Programs TRAMO and SEATS: Instructions for the User*. Working Paper No. 9628. Banco de España.

Hamilton, James D. 2018. «Why You Should Never Use the Hodrick-Prescott Filter». *The Review of Economics and Statistics* 100 (5): 831-43.

Harvey, Andrew C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Hodrick, Robert J., y Edward C. Prescott. 1997. «Postwar U.S. Business Cycles: An Empirical Investigation». *Journal of Money, Credit and Banking* 29 (1): 1-16.

Ivanov, Ventzislav, y Lutz Kilian. 2005. «A Practitioner's Guide to Lag Order Selection for VAR Impulse Response Analysis». *Studies in Nonlinear Dynamics & Econometrics* 9 (1).

Johansen, Søren. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

Jordà, Òscar. 2005. «Estimation and Inference of Impulse Responses by Local Projections». *The American Economic Review* 95 (1): 161-82.

Kilian, Lutz. 1998. «Small-Sample Confidence Intervals for Impulse Response Functions». *The Review of Economics and Statistics* 80 (2): 218-30.

Kilian, Lutz, y Helmut Lütkepohl. 2017. *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Lenza, Michele, y Giorgio E. Primiceri. 2022. «How to Estimate a Vector Autoregression after March 2020». *Journal of Applied Econometrics* 37 (4): 688-99.

Litterman, Robert B. 1986. «Forecasting with Bayesian Vector Autoregressions: Five Years of Experience». *Journal of Business & Economic Statistics* 4 (1): 25-38.

Lütkepohl, Helmut. 2005. *New Introduction to Multiple Time Series Analysis*. Springer.

Lütkepohl, Helmut, y Fang Xu. 2012. «The Role of the Log Transformation in Forecasting Economic Variables». *Empirical Economics* 42 (3): 619-38.

Mertens, Karel, y Morten O. Ravn. 2013. «The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States». *The American Economic Review* 103 (4): 1212-47.

Montiel Olea, José Luis, y Mikkel Plagborg-Møller. 2019. «Simultaneous Confidence Bands: Theory, Implementation, and an Application to SVARs». *Journal of Applied Econometrics* 34 (1): 1-17.

Nelson, Charles R., y Charles I. Plosser. 1982. «Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications». *Journal of Monetary Economics* 10 (2): 139-62.

Plagborg-Møller, Mikkel, y Christian K. Wolf. 2021. «Local Projections and VARs Estimate the Same Impulse Responses». *Econometrica* 89 (2): 955-80.

Ramey, Valerie A. 2016. «Macroeconomic Shocks and Their Propagation». En *Handbook of Macroeconomics*, editado por John B. Taylor y Harald Uhlig, vol. 2. Elsevier.

Rigobon, Roberto. 2003. «Identification through Heteroskedasticity». *The Review of Economics and Statistics* 85 (4): 777-92.

Sims, Christopher A. 1980. «Macroeconomics and Reality». *Econometrica* 48 (1): 1-48.

Sims, Christopher A., James H. Stock, y Mark W. Watson. 1990. «Inference in Linear Time Series Models with Some Unit Roots». *Econometrica* 58 (1): 113-44.

Sims, Christopher A., y Tao Zha. 1999. «Error Bands for Impulse Responses». *Econometrica* 67 (5): 1113-55.

Stock, James H., y Mark W. Watson. 2001. «Vector Autoregressions». *Journal of Economic Perspectives* 15 (4): 101-15.

Stock, James H., y Mark W. Watson. 2018. «Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments». *The Economic Journal* 128 (610): 917-48.

Uhlig, Harald. 2005. «What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure». *Journal of Monetary Economics* 52 (2): 381-419.

APÉNDICE A

Notación

Apéndice en preparación.

A.1. Convenciones generales

Símbolo	Significado
$\mathbf{y}_t$	vector $n \times 1$ de variables endógenas en el periodo t
y_t	PBI real (escalar); las variables escalares van sin negrita
Φ_ℓ	matriz $n \times n$ de coeficientes del rezago ℓ
$\mathbf{u}_t$	innovaciones de forma reducida, $\mathbf{u}_t \sim N(0, \Sigma)$
ε_t	choques estructurales, $\mathbb{E}[\varepsilon_t \varepsilon_t'] = I_n$
Σ	matriz de covarianzas de las innovaciones de forma reducida
B	matriz $k \times n$ de coeficientes en forma compacta, $k = np + 1$
b	$\mathrm{vec}(B)$, apilado por columnas (ecuación por ecuación)
T, n, p, h	tamaño de muestra, número de variables, rezagos y horizonte
$\otimes$	producto de Kronecker
F, J	matriz companion $np \times np$ y selector $n \times np$
Ψ_j	pesos de Wold de la forma reducida, $\Psi_j = JF^j J'$
μ	media incondicional, $(I_n - \Phi_1 - \cdots - \Phi_p)^{-1} c$

A.2. Datos y descomposición

Símbolo	Significado
τ_t, c_t, s_t, η_t	tendencia, ciclo, componente estacional e irregular de $\ln y_t$
T_t, C_t, S_t, I_t	los mismos componentes en la versión multiplicativa, $y_t = T_t C_t S_t I_t$
g_t	tasa de crecimiento ordinaria, $g_t = y_t / y_{t-1} - 1$
$\Delta \ln y_t$	diferencia logarítmica, $\ln(1 + g_t)$
Δ_s	diferencia de orden s, $\Delta_s x_t = x_t - x_{t-s}$

El ciclo c_t lleva subíndice temporal y no debe confundirse con el vector de constantes c de la forma reducida, que no lo lleva. El irregular es η_t y no ε_t, reservado para los choques estructurales.

A.3. Análisis estructural

Símbolo	Significado
S	matriz de impacto, $\mathbf{u}_t = S\varepsilon_t$ y $\Sigma = SS'$
Θ_j	respuesta estructural en el horizonte j, $\Theta_j = \Psi_j S$
Θ_∞	efecto acumulado de largo plazo, $A^{-1}S$ con $A = I_n - \sum_\ell \Phi_\ell$
Q	matriz ortonormal, $QQ' = I_n$, usada para recorrer el conjunto identificado
$\mathrm{FEVD}_{i,k}(h)$	parte de la varianza del error de pronóstico de i atribuida al choque k

A.4. Distribuciones

Símbolo	Significado
$N(\mu, \Omega)$	normal multivariada
$\mathcal{IW}(S, \nu)$	inversa-Wishart con matriz de escala S y ν grados de libertad
b_0, H	media y varianza a priori de b
S_0, ν_0	escala y grados de libertad a priori de Σ

A.5. Hiperparámetros del prior de Minnesota

Símbolo	Significado
λ_1	intensidad general del encogimiento
λ_2	encogimiento de los rezagos cruzados
λ_3	decaimiento con el rezago
λ_4	dispersión de los términos deterministas
δ_i	media a priori del primer rezago propio de la variable i

APÉNDICE B

Repaso bayesiano

Apéndice en preparación.

B.1. Qué se necesita saber

Por escribir: verosimilitud, prior, posterior y el papel de la constante de normalización; por qué en este libro casi nunca hace falta calcularla.

B.2. Distribuciones de uso frecuente

Por escribir: normal multivariada, inversa-Wishart, gamma inversa y la normal-inversa-Wishart conjugada del Capítulo 5.

B.3. Muestreo

Por escribir: Gibbs, Metropolis-Hastings dentro de Gibbs y el muestreo del estado en modelos de espacio de estados.

B.4. Diagnóstico de convergencia

Por escribir: trazas, media acumulada, tamaño de muestra efectivo y el diagnóstico de cadenas múltiples.

B.5. Elección de priors

Por escribir: priors informativos frente a difusos, sensibilidad y verosimilitud marginal.