# Vectores Autoregresivos desde una Perspectiva Bayesiana

## Macroeconometría Aplicada

Juan Guerra-Salas

Pontificia Universidad Católica de Chile

Segundo Semestre 2020

# Introducción

- Sacaré contenido de **Blake y Mumtaz (2017), cáp. 2**.
- Las técnicas Bayesianas que discutimos en el contexto de una regresión lineal pueden extenderse para modelos VAR.
- Los modelos VAR Bayesianos combinan información externa (*priors*) con la información contenida en los datos (*likelihood*), para obtener la distribución posterior de los parámetros.
- Las técnicas de muestreo que se usan en estimación Bayesiana permiten caracterizar fácilmente la incertidumbre alrededor de objetos como funciones impulso-respuesta.

# Contenido

1. El VAR a estimar.
2. *Priors* y *posteriors* condicionales de parámetros y matriz de varianza-covarianza: descripción general.
3. Algoritmo de muestreo de Gibbs.
4. Formulación de *priors*: descripción específica. El *Minnesota prior*.
5. Aplicación 1: *Shocks* monetarios en Chile. Identificación por estructura recursiva.
6. Aplicación 2: *Shocks* monetarios en Chile. Identificación por restricciones de signo.

# El VAR a estimar

- Consideren el siguiente VAR(p):

$$
\begin{aligned}
Y_t &= c + Y_{t-1}B_1 + Y_{t-2}B_2 + \ldots + Y_{t-P}B_P + v_t, \\
E(v_t' v_s) &= \Sigma \quad si \quad t = s, \\
E(v_t' v_s) &= 0 \quad si \quad t \neq s, \\
E(v_t) &= 0.
\end{aligned}
$$

- $Y_t$ es un vector $(1 \times N)$ de variables endógenas, $c$ es un vector $(1 \times N)$ de constantes, $B_i$ son matrices $(N \times N)$ de coeficientes, y $v_t$ es un vector de errores. Cada columna representa una ecuación del sistema.

- El VAR se puede escribir de manera más compacta:

$$Y_t = X_t B + v_t,$$

donde $X_t = [1 \; Y_{t-1} \; Y_{t-2} \; \ldots \; Y_{t-P}]$ es un vector de dimensión $[1 \times (N * P + 1)]$ y $B$ es una matriz de dimensión $[(N * P + 1) \times N]$.

# Ejemplo de la notación (1/2)

- Consideren un VAR bivariado con dos rezagos para la inflación $\pi_t$ y la tasa de política monetaria $R_t$.

$$\pi_t = c_\pi + b_{11}\pi_{t-1} + b_{12}R_{t-1} + d_{11}\pi_{t-2} + d_{12}R_{t-2} + v_{\pi t},$$
$$R_t = c_R + b_{21}\pi_{t-1} + b_{22}R_{t-1} + d_{21}\pi_{t-2} + d_{22}R_{t-2} + v_{Rt}.$$

- En términos de vectores y matrices, donde cada **columna** representa una ecuación, el sistema puede escribirse así:

$$\begin{bmatrix}\pi_t & R_t\end{bmatrix} = \begin{bmatrix}c_\pi & c_R\end{bmatrix} + \begin{bmatrix}\pi_{t-1} & R_{t-1}\end{bmatrix}\begin{bmatrix}b_{11} & b_{21}\\ b_{12} & b_{22}\end{bmatrix} + \begin{bmatrix}\pi_{t-2} & R_{t-2}\end{bmatrix}\begin{bmatrix}d_{11} & d_{21}\\ d_{12} & d_{22}\end{bmatrix} + \begin{bmatrix}v_{\pi t} & v_{Rt}\end{bmatrix}.$$

# Ejemplo de la notación (2/2)

- De manera más compacta, el sistema puede escribirse así:

$$\underbrace{\begin{bmatrix} \pi_t & R_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} 1 & \pi_{t-1} & R_{t-1} & \pi_{t-2} & R_{t-2} \end{bmatrix}}_{X_t} \underbrace{\begin{bmatrix} c_\pi & c_R \\ b_{11} & b_{21} \\ b_{12} & b_{22} \\ d_{11} & d_{21} \\ d_{12} & d_{22} \end{bmatrix}}_{B} + \underbrace{\begin{bmatrix} v_{\pi t} & v_{Rt} \end{bmatrix}}_{v_t},$$

$$Y_t = X_t B + v_t.$$

# Volviendo a nuestro VAR genérico...

$$Y_t = c + Y_{t-1}B_1 + Y_{t-2}B_2 + \ldots + Y_{t-P}B_P + v_t,$$
$$Y_t = X_t B + v_t.$$

- También puede escribirse así:

$$y_t = (I_N \otimes X_t)b + V_t.$$

- $I_N$ es una matriz identidad de dimensión $(N \times N)$.
- $y = vec(Y_t)$, $b = vec(B)$, $V = vec(v_t)$.
- *vec* es un operador que "vectoriza" una matriz, poniendo cada columna debajo de la anterior.
- $\otimes$ es el producto de Kronecker: $I_N \otimes X$ indica que cada elemento de $I_N$ se multiplica por toda la matriz $X$.
- Demostración en apéndice A.

# *Prior* y *posterior* de los parámetros del VAR (1/3)

- Nuestro objetivo es estimar los coeficientes $b$ y la matriz de varianza-covarianza $\Sigma$.
- El *prior* conjugado para los coeficientes $b$ es normal:

$$P(b) \sim N(\tilde{b}_0, H).$$

- $\tilde{b}_0$ es la media del *prior*; un vector $[N \times (N \times P + 1)] \times 1$.
- $H$ es una matriz cuyos elementos de la diagonal principal son la varianza del *prior*; $H$ tiene dimensión $[N \times (N \times P + 1)] \times [N \times (N \times P + 1)]$.
- Como para una regresión lineal (tema anterior del curso), se usa distribución posterior de $b$ condicional en $\Sigma$ para aproximar distribución posterior marginal de $b$.

# *Prior* y *posterior* de los parámetros del VAR (2/3)

- La distribución posterior de $b$ condicional en $\Sigma$ también es normal:

$$H(b|\Sigma, Y_t) \sim N(M^*, V^*),$$

$$\begin{aligned}
M^* &= \left(H^{-1} + \Sigma^{-1} \otimes X_t'X_t\right)^{-1} \left(H^{-1}\tilde{b}_0 + \Sigma^{-1} \otimes X_t'X_t\hat{b}\right), \\
V^* &= \left(H^{-1} + \Sigma^{-1} \otimes X_t'X_t\right)^{-1}.
\end{aligned}$$

- $\hat{b}$ contiene los estimadores OLS de los parámetros del VAR.
- La media de la *posterior* es un promedio ponderado de la media del *prior* $\tilde{b}_0$ y del estimador OLS $\hat{b}$.
- Noten el rol de la varianza del *prior*. Mientras más grande sea, menor será la importancia del *prior*.

# *Prior* y *posterior* de los parámetros del VAR (3/3)

- El *prior* conjugado de la matriz de varianza-covarianza de los residuos tiene una distribución *inverse Wishart* con matriz de escala $\bar{S}$ y grados de libertad $\alpha$:

$$P(\Sigma) = IW(\bar{S}, \alpha).$$

- La distribución *inverse Wishart* es un análogo multivariado de la distribución *inverse Gamma*.
- La *posterior* de $\Sigma$ condicional en $b$ también es *inverse Wishart*:

$$H(\Sigma | b, Y_t) \sim IW(\bar{\Sigma}, T + \alpha).$$

- $\bar{\Sigma}$ es la matriz de escala de la *posterior* condicional; $T + \alpha$ son los grados de libertad. $T$ es el número de observaciones.

$$\bar{\Sigma} = \bar{S} + (Y_t - X_t B)'(Y_t - X_t B).$$

- Recuerden que $B$ son los coeficientes del VAR en forma de una matriz de dimensión $(N \times P + 1) \times N$.

# El algoritmo de muestreo de Gibbs para un VAR

1. Formular *priors* para los coeficientes y la matriz de varianza-covarianza de los residuos: $P(b) \sim N(\tilde{b}_0, H)$, $P(\Sigma) \sim IW(\bar{S}, \alpha)$. Definir un valor inicial para $\Sigma$.
2. Tomar una muestra de la distribución posterior condicional de los coeficientes $H(b|\Sigma, Y_t) \sim N(M^*, V^*)$. Con $M^*$ y $V^*$ se obtiene una muestra de la distribución normal:

   $$b^1 = M^* + [\bar{b} \times (V^*)^{1/2}],$$

   donde $\bar{b}$ es una muestra de la distribución multivariada normal estándar.
3. Tomar una muestra de la dist. posterior condicional de $\Sigma$, $H(\Sigma|b, Y_t) \sim IW(\bar{\Sigma}, T + \alpha)$, usando la muestra previa de parámetros $b^1$ como insumo. Se obtiene $\Sigma^1$.
4. Repetir los dos pasos anteriores $M$ veces, lo que ofrece $B^1, \ldots, B^M$ y $\Sigma^1, \ldots, \Sigma^M$. Se usan los últimos $H$ valores de $B$ y $\Sigma$ para formar la distribución empírica de los parámetros, y para calcular otros objetos como pronósticos y funciones impulso-respuesta.

# Formulación de *priors* (1/6)

- Aquí revisamos exactamente cómo se definen los priors.
- Volvamos a nuestro ejemplo de dos variables y dos rezagos:

$$\begin{aligned}\begin{bmatrix}\pi_t & R_t\end{bmatrix} = \begin{bmatrix}c_\pi & c_R\end{bmatrix} + \begin{bmatrix}\pi_{t-1} & R_{t-1}\end{bmatrix}\begin{bmatrix}b_{11} & b_{21}\\ b_{12} & b_{22}\end{bmatrix} \\ + \begin{bmatrix}\pi_{t-2} & R_{t-2}\end{bmatrix}\begin{bmatrix}d_{11} & d_{21}\\ d_{12} & d_{22}\end{bmatrix} + \begin{bmatrix}v_{\pi t} & v_{Rt}\end{bmatrix}.\end{aligned}$$

- El *Minnesota prior* consiste en que las variables del sistema siguen un proceso AR(1) o de caminata aleatoria. Bajo la media de este *prior*, el sistema toma la siguiente forma:

$$\begin{aligned}\begin{bmatrix}\pi_t & R_t\end{bmatrix} = \begin{bmatrix}0 & 0\end{bmatrix} + \begin{bmatrix}\pi_{t-1} & R_{t-1}\end{bmatrix}\begin{bmatrix}b_{11}^0 & 0\\ 0 & b_{22}^0\end{bmatrix} \\ + \begin{bmatrix}\pi_{t-2} & R_{t-2}\end{bmatrix}\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix} + \begin{bmatrix}v_{\pi t} & v_{Rt}\end{bmatrix}.\end{aligned}$$

- $b_{11}^0$ y $b_{22}^0$ son los *priors* AR(1) si son menores que 1, o de caminata aleatoria si son iguales a 1.

# Formulación de *priors* (2/6)

- Por tanto, la media del prior de los parámetros $P(b) \sim N(\tilde{b}_0, H)$ toma la siguiente forma:

$$\tilde{b}_0 = \begin{bmatrix} 0 \\ b_{11}^0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ b_{22}^0 \\ 0 \\ 0 \end{bmatrix}.$$

- Las primeras 5 filas de este vector corresponden a los parámetros de la primera ecuación y las últimas 5 filas a los de la segunda ecuación.

# Formulación de *priors* (3/6)

- La varianza de este prior $H$ toma la siguiente forma (intuición más adelante):

$$H_{(10\times10)} = \begin{bmatrix} H_1 & 0_{(5\times5)} \\ 0_{(5\times5)} & H_2 \end{bmatrix},$$

$$H_1 = \begin{bmatrix} (\sigma_1\lambda_4)^2 & 0 & 0 & 0 & 0 \\ 0 & (\lambda_1)^2 & 0 & 0 & 0 \\ 0 & 0 & \left(\frac{\sigma_1\lambda_1\lambda_2}{\sigma_2}\right)^2 & 0 & 0 \\ 0 & 0 & 0 & (\frac{\lambda_1}{2^{\lambda_3}})^2 & 0 \\ 0 & 0 & 0 & 0 & \left(\frac{\sigma_1\lambda_1\lambda_2}{\sigma_2 2^{\lambda_3}}\right)^2 \end{bmatrix},$$

# Formulación de *priors* (4/6)

$$H_2 = \begin{bmatrix} (\sigma_2\lambda_4)^2 & 0 & 0 & 0 & 0 \\ 0 & \left(\frac{\sigma_2\lambda_1\lambda_2}{\sigma_1}\right)^2 & 0 & 0 & 0 \\ 0 & 0 & (\lambda_1)^2 & 0 & 0 \\ 0 & 0 & 0 & \left(\frac{\sigma_2\lambda_1\lambda_2}{\sigma_1 2^{\lambda_3}}\right)^2 & 0 \\ 0 & 0 & 0 & 0 & (\frac{\lambda_1}{2^{\lambda_3}})^2 \end{bmatrix}.$$

- $\sigma_1$, $\sigma_2$ son desvíos estándar de términos de error estimados por regresiones AR(1) (estimadas por OLS) para las variables 1 y 2. Se necesitan por si las variables tienen diferentes escalas.

# Formulación de *priors* (5/6)

- Los parámetros $\lambda_1$, $\lambda_2$, $\lambda_3$, $\lambda_4$ controlan la estrechez del *prior*, y se determinan, para cada coeficiente $b_{ij}$, así:

$$\left(\frac{\lambda_1}{l^{\lambda_3}}\right)^2 \quad si \quad i = j$$

$$\left(\frac{\sigma_i \lambda_1 \lambda_2}{\sigma_j l^{\lambda_3}}\right)^2 \quad si \quad i \neq j$$

  $(\sigma_i \lambda_4)^2$ para la constante.

- $i$ se refiere a la variable dependiente de la ecuación $i$-ésima; $j$ a la variable independiente de esa ecuación. Cuando $i = j$, se trata de coeficientes sobre rezagos de la misma variable $i$. $l$ se refiere al número de rezago asociado al coeficiente $b_{ij}$.

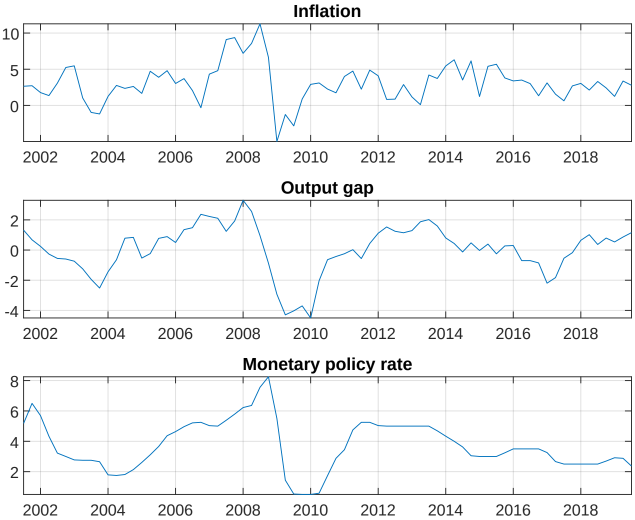
# Formulación de *priors* (6/6)

- $\lambda_1$ controla el desvío estándar del *prior* sobre los propios rezagos de cada variable. Cuando $\lambda_1 \to 0$, $b_{11}, b_{22} \to b_{11}^0, b_{22}^0$ y los parámetros asociados a otros rezagos $\to 0$.
- $\lambda_2$ controla el desvío estándar del *prior* sobre variables que no son la dependiente ($b_{12}, b_{21}$, etc.) $\lambda_2 \to 0$, $b_{ij}, d_{ij} \to 0$. Cuando $\lambda_2 = 1$, rezagos de variables que no son la dependiente en efecto se tratan igual que rezagos de la variable dependiente.
- $\lambda_3$ controla la medida en que coeficientes de rezagos mayores que 1 son 0. Mientras mayor es $\lambda_3$, más probable que estos coeficientes tomarán valores de 0. Noten que $l$ describe el rezago al que se asocia cada coeficiente.
- $\lambda_4$ controla la varianza del *prior* sobre la constante. Cuando $\lambda_4 \to 0$, las constantes $\to 0$.

# Aplicación 1: *Shocks* monetarios en Chile; identificación recursiva (1/4)

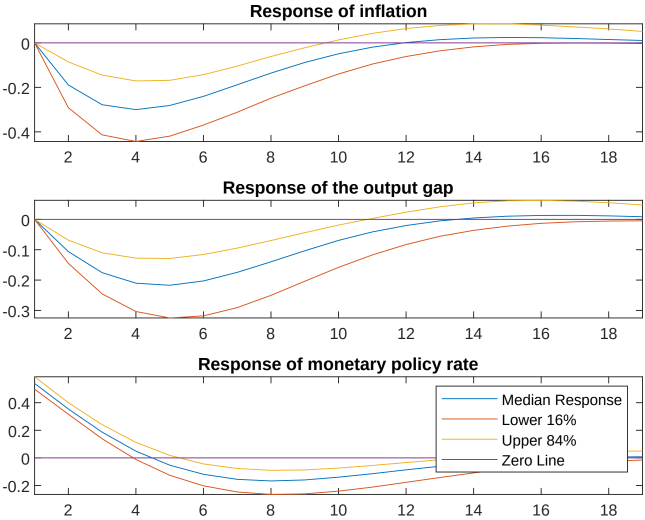
- En esta aplicación usamos datos sobre inflación, brecha del producto y tasa de política monetaria (TPM) para estimar el efecto de un *shock* a la TPM.
- La muestra cubre el periodo 2001:T3–2019:T3.
- Para identificar *shocks* estructurales, suponemos una estructura recursiva, según la cual: (i) la TPM responde de manera contemporánea a la inflación y la brecha del producto, (ii) la brecha responde de manera contemporánea a la inflación, pero no a la TPM, y (iii) la inflación no responde a ninguna de las otras dos variables de manera contemporánea.
- El VAR tiene un rezago.
- Como puede verse, el ejercicio es igual al discutido previamente en el curso, excepto por el método de estimación: Bayesiano en lugar de clásico.

- Los datos para estimar el modelo son los siguientes:



# Aplicación 1: *Shocks* monetarios en Chile; identificación recursiva (3/4)

- Respuesta a un *shock* de un desvío estándar a la TPM:



# Aplicación 1: *Shocks* monetarios en Chile; identificación recursiva (4/4)

- La estimación Bayesiana del modelo arroja resultados similares a los de la estimación clásica:
  - Un *shock* típico a la TPM (un desvío estándar) en esta muestra hace que aumente en torno a 50 puntos base en impacto.
  - La contracción monetaria genera caídas graduales de la inflación y la actividad económica.
  - El efecto máximo del *shock* de política monetaria sobre la inflación y brecha del producto y el rezago con el que ocurre es similar al encontrado mediante estimación clásica.
- La razón por la que los resultados de la estimación Bayesiana son similares a los de la estimación clásica es que los *priors* son muy poco informativos (son "bastante amplios").

# Aplicación 2: *Shocks* monetarios en Chile; identificación por restricciones de signo (1/5)

- La idea de la identificación por restricciones de signo es admitir funciones impulso-respuesta (IRFs) con signos que son consistentes con la teoría económica.
- Por ejemplo, para identificar un *shock* de política monetaria, sería deseable que la inflación y la brecha del producto respondan de manera negativa a un *shock* que genera un aumento de la TPM.
- El algoritmo para ejecutar esta identificación consiste en "alterar" los elementos de una descomposición de Cholesky, generar IRFs, y verificar si satisfacen las restricciones de signo.

## Aplicación 2: *Shocks* monetarios en Chile; identificación por restricciones de signo (2/5)

- Dentro del algoritmo de *Gibbs sampling*, después de pasar la etapa de iteraciones que se descartan (*burn-in*), los pasos que se siguen son los siguientes:

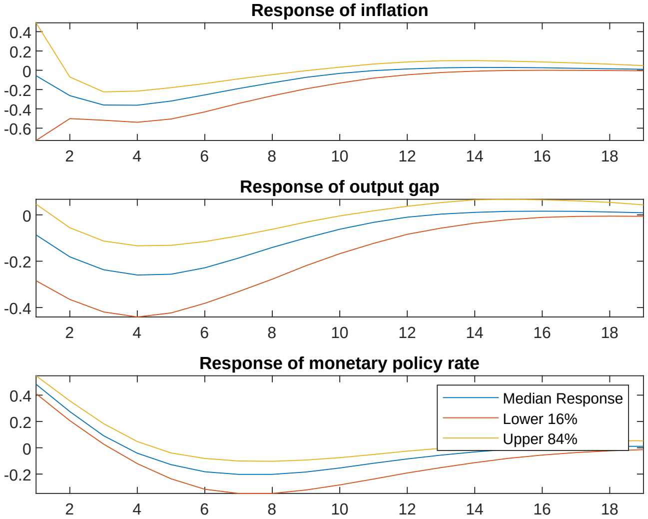
  1. De la distribución normal estándar, se obtiene una matriz $K$ de dimensión $N \times N$.
  2. Se calcula la matriz $Q$ de la descomposición QR de K. Según esta descomposición, $Q$ es ortonormal, es decir, $Q'Q = I$.
  3. Se calcula la descomposición de Cholesky ($\tilde{A}_0$) de la matriz de varianza-covarianza ($\Sigma$) de la iteración actual del algoritmo de *Gibbs*. La descomposición de Cholesky satisface $\Sigma = \tilde{A}_0'\tilde{A}_0$.
  4. Se calcula la matriz $A_0 = Q\tilde{A}_0$. Esta matriz con efectos contemporáneos se usa para generar IRFs (tal como se generan en el caso en que usamos solo la descomposición de Cholesky) y se evalúa si las IRFs satisfacen las restricciones de signo. Si lo hacen, se guardan. Si no, se descartan y se repiten estos pasos.

# Aplicación 2: *Shocks* monetarios en Chile; identificación por restricciones de signo (3/5)

- Noten que en este procedimiento se altera la descomposición de Cholesky $\tilde{A}_0$ multiplicándola por $Q$ para obtener $A_0$, pero como $Q$ es ortonormal, $A_0$ también satisface $\Sigma = A_0' A_0$, porque $A_0' A_0 = (Q\tilde{A}_0)' Q\tilde{A}_0 = \tilde{A}_0' Q' Q\tilde{A}_0 = \tilde{A}_0' I \tilde{A}_0 = \tilde{A}_0' \tilde{A}_0 = \Sigma$.
- En el análisis de *shocks* monetarios en Chile, vamos a retener las iteraciones cuyas IRFs satisfacen los siguientes criterios:
  - Un *shock* a la TPM hace que la TPM aumente en el periodo en que golpea a la economía y los siguientes dos periodos.
  - La inflación cae en respuesta a este *shock* en los dos trimestres después de que el *shock* golpea a la economía. La respuesta en el periodo en que el *shock* golpea se deja libre por la posibilidad de que se produzca un efecto "*price puzzle*" (que la inflación aumente al inicio).
  - La brecha del producto cae en los dos trimestres después de que el *shock* golpea a la economía, al igual que la inflación.

# Aplicación 2: *Shocks* monetarios en Chile; identificación por restricciones de signo (4/5)

- Respuesta a un *shock* de un desvío estándar a la TPM:



# Aplicación 2: *Shocks* monetarios en Chile; identificación por restricciones de signo (5/5)

- La estimación de los efectos de un *shock* a la TPM por restricciones de signo y mediante técnicas Bayesianas es razonablemente similar a la anterior.
- A diferencia de la identificación por estructura recursiva, en este caso la inflación y la brecha pueden responder libremente en el periodo en que el *shock* golpea a la economía (en la identificación por estructura recursiva, la inflación y la brecha no responden de manera contemporánea a la TPM por construcción).

# Apéndice A (1/2)

- $Y_t = X_t B + v_t$ puede escribirse como $y_t = (I_N \otimes X_t)b + V_t$, donde $y_t = vec(Y_t)$, $b = vec(B)$, $V_t = vec(v_t)$.
- Usemos el ejemplo de dos variables $(\pi_t,\ R_t)$ y dos rezagos.

$$y_t = (I_N \otimes X_t)b + V_t,$$

$$\begin{bmatrix} \pi_t \\ R_t \end{bmatrix} = \left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & \pi_{t-1} & R_{t-1} & \pi_{t-2} & R_{t-2} \end{bmatrix} \right) \begin{bmatrix} c_\pi \\ b_{11} \\ b_{12} \\ d_{11} \\ d_{12} \\ c_R \\ b_{21} \\ b_{22} \\ d_{21} \\ d_{22} \end{bmatrix} + \begin{bmatrix} v_{\pi t} \\ v_{Rt} \end{bmatrix},$$

# Apéndice A (2/2)

$$\begin{bmatrix}\pi_t \\ R_t\end{bmatrix} =$$

$$\begin{bmatrix}1 & \pi_{t-1} & R_{t-1} & \pi_{t-2} & R_{t-2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \pi_{t-1} & R_{t-1} & \pi_{t-2} & R_{t-2}\end{bmatrix} \begin{bmatrix} c_\pi \\ b_{11} \\ b_{12} \\ d_{11} \\ d_{12} \\ c_R \\ b_{21} \\ b_{22} \\ d_{21} \\ d_{22}\end{bmatrix} + \begin{bmatrix} v_{\pi t} \\ v_{Rt}\end{bmatrix}.$$

$$\begin{bmatrix}\pi_t \\ R_t\end{bmatrix} = \begin{bmatrix} c_\pi + b_{11}\pi_{t-1} + b_{12}R_{t-1} + d_{11}\pi_{t-2} + d_{12}R_{t-2} \\ c_R + b_{21}\pi_{t-1} + b_{22}R_{t-1} + d_{21}\pi_{t-2} + d_{22}R_{t-2}\end{bmatrix} + \begin{bmatrix} v_{\pi t} \\ v_{Rt}\end{bmatrix}.$$