

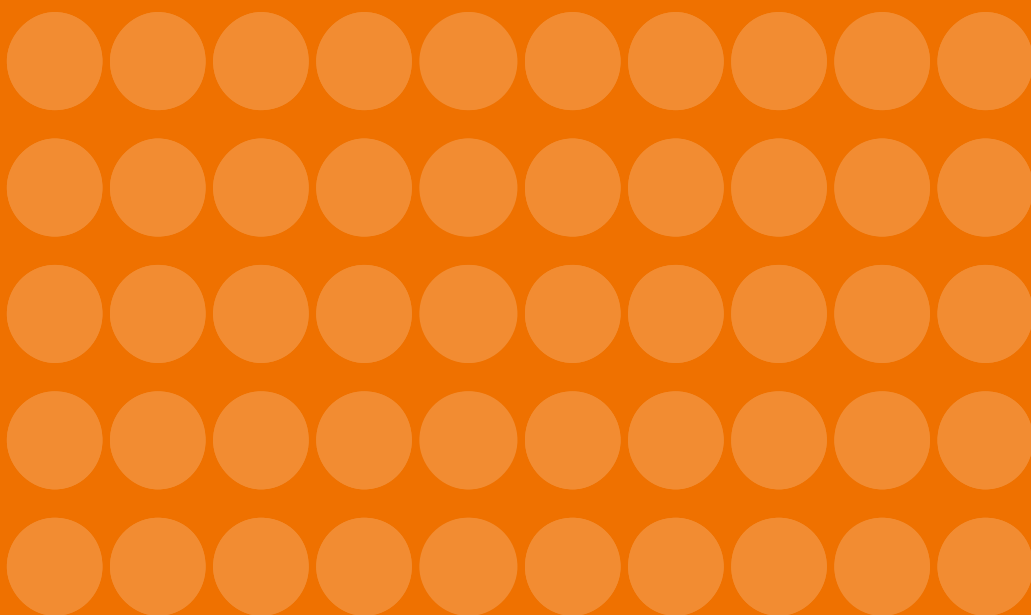
Sergio Alfonso Lago Alves  
Angelo Marsiglia Fasolo

**No es sólo un estudio más  
sobre frecuencia mixta**

Premio de Banca Central Rodrigo Gómez, 2016


**Not Just Another Mixed Frequency Paper**

Central Bank Award Rodrigo Gómez, 2016









No es sólo un  
estudio más sobre  
frecuencia mixta





Sergio Alfonso Lago Alves  
Angelo Marsiglia Fasolo

*No es sólo un estudio más  
sobre frecuencia mixta*

PREMIO DE BANCA CENTRAL RODRIGO GÓMEZ,  
2016

---

CENTRO DE ESTUDIOS MONETARIOS LATINOAMERICANOS

Ciudad de México, 2019

Primera edición, en español e inglés, 2019

---

D.R. © 2019 Centro de Estudios Monetarios Latinoamericanos  
Durango 54, colonia Roma Norte, delegación Cuahutémoc,  
06700 Ciudad de México, México.

ISNI 0000 0001 2315 4270

Derechos reservados conforme a la ley. *Queda prohibida la reproducción total o parcial de esta obra por cualquier medio o procedimiento comprendidos, la reprografía y el tratamiento informático, la fotocopia o la grabación, sin la previa autorización por escrito de la Dirección General del CEMLA.*

Coordinación editorial realizada por la Gerencia  
de Servicios de Información del CEMLA.

ISBN  
En trámite (electrónico)

Hecho en México  
*Made in Mexico*



# Índice





Presentación	11
Sobre los autores	12
Agradecimientos	13
<b>1. Introducción .....</b>	<b>14</b>
<b>2. El Modelo .....</b>	<b>23</b>
<b>3. Muestreo de Gibbs.....</b>	<b>27</b>
3.1 El modelo lineal dinámico .....	29
3.1.1 Algoritmo FPMR adaptado .....	34
3.2 VAR bayesiano .....	40
3.2.1 Cómo calcular la función de verosimilitud .....	42
3.2.2 Distribuciones conjuntas a priori y a posteriori.....	43
3.2.3 Cómo especificar la distribución conjunta a priori ...	44
<b>4. Propiedades del VAR de frecuencia mixta: simulaciones ..</b>	<b>47</b>
<b>5. Análisis empírico .....</b>	<b>54</b>
5.1 Cómo calcular el PIB real con frecuencia mensual.....	58

5.2 Cómo estimar la tasa de desempleo con datos irregularmente espaciados .....	61
<b>Conclusión .....</b>	<b>67</b>
<b>Anexos .....</b>	<b>71</b>
Anexo A. Algunos resultados.....	73
Álgebra matricial.....	74
Estimador de máxima verosimilitud-distribución gaussiana multivariado .....	75
Reescritura de la función de verosimilitud .....	75
La FDP conjunta a posteriori del VAR bayesiano .....	76
Anexo B. Pruebas .....	77
Proposición 1.....	77
Prueba.....	78
Proposición 2.....	80
Prueba.....	80
<b>Bibliografía.....</b>	<b>81</b>

## PRESENTACIÓN

Este estudio introduce un nuevo algoritmo, basado en un muestreo de Gibbs en dos pasos con un método de filtrado prospectivo con muestreo retrospectivo (FPMR), que tiene como propósito recuperar la distribución conjunta de observaciones faltantes en un conjunto de datos de frecuencia mixta. El nuevo algoritmo atenúa la mayoría de las restricciones que usualmente aparecen en estudios previos, a saber: 1) no requiere que se observe al menos una serie de tiempo cada periodo; 2) permite añadir fácilmente restricciones lineales basadas en la representación de estados de los espacios del VAR; 3) no requiere de series de tiempo regularmente espaciadas a bajas frecuencias; y, 4) evita los problemas de degeneración que surgen cuando los estados, o una combinación lineal de estados, son realmente observados. Asimismo, el algoritmo es idóneo para incorporar información de alta frecuencia en tiempo real que permita mejorar los pronósticos inmediatos y a futuro de series de tiempo de baja frecuencia. Evaluamos las propiedades del algoritmo utilizando datos simulados. Por otra parte, a manera de aplicaciones empíricas, simulamos el PIB mensual de Brasil, comparamos los resultados

con los del IBC-BR de Brasil, y recuperamos lo que serían las tasas de desempleo históricas de la PNAD-C antes de 2012.

Palabras clave: muestreo de Gibbs, filtrado prospectivo con muestreo retrospectivo (FPMR), datos de frecuencia mixta, métodos bayesianos, datos en tiempo real, pronóstico macroeconómico, autorregresiones vectoriales, PNAD.

Clasificación JEL: C11, C32, C53, E27.

## **SOBRE LOS AUTORES**

**S**ergio Afonso Lago Alvez actualmente ocupa el cargo de sub-jefe del Departamento de Investigación del Banco Central de Brasil. Posee una licenciatura en Ingeniería Mecánica por parte del Instituto Tecnológico de Aeronáutica (1996), una maestría en Transportación Aérea por parte del Instituto Tecnológico de Aeronáutica (2000), una maestría en Economía por parte de la Universidad de Brasilia (2005), así como un doctorado en Economía por parte de la Universidad de California Santa Cruz (2011). Su investigación se ha enfocado en los campos de la economía monetaria, la inflación tendencial, el desempleo, el límite inferior cero de las tasas de interés nominales, la inferencia bayesiana, los modelos DSGE y los métodos cuantitativos.

<sergio.lago@bcb.gov.br>

**A**ngelo Marsiglia Fasolo actualmente se desempeña como asesor sénior en el Departamento de Investigación del Banco Central de Brasil. Posee una maestría en Economía por parte de la Universidad Federal de Río Grande do Sul (2003) y un doctorado en Economía por parte de la Universidad de Duke (2010). Tiene experiencia en métodos cuantitativos y modelación macroeconómica, y ha publicado sobre la solución y estimación de modelos DSGE, la economía internacional, y la coordinación entre las políticas monetaria y fiscal.

<angelo.fasolo@bcb.gov.br>

## AGRADECIMIENTOS

Los autores agradecen, por un lado, a Hernán D. Seoane y Andrew Foerster por sus comentarios iniciales sobre las ideas principales de esta investigación, y, por el otro, a los participantes de los seminarios impartidos en el Banco Central do Brasil. Las opiniones aquí expresadas son responsabilidad de los autores y no representan las del Banco Central do Brasil.





1

# Introducción





Muchas series de tiempo económicas de países subdesarrollados y en desarrollo presentan problemas que impiden explorar a fondo su contenido informativo. Estos problemas comunes generalmente son resultado de: 1) su *corta duración*; 2) la medición sólo está disponible en *bajas frecuencias*, como trimestral o anual; 3) la *discontinuidad* debido a cambios de una metodología de medición a otra; 4) los problemas técnicos o de medición que producen *observaciones faltantes espaciadas aleatoriamente* en series de tiempo; y 5) el *cambio de frecuencia*, cuando la frecuencia de las observaciones se modifica a lo largo del periodo de muestra, por ejemplo, cuando la serie comienza siendo anual pero, después de un par de años, se convierte en trimestral.

La corta duración es un problema común que puede afectar también a los países desarrollados, si bien es más frecuente en los países en desarrollo. En pocas palabras, el problema surge cuando una serie de tiempo deja de medirse o apenas se empezó a medir recientemente. Es sencillo comprender las limitaciones de la inferencia generadas por las series de tiempo de baja frecuencia debido al reducido número de observaciones disponibles. El problema con la discontinuidad surge cuando se cambia el método de medición de una serie de tiempo sin actualizar el método para las observaciones pasadas. En este caso, las propiedades de los datos podrían verse significativamente afectadas debido al nuevo método de medición, de tal manera que las observaciones obtenidas

a partir del método anterior y del nuevo deben considerarse como dos series de tiempo diferentes.

Las observaciones faltantes espaciadas aleatoriamente pueden presentarse ya sea por fallas en la medición al recopilar y tratar datos brutos, o al excluir deliberadamente aquellas observaciones consideradas como valores atípicos o que no concuerdan con el efecto de las hipótesis que uno desea medir y demostrar. Por último, las series con cambios de frecuencia podrían considerarse como un caso especial de series con observaciones faltantes, excepto que el espacio entre las observaciones se haya establecido a partir de una elección discrecional. Por simplicidad, tratamos este caso de la misma forma en que tratamos las observaciones faltantes espaciadas aleatoriamente.<sup>1</sup> En el contexto del análisis multivariado, se produce una situación problemática cuando una combinación de los problemas mencionados anteriormente produce lo que se conoce como conjuntos de datos de *frecuencia mixta desigualmente espaciados*, en los que, en un caso extremo, se muestran periodos sin ninguna observación disponible; es decir, todos los elementos de la serie de tiempo podrían estar faltando en un momento dado.

En un marco de frecuencia mixta multivariado, las decisiones habituales pueden poner en riesgo la inferencia estadística, lo cual daría como resultado conclusiones erróneas. Por ejemplo, el analista puede llenar los vacíos en las series de tiempo con valores *ad hoc*, es decir, enfoques mecánicos sin sustento estadístico formal, tales como interpolaciones u observaciones recortadas en periodos en los que una o más series de tiempo presentan observaciones faltantes. De manera alternativa, se podría considerar reducir la frecuencia de las series de tiempo de todo el conjunto de datos para evitar atender los problemas de frecuencia mixta. Sin embargo, recortar los periodos y reducir las frecuencias puede provocar la pérdida de información relevante. De ahí que surja la necesidad de encontrar formas de analizar simultáneamente todo el contenido informativo disponible en el conjunto de datos de frecuencia mixta, evitando

---

<sup>1</sup> Ver Prado y West (2010, sección 4.3.3) para conocer más acerca de las series con observaciones faltantes con espaciado discrecional.

al mismo tiempo el uso de métodos *ad hoc* para llenar los vacíos de información.

En economía, la bibliografía ha evolucionado hacia el desarrollo de muestreos bayesianos de Gibbs para recuperar toda la distribución conjunta de las observaciones faltantes, en vez de identificar valores únicos para llenar los vacíos. En este estudio, contribuimos a la bibliografía proponiendo una forma eficiente e intuitiva de recuperar dicha distribución conjunta mediante un muestreo de Gibbs en dos pasos, desarrollado en su mayoría a partir de importantes contribuciones sobre el filtrado de frecuencia mixta hechas de manera reciente por Schorfheide y Song (2015) y por Eraker *et al.* (2015), y, en el campo de la estadística, a partir del eficiente algoritmo descrito en West (1996, 1997) y en West y Harrison (1997, Sección 15.3.2).

Cabe resaltar el hecho de que en la investigación estadística se ha considerado el problema de recuperar toda la distribución conjunta de observaciones faltantes mucho antes de que lo hiciera la investigación económica, con algoritmos eficientes que usan el muestreo de Gibbs, como el creado por West y que se describe más adelante, desarrollados durante los noventa. Por tanto, el punto central de este artículo es adaptar y llevar algunos de los primeros avances en el campo de la estadística a las aplicaciones en economía. Dicho esto, en nuestra opinión, el método propuesto por West (1996, 1997) es más eficiente, sencillo e intuitivo que aquellos propuestos por Schorfheide y Song (2015) y Eraker *et al.* (2015), aunque arroja más o menos los mismos resultados.

Al igual que Schorfheide y Song (2015) y Eraker *et al.* (2015), suponemos que un análisis de vectores autorregresivos (VAR) con parámetros desconocidos puede describir la dinámica de las series de tiempo multivariadas si todos sus valores faltantes fueron observados. La tarea del muestreo de Gibbs es recuperar la distribución conjunta de los valores faltantes y los parámetros desconocidos del VAR. Al igual que en West (1996, 1997) y en West y Harrison (1997, sección 15.3.2), nuestra estrategia para muestrear las series de tiempo es usar un enfoque de incremento de datos basado en Carter y Kohn (1994) y en Frühwirth-Schnatter (1994). Frühwirth-Schnatter (1994) llama a este enfoque algoritmo de *filtrado prospectivo con muestreo retrospectivo* (FPMR), utilizado también en Schorfheide y Song

(2015). Este es un paso del muestreo de Gibbs, el cual toma parámetros muestreados y una matriz de covarianzas de un paso anterior del muestreo de Gibbs, basado en un VAR bayesiano con un prior tipo Minnesota modificado (ver Doan *et al.*, 1984, y Litterman, 1986, para obtener más información sobre el prior de Minnesota original).

En su forma general, el algoritmo FPMR es fácil de llevar a cabo. Sin embargo, se presenta un problema en modelos con componentes autorregresivos (AR), o en cualquier otro modelo en el que los vectores de estado consecutivo contengan componentes comunes. En este caso, lo que en realidad se observa son combinaciones lineales de elementos del estado actual, y el algoritmo FPMR básico se degenera. En este punto, nos apartamos de Schorfheide y Song (2015) al proponer un algoritmo diseñado para realizar el FPMR para un caso VAR, evitando problemas de degeneración. El algoritmo se basa en el método de West para el muestreo eficiente de vectores de estado en modelos AR, adaptado a nuestro caso VAR. De igual forma, diferimos de Schorfheide y Song (2015) al permitir irregularidades en la frecuencia de las observaciones, generalizando la forma de la matriz de selección presentada por los autores para series de tiempo de frecuencia mixta regularmente espaciadas.

Eraker *et al.* (2015) también trabajan sobre series de tiempo de frecuencia mixta desigualmente espaciadas, pero no aprovechan el uso de una especificación de espacio de estados, la cual permite restricciones lineales adicionales en la relación entre las series de tiempo observadas y las variables latentes. Normalmente, manejar valores faltantes en un marco de espacios de estados, como el que usamos en nuestro paso de Gibbs para el muestreo de observaciones faltantes condicional sobre el conjunto de parámetros, es relativamente fácil (véase, por ejemplo, Prado y West, 2010, sección 4.3.3, y Durbin y Koopman, 2012, sección 4.10). El enfoque es lo suficientemente flexible como para no necesitar que haya por lo menos una serie de tiempo que sea observada en cada periodo. De hecho, funciona incluso cuando no hay observaciones en periodos específicos. Al incorporar este enfoque, también nos apartamos de Schorfheide y Song (2012) y de Eraker *et al.* (2015), pues estos requieren de al menos una serie de tiempo completamente observada. Por lo tanto, nuestro enfoque aprovecha las mejores

características de los procedimientos tanto de Schorfheide y Song (2015) como de Eraker *et al.* (2015).

La flexibilidad que brinda el hecho de no necesitar que haya al menos una serie de tiempo observada cada periodo resulta de suma utilidad al realizar pronósticos (*forecasts*) y pronósticos retrospectivos (*backcasts*), ya que el algoritmo sólo tiene que tomar en cuenta estos ejercicios como una muestra ampliada para la que no hay ninguna observación. En otras palabras, el algoritmo FPMR muestrea los valores faltantes en ejercicios de pronóstico y pronóstico retrospectivo. Por lo tanto, nuestro enfoque es eficiente, intuitivo e idóneo para resolver los cuatro problemas comunes antes descritos: la corta duración, las bajas frecuencias, la discontinuidad, las observaciones faltantes aleatoriamente espaciadas y las series de tiempo multivariadas con periodos en los cuales no se cuenta con observaciones disponibles.

El algoritmo también es ideal para manejar información de alta frecuencia en tiempo real con el propósito de mejorar los pronósticos inmediatos (*nowcasts*) y futuros de series de tiempo de frecuencia más baja. Un ejercicio empírico en la sección 5 muestra la mejora en los pronósticos inmediatos del PIB trimestral de Brasil a medida que llega información de alta frecuencia en tiempo real y compara los resultados con aquellos del IBC-BR de Brasil. El ejercicio también resalta el papel que desempeña una configuración adecuada de la representación del espacio de estados del modelo, a fin de establecer la relación entre las observaciones mensuales y trimestrales. En otro ejercicio, usamos el algoritmo para recuperar la distribución conjunta de las tasas de desempleo históricas de la PNAD-C de Brasil anteriores a 2012, utilizando información de otras encuestas como la PNAD anual y la PME. La importancia de este último ejercicio radica en el hecho de que la PNAD-C es una nueva serie de tiempo mensual, cuyos valores sólo han estado disponibles a partir de 2012.

El resto del artículo está estructurado de la siguiente manera: el modelo se describe en la sección 2; el muestreo de Gibbs y la información sobre cómo recuperar toda la distribución conjunta de las observaciones faltantes se describe en la sección 3. La información sobre el algoritmo FPMR eficiente y el VAR bayesiano se describe en las secciones 3.1 y 3.2. El desempeño de nuestro

enfoque con conjuntos de datos simulados se prueba en la sección 4. En la sección 5, se muestran dos ejercicios empíricos basados en la información del PIB y el desempleo en Brasil. Los ejercicios empíricos exploran tanto la cuestión de la regularidad de la información faltante como el uso de restricciones lineales en el modelo. Por último, en la sección 6, se hace un resumen de las conclusiones del estudio.



# 2

# El Modelo





**A**sumimos la existencia de una muestra  $\mathbf{Y}^{ob} \equiv \{y_1^{ob}, \dots, y_T^{ob}\}$  de tamaño  $T$ . En cada periodo,  $y_t^{ob}$  es un vector ( $m_t^{ob} \times 1$ ) de variables endógenas observadas, cuya dimensión  $m_t^{ob}$  cambia con el paso del tiempo debido a la frecuencia mixta irregularmente espaciada. Esta definición de  $m_t^{ob}$  no excluye la posibilidad de que, en ciertos periodos, no haya observaciones disponibles, es decir,  $m_t^{ob} = 0$ .

En este contexto, definimos primero  $y_t$  como un vector ( $m \times 1$ ) de variables endógenas de interés, cuya dimensión  $m \geq m_t^{ob}$  ( $m > m_t^{ob}$ , si  $m_t^{ob} = 0$ ) es invariante en el tiempo. Nótese que es posible que el conjunto de variables endógenas de interés pudiera no ser directamente observado en lo absoluto en la muestra. Por ejemplo,  $y_t^{ob}$  podría incluir promedios móviles de  $y_t$ , los cuales son transformaciones afines de  $y_t$  y sus valores rezagados. En este caso,  $y_t^{ob}$  ni siquiera es un subconjunto de  $y_t$ . La relación entre  $y_t^{ob}$  y  $y_t$  se aborda formalmente en la sección 3.1.

Nuestro objetivo es entonces inferir la distribución conjunta de la secuencia  $\mathbf{Y} \equiv [y_1, \dots, y_T]'$  ( $T \times m$ ) de variables endógenas de interés. Para realizar esta tarea, se asume que la dinámica de  $y_t$  puede representarse por un VAR( $q$ ):

$$\mathbf{1} \quad y_t = \Phi_1 y_{t-1} + \dots + \Phi_q y_{t-q} + \bar{\Phi}_c \bar{x}_{ct} + e_t, \quad e_t \stackrel{iid}{\sim} N(0, \Sigma_e),$$

donde  $\bar{x}_{ct}$  es una matriz de variables deterministas ( $m_c \times 1$ ), tales como la constante, la tendencia y las variables ficticias estacionales,

$\Phi_\ell$  es una matriz de coeficientes ( $m \times m$ ) para  $\ell \in \{1, \dots, q\}$ ,  $\bar{\Phi}_c$  es una matriz de coeficientes ( $m \times m_c$ ), y  $\Sigma_e$  es una matriz de covarianzas definida positiva ( $m \times m$ ). Cada ecuación tiene  $k \equiv mq + m_c$  regresores, el VAR tiene coeficientes  $mk$ , y todo el sistema tiene parámetros  $mk + m^2$ .

$$y_t = [y_{1t} \dots y_{mt}]'$$

$$e_t = [e_{1t} \dots e_{mt}]'$$

$$\bar{x}_{ct} = [\bar{x}_{c;1t} \dots \bar{x}_{c;m_c t}]'$$

$$\Phi_\ell = \begin{pmatrix} \Phi_{\ell;11} & \dots & \Phi_{\ell;m1} \\ \vdots & & \vdots \\ \Phi_{\ell;1m} & \dots & \Phi_{\ell;mm} \end{pmatrix}$$

$$\bar{\Phi}_c = \begin{pmatrix} \bar{\Phi}_{c;11} & \dots & \bar{\Phi}_{c;m1} \\ \vdots & & \vdots \\ \bar{\Phi}_{c;1m_c} & \dots & \bar{\Phi}_{c;mm_c} \end{pmatrix}$$

Por conveniencia analítica, cada parte del muestreo de Gibbs, descrito en la sección 3, considera diferentes formas compañeras al tratar con el sistema lineal. Se brinda más información de ello en las secciones 3.1 y 3.2.



# 3

## Muestreo de Gibbs



Esta sección describe los pasos de un algoritmo de Gibbs para extraer una muestra de tamaño  $S$ , dado el conjunto de información descrito por  $\mathbf{Y}^{ob}$ , de la distribución conjunta de  $(\mathbf{Y}, \Phi | \mathbf{Y}^{ob})$ . Cada paso del muestreo consiste en dos partes; 1) extractos de muestra de  $\mathbf{Y}$  de la distribución condicional de  $(\mathbf{Y} | \mathbf{Y}^{ob}, \Phi)$ , mediante el filtro de Kalman y una forma modificada del algoritmo de *filtrado prospectivo con muestreo retrospectivo* (FPMR), tal como se describe en la sección 3.1; y 2) extractos de muestra de  $\Phi$  de la distribución condicional de  $(\Phi | \mathbf{Y}^{ob}, \mathbf{Y})$ , mediante un VAR bayesiano, tal como se describe en la sección 3.2

El muestreo se inicia, en la parte 1) del paso 0, estableciendo  $\Phi_{mit}$  para implicar una dinámica de paseo aleatorio a  $\mathbf{Y}$ , es decir,  $\Phi_\ell = \frac{1}{q} \mathbf{I}_{(m \times m)}$ , para  $\ell \in \{1, \dots, q\}$ , y  $\bar{\Phi}_c = 0$ . El algoritmo FPMR de la parte 1) suaviza la trayectoria de  $\mathbf{Y}_0$  de  $(\mathbf{Y} | \mathbf{Y}^{ob}, \Phi_{mit})$  para usarla en la parte 2), donde un VAR bayesiano genera el conjunto de parámetros iniciales  $\Phi_0$  de  $(\Phi | \mathbf{Y}^{ob}, \mathbf{Y}_0)$ , aún en el paso 0. Del paso  $s=1$  a  $s=S$ , el algoritmo de Gibbs evoluciona de la manera habitual; 1) extrae  $\mathbf{Y}_s$  de  $(\mathbf{Y} | \mathbf{Y}^{ob}, \Phi_{s-1})$  y 2) extrae  $\Phi_s$  de  $(\Phi | \mathbf{Y}^{ob}, \mathbf{Y}_s)$ .

### 3.1 El modelo lineal dinámico

A continuación, caracterizamos el sistema  $VAR_m(q)$  1 como un modelo lineal dinámico (MLD), al igual que en Schorfheide y Song (2015),

y describe la actualización secuencial, es decir, ecuaciones de filtro de Kalman (1960).<sup>2</sup> Una forma compañera del sistema  $VAR_m(q)$  es:

$$\begin{aligned}
 \text{Ecuaci } n \\ \text{de observaci } n \\
 y_t^{ob} &= F_t z_t; \quad F_t \equiv M_t \Lambda_z \\
 \text{Ecuaci } n \\ \text{de transici } n \\
 z_t &= G_1 z_{t-1} + \omega_t; \quad \omega_t \sim N(G_2 \bar{x}_{ct}, \Omega_e)
 \end{aligned}$$

donde  $y_t^{ob}$  es un vector ( $m_t^{ob} \times 1$ ) de variables endógenas observadas, cuya dimensión  $m_t^{ob}$  cambia debido a observaciones de frecuencia mixta o irregularmente espaciadas,  $M_t$  y  $\Lambda_z$  son generalizaciones de lo que se hizo en Schorfheide y Song (2015), es decir,  $M_t$  es una *matriz de selección* ( $m_t^{ob} \times m$ ) y  $\Lambda_z$  es una *matriz de transformación* ( $m \times m\bar{q}$ ),<sup>3</sup> útil cuando la variable observada es una transformación afín conocida de los estados (por ejemplo, promedios móviles de tres meses o cualquier otra combinación lineal),  $z_t \equiv [y'_t, \dots, y'_{t-\bar{q}+1}]'$  es un vector de estados ( $m\bar{q} \times 1$ ),  $G_1$  es una matriz de coeficientes ( $m\bar{q} \times m\bar{q}$ ) para variables endógenas,  $G_2$  es una matriz de coeficientes para variables endógenas ( $m\bar{q} \times m_c$ ), y  $\Omega_e$  es una matriz de covarianzas semidefinida positiva ( $m\bar{q} \times m\bar{q}$ ):

<sup>2</sup> Excelentes referencias sobre modelos lineales dinámicos (MLD) e interferencia mediante el filtro de Kalman son Hamilton (1994), Prado y West (2010) y West y Harrison (1997). Para obtener más información sobre las aplicaciones MLD en modelos macroeconómicos, ver Basdevant (2003).

<sup>3</sup> En este sentido,  $\Lambda_z$  se debe diseñar con cuidado para evitar *problemas de solapamiento* cuando sólo observamos promedios, o cualquier otra combinación lineal, de variables latentes de interés. Por ejemplo, supóngase que observamos tres series de tiempo trimestrales  $y_t = [y_{1t}, y_{2t}, y_{3t}]'$ , que en realidad agregan o promedian las variables mensuales latentes durante cada trimestre, como el PIB sectorial. En este caso, hay un conjunto infinito de posibles series de tiempo mensuales del PIB sectorial que son congruentes con las variables observadas. Esos conjuntos incluyen series de tiempo mensuales con variaciones extremada e irracionalmente marcadas. En este caso, el algoritmo no convergerá hacia una distribución estacionaria única.

$$G_1 = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{\bar{q}-1} & \Phi_{\bar{q}} \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \quad G_2 = \begin{bmatrix} \bar{\Phi}_c \\ 0_{m \times m_c} \\ 0_{m \times m_c} \\ \vdots \\ 0_{m \times m_c} \end{bmatrix} \quad \Omega_e = \begin{bmatrix} \Sigma_e & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Nótese que  $\bar{q}$  no necesariamente coincide con  $q$ , el número de rezagos en el  $VAR_m(q)$  que describe la dinámica de  $y_t$ . Por ejemplo, supóngase que  $y_t^{ob}$  representa los promedios trimestrales ( $\bar{q} = 3$ ) de valores mensuales no observados de  $y_t$ , cuya dinámica se describe mediante un  $VAR$  con cuatro rezagos ( $q = 4$ ). Por otra parte, si la dinámica se describe mejor con un  $VAR_m(1)$ , surge un problema similar. Tales casos se abordan con facilidad de la siguiente manera: 1) si  $\bar{q} < q$ , defina  $\bar{q} = q$  y amplíe  $\Lambda_z$  con matrices cero, es decir,  $\Lambda_z = [\Lambda_z, 0_{m \times m(q-\bar{q})}]$ ; 2) si  $\bar{q} > q$ , considere matrices cero adicionales  $\Phi_{\bar{\ell}} = 0_{m \times m}$  para  $\bar{\ell} \in \{q+1, \dots, \bar{q}\}$ .

En este contexto,  $M_t$  es una matriz de selección variable en el tiempo en la que el número de filas se ajusta para coincidir con el número de observaciones ausentes en cada periodo. Desde esta perspectiva, partiendo de una matriz de identidad  $M$  de tamaño  $(m \times m)$ ,  $M_t$  se define en cada periodo eliminando las filas  $m - m_t^{ob}$  de  $M$ , las cuales corresponden a las variables no observadas del vector  $y_t^{ob}$ .

Sea  $\mathcal{D}_t \equiv \{\mathcal{D}_{t-1}, y_t^{ob}\}$  que denota el conjunto de información en cada periodo  $t$ . Si no hay valores observados en el periodo  $t$ , es decir,  $y_t^{ob}$  es un vector vacío, entonces  $M_t$  es una matriz vacía y  $\mathcal{D}_t = \mathcal{D}_{t-1}$ .<sup>4</sup>

Considérese que la siguiente densidad de probabilidad gaussiana prior inicial para  $z_0$  es  $(z_0 | \mathcal{D}_0) \sim N(\mu_0, C_0)$ . Por lo tanto, las ecuaciones del filtro de Kalman (FK) son:

<sup>4</sup> Se puede hallar un análisis más profundo sobre modelos con datos faltantes y desigualmente espaciados en, por ejemplo, Prado y West (2010, sección 4.3.3) y Durbin y Koopman (2012, sección 4.10).



a) Densidad a priori para  $z_t$ , con información hasta el periodo  $(t-1)$ :

$$(z_t | \mathcal{D}_{t-1}) \sim N(a_t, R_t)$$

$$a_t \equiv G_1 \mu_{t-1} + G_2 \bar{x}_{ct}; \quad R_t \equiv G_1 C_{t-1} G_1' + \Omega_e$$

b) Pronóstico en un paso de  $y_t^{ob}$  con información hasta el periodo  $(t-1)$ :

$$(y_t^{ob} | \mathcal{D}_{t-1}) \sim N(f_t, Q_t)$$

$$f_t \equiv F_t a_t; \quad Q_t \equiv F_t R_t F_t'$$

c) Densidad a posteriori para  $z_t$  con información hasta el periodo  $t$ :

$$(z_t | \mathcal{D}_t) \sim N(\mu_t, C_t).$$

Si hay observaciones en el periodo  $t$ :

$$\mu_t \equiv a_t + A_t v_t; \quad C_t \equiv R_t - A_t Q_t A_t'; \quad v_t \equiv y_t^{ob} - f_t; \quad A_t \equiv R_t F_t' Q_t^{-1}.$$

De lo contrario:

$$\mathcal{D}_t = \mathcal{D}_{t-1}; \quad (z_t | \mathcal{D}_t) \equiv (z_t | \mathcal{D}_{t-1}); \quad \mu_t \equiv a_t; \quad C_t \equiv R_t.$$

Nuestra estrategia para completar la serie con observaciones de frecuencia mixta o irregularmente espaciadas es usar un enfoque de incremento de datos basado en Carter y Kohn (1994) y en FrühwirthSchnatter (1994) para muestrear secuencias completas de variables de estado. De igual modo, seguimos a FrühwirthSchnatter (1994) al denominar este enfoque un algoritmo de FPMR. Sea  $Z_T \equiv [z_T, z_{T-1}, \dots, z_1]$  que denota la secuencia completa

de vectores de estado. La naturaleza del algoritmo FPMR proviene del análisis de la estructura de Markov sobre la ecuación de evolución de cualquier MLD, la cual nos permite escribir

$$\Pr(Z_T | \mathcal{D}_T) = \Pr(z_1 | z_0, \mathcal{D}_1) \dots \Pr(z_t | z_{t+1}, \mathcal{D}_t) \dots \Pr(z_{T-1} | z_T, \mathcal{D}_{T-1}) \Pr(z_T | \mathcal{D}_T).$$

La comprobación se muestra en el anexo.

Por lo tanto, el algoritmo FPMR consiste en muestrear  $z_T$  a partir de  $\Pr(z_T | \mathcal{D}_T)$ , y muestrear secuencialmente  $z_t$  a partir de  $\Pr(z_t | z_{t+1}, \mathcal{D}_t)$  mediante ecuaciones directamente derivadas del filtro de Kalman.

En un artículo reciente, similar al nuestro, Schorfheide y Song (2015) también utilizan el enfoque de incremento de datos basado en Carter y Kohn (1994). Nos apartamos de Schorfheide y Song (2015) al permitir irregularidades en la frecuencia de las observaciones, en vez de sencillas series de tiempo regularmente espaciadas de frecuencia mixta. Esta ampliación, realizada también en Eraker *et al.* (2015), se obtiene al generalizar la forma de la matriz  $M_t$ . Eraker *et al.* (2015), por otro lado, no aprovechan el uso de una especificación de espacio de estados, la cual permite restricciones lineales adicionales en la relación entre observables y estados, representada por la matriz  $\Lambda_z$ . Por lo tanto, nuestro enfoque aprovecha las mejores características de ambos procedimientos.

Diferimos, asimismo, de Schorfheide y Song (2015) y de Eraker *et al.* (2015), ya que no necesitamos que exista por lo menos una serie cuyos valores sean observados en todos los periodos. En este sentido, nuestro enfoque es idóneo para los siguientes casos: 1) conjuntos de datos de países en los que las series sean relativamente cortas; 2) series nuevas o series cuyo método de medición cambie sin aplicarlo a periodos previos.

En su forma general, el algoritmo FPMR es bastante fácil de implementar y requiere de una matriz inversa  $R_t$ , es decir, la matriz de covarianzas de  $(z_t | \mathcal{D}_{t-1})$ .<sup>5</sup> Sin embargo, se presenta un problema

---

<sup>5</sup> En efecto, por cada  $t \in \{(T-1), (T-2), \dots, 1\}$  y después de haber muestreado  $z_{t+1}$ , obtendríamos la distribución condicional  $(z_t | z_{t+1}, \mathcal{D}_t) \sim N(h_t, H_t)$ , donde  $h_t = \mu_t + B_t(z_{t+1} - a_{t+1})$ ,  $H_t = C_t - B_t R_{t+1} B_t'$  y  $B_t \equiv C_t G_1' R_{t+1}^{-1}$ .

en modelos con componentes AR, o en cualquier otro modelo en el que los vectores de estados consecutivos contengan componentes comunes. En este caso, lo que realmente se observa son las combinaciones lineales de elementos del estado actual. Este hecho implica que habrá muchas instancias en las que  $R_t$  sea singular, por lo que el algoritmo básico FPMR se degenerará.

Por lo tanto, también diferimos de dichos autores al proponer un algoritmo diseñado para realizar un FPMR para un caso VAR, evitando problemas de degeneración. Este se basa en West (1996, 1997) y en West y Harrison (1997, sección 15.3.2) sobre el muestreo eficiente de vectores de estados en modelos AR. Adaptamos el enfoque de West a nuestro caso VAR, el cual se detalla más adelante.

### 3.1.1 Algoritmo FPMR adaptado

La idea detrás del enfoque de West tiene su origen en la estructura representada en la ecuación 3. Nótese que, dado un vector previamente muestreado  $z_{t+1}$ , el procedimiento de muestreo retrospectivo implica que sólo hay un elemento del vector  $z_t$  que aún tiene que muestrearse: el vector  $(m \times 1)$ ,  $y_{t-\bar{q}+1}$ . Todo lo que debemos hacer es muestrear valores de la distribución condicional de  $(y_{t-\bar{q}+1} | z_{t+1}, \mathcal{D}_t)$  y completar el vector  $z_t$  mediante los elementos correspondientes del vector  $z_{t+1}$ . El caso en que  $\bar{q} = 1$ , el cual no encaja en la estructura representada, puede resolverse con facilidad imponiendo  $\bar{q} = 2$  y definiendo  $\Phi_2 = 0_{(m \times m)}$ .

$$\begin{matrix} \mathbf{3} & z'_{t+1} = & [y'_{t+1}, & y'_t, & y'_{t-1}, & \dots, & y'_{t-\bar{q}+2}] \\ & & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \\ & z'_t = & [y'_t, & y'_{t-1}, & \dots, & y'_{t-\bar{q}+2}, & y'_{t-\bar{q}+1}] \end{matrix} .$$

Adaptar la modificación de West del algoritmo FPMR para muestrear  $z_t$  es sencillo en nuestro caso. Primero, sea  $\tilde{z}_t \equiv [y'_t, \dots, y'_{t-\bar{q}+2}]'$  que denota el vector  $(m(\bar{q}-1) \times 1)$  ya muestreado de los elementos comunes de  $z_{t+1}$  y  $z_t$ , es decir, separamos ambos vectores como

$z_{t+1} = [y'_{t+1}, \tilde{z}'_t]'$  y  $z_t = [\tilde{z}'_t, y'_{t-\bar{q}+1}]'$ . En este contexto, sea  $\tilde{z}_{j,t}$  que denote cada elemento singular de  $\tilde{z}_t$  para  $j \in \{1, \dots, m(\bar{q}-1)\}$ .

Por simplicidad analítica, considérese el subvector  $\delta_{j,t} \equiv [\tilde{z}'_{j:m(\bar{q}-1),t}, y'_{t-\bar{q}+1}]'$  donde  $\tilde{z}_{j_1:j_2,t} \equiv [\tilde{z}_{j_1,t}, \dots, \tilde{z}_{j_2,t}]'$  para  $j_2 \geq j_1$  y  $\tilde{z}_{j_1:j_2,t} \equiv \text{vacío}$  para  $j_2 < j_1$ ,<sup>6</sup> así como las siguientes particiones para  $j \leq i$ :

$$\delta_{j,t} = \begin{bmatrix} \tilde{z}_{j,t} \\ \delta_{j+1,t} \end{bmatrix}, \quad \mu_{j,j,t} = \begin{bmatrix} \tilde{\mu}_{j,t} \\ \mu_{j,j+1,t} \end{bmatrix}, \quad C_{j,j,t} = \begin{bmatrix} \tilde{\sigma}_{j,t}^2 & \Sigma'_{j,j+1,t} \\ \Sigma_{j,j+1,t} & C_{j,j+1,t} \end{bmatrix},$$

donde  $\mu_{j,i,t} \equiv E(\delta_{i,t} | \tilde{z}_{1:(j-1),t}, \mathcal{D}_t)$ ,  $C_{j,i,t} \equiv \text{Var}(\delta_{i,t} | \tilde{z}_{1:(j-1),t}, \mathcal{D}_t)$ ,  
 $\tilde{\mu}_{j,t} \equiv E(\tilde{z}_{j,t} | \tilde{z}_{1:(j-1),t}, \mathcal{D}_t)$ ,  $\tilde{\sigma}_{j,t}^2 \equiv \text{Var}(\tilde{z}_{j,t} | \tilde{z}_{1:(j-1),t}, \mathcal{D}_t)$   
y  $\Sigma_{j,i,t} \equiv \text{Cov}(\delta_{i,t}, \tilde{z}_{j,t} | \tilde{z}_{1:(j-1),t}, \mathcal{D}_t)$ . Las particiones implican que  $\tilde{z}_{j,t}$ ,  
 $\tilde{\mu}_{j,t}$  y  $\tilde{\sigma}_{j,t}^2$  son escalares.

Los siguientes pasos describen la estrategia de muestreo:

- 1) Muestree un valor de  $z_T$  a partir de  $(z_T | \mathcal{D}_T) \sim N(\mu_T, C_T)$ ;
- 2) Por cada  $t \in \{(T-1), (T-2), \dots, 1\}$ , muestree secuencialmente los valores de la distribución condicional de  $(y_{t-\bar{q}+1} | z_{t+1}, \mathcal{D}_t)$  y complete el vector  $z_t$  mediante los elementos relevantes del vector  $z_{t+1}$ . Esto se logra al obtener las distribuciones de  $(y_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t)$  y  $(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$ . Nótese también, como se muestra en el anexo, que  $\Pr(y_{t-\bar{q}+1} | z_{t+1}, \mathcal{D}_t) \propto \Pr(y_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t) \Pr(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$ .

<sup>6</sup> Nótese que  $\delta_{1,t} = z_t$  y  $\delta_{(m(\bar{q}-1)+1),t} = y_{t-\bar{q}+1}$ .

Dado que  $z_t = [\tilde{z}'_t, y'_{t-\bar{q}+1}]'$ , la primera opción para recuperar  $\Pr(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$  sería utilizar la distribución del FK  $(z_t | \mathcal{D}_t) \sim N(\mu_t, C_t)$  para obtener directamente la distribución condicional de  $(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$ . Sin embargo, este método implica invertir  $\text{Var}(\tilde{z}_t | \mathcal{D}_t)$ , el cual suele ser singular.<sup>7</sup> Por lo tanto, con base en el método de West, obtenemos eficientemente la distribución condicional  $(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$  al condicionar de manera secuencial cada elemento escalar  $\tilde{z}_{j,t}$  de  $\tilde{z}_t$ , reduciendo la dimensión de la distribución por uno en cada etapa. Los pasos se describen a continuación:

- a) Dado que  $\hat{\mathfrak{z}}_{1,t} = z_t$ , nótese que  $\mu_{1,1,t} = \mu_t$  y  $C_{1,1,t} = C_t$ .
- b) Para  $j \in \{1, \dots, m(\bar{q}-1)\}$ , calcule la distribución de  $(\hat{\mathfrak{z}}_{(j+1),t} | \tilde{z}_{1:j,t}, \mathcal{D}_t)$ :<sup>8</sup>

$$\left( \hat{\mathfrak{z}}_{(j+1),t} | \tilde{z}_{1:j,t}, \mathcal{D}_t \right) \sim N\left( \mu_{(j+1),(j+1),t}, C_{(j+1),(j+1),t} \right)$$

$$\mu_{(j+1),(j+1),t} = \begin{cases} \mu_{j,j+1,t} + \frac{1}{\tilde{\sigma}_{j,t}^2} \Sigma_{j,j+1,t} (\tilde{z}_{j,t} - \tilde{\mu}_{j,t}), & \text{si } \tilde{\sigma}_{j,t}^2 \neq 0 \\ \mu_{j,j+1,t}, & \text{si } \tilde{\sigma}_{j,t}^2 = 0 \end{cases},$$

<sup>7</sup> Las mismas razones por las que  $R_t$  eventualmente es singular podrían causar que  $\text{Var}(\tilde{z}_t | \mathcal{D}_t)$  también lo sea.

<sup>8</sup> Si  $\tilde{\sigma}_{j,t}^2 = 0$ , entonces  $\Sigma_{j,j+1,t} = 0$  y  $\tilde{z}_{j,t} = \tilde{\mu}_{j,t}$ . En este caso, el valor límite de  $\frac{1}{\tilde{\sigma}_{j,t}^2} \Sigma_{j,j+1,t} (\tilde{z}_{j,t} - \tilde{\mu}_{j,t})$  y  $\frac{1}{\tilde{\sigma}_{j,t}^2} \Sigma_{j,j+1,t} \Sigma'_{j,j+1,t}$  es cero en ambos casos. Para fines de codificación, aún podría haber alguna inestabilidad numérica en casos en los que  $\tilde{\sigma}_{j,t}^2$  es positivo, si bien muy poca. En este caso, la restricción  $\tilde{\sigma}_{j,t}^2 > 0$  podría convertirse en algo como, por ejemplo,  $\tilde{\sigma}_{j,t}^2 > \text{Toler} \cdot \tilde{\mu}_{j,t}$ , donde *Toler* es un parámetro de tolerancia relativa.

$$C_{(j+1),(j+1),t} = \begin{cases} C_{j,j+1,t} - \frac{1}{\tilde{\sigma}_{j,t}^2} \Sigma_{j,j+1,t} \Sigma'_{j,j+1,t}, & \text{si } \tilde{\sigma}_{j,t}^2 \neq 0 \\ C_{j,j+1,t} & \text{si } \tilde{\sigma}_{j,t}^2 = 0 \end{cases}.$$

c) Dado que  $\delta_{(m(\bar{q}-1)+1),t} = y_{t-\bar{q}+1}, (y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$  se distribuye como  $(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t) \sim N(\tilde{\mu}_{t-\bar{q}+1}, \tilde{C}_{t-\bar{q}+1})$ , donde

$$\tilde{\mu}_{t-\bar{q}+1} = \mu_{(m(\bar{q}-1)+1), (m(\bar{q}-1)+1), t},$$

$$\tilde{C}_{t-\bar{q}+1} = C_{(m(\bar{q}-1)+1), (m(\bar{q}-1)+1), t}.$$

3) Basándose en el vector ya muestreado  $\tilde{z}_{t+1}$  y el vector exógeno  $\bar{x}_{ct+1}$ , calcule el residuo parcial

$$\tilde{e}_{t+1} \equiv y_{t+1} - (\Phi_1 y_t + \dots + \Phi_{\bar{q}-1} y_{t-\bar{q}+2} + \bar{\Phi}_c \bar{x}_{ct+1})$$

Nótese que  $\Pr(\tilde{e}_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t) = \Pr(y_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t)$ , cuya distribución condicional es  $(y_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t) \sim N(\Phi_{\bar{q}} y_{t-\bar{q}+1}, \Sigma_e)$ .

4) Considérese

que  $\Pr(y_{t-\bar{q}+1} | z_{t+1}, \mathcal{D}_t) \propto \Pr(y_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t) \Pr(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$ , tal como se muestra en el anexo. Si  $\tilde{C}_{t-\bar{q}+1}$  no es singular, todo lo que debemos hacer es muestrear  $y_{t-q+1}$  a partir de

$$4 \quad (y_{t-\bar{q}+1} | z_{t+1}, \mathcal{D}_t) \sim N(\mu_{t-q+1}^y, C_{t-q+1}^y)$$

donde

$$\mu_{t-q+1}^y \equiv \left( \Phi_{\bar{q}}' \Sigma_e^{-1} \Phi_{\bar{q}} + (\tilde{C}_{t-\bar{q}+1})^{-1} \right)^{-1} \left( \Phi_{\bar{q}}' \Sigma_e^{-1} \tilde{e}_{t+1} + (\tilde{C}_{t-\bar{q}+1})^{-1} \tilde{\mu}_{t-\bar{q}+1} \right),$$

$$C_{t-q+1}^y \equiv \left( \Phi_{\bar{q}}' \Sigma_e^{-1} \Phi_{\bar{q}} + (\tilde{C}_{t-\bar{q}+1})^{-1} \right)^{-1}.$$

Hasta este punto, hemos observado fielmente las recomendaciones de West. Sin embargo, podría haber un problema al muestrear  $(y_{t-\bar{q}+1} | z_{t+1}, \mathcal{D}_t)$  basándonos en el último resultado, es decir, podría haber casos en los que  $\tilde{C}_{t-\bar{q}+1}$  sea único. En este caso, proponemos un procedimiento similar al que se realizó en 2.b. Nuestra recomendación es muestrear de manera secuencial cada elemento escalar  $y_{j,t-\bar{q}+1}$  de  $y_{t-\bar{q}+1}$ , para  $j \in \{1, \dots, m\}$ , con lo cual se reduce la dimensión de la distribución  $(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$  por uno en cada etapa. Para tal efecto, considérense los subvectores  $\check{y}_{j,t} \equiv y_{j:m,t-\bar{q}+1}$  y  $\hat{y}_{j,t} \equiv y_{1:j,t-\bar{q}+1}$ , la distribución condicional  $\Pr(\check{y}_{j,t} | y_{(j-1)}, z_{t+1}, \mathcal{D}_t) \propto \Pr(\tilde{e}_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t) \Pr(\check{y}_{j,t} | y_{(j-1)}, \tilde{z}_t, \mathcal{D}_t)$ , donde  $\tilde{y}_{j,t} \equiv y_{j,t-\bar{q}+1}$  es una notación simplificadora, así como las siguientes divisiones:

$$\check{y}_{j,t} = \left[ \frac{y_{j,t-\bar{q}+1}}{\check{y}_{(j+1),t}} \right], \quad \check{\mu}_{j,j,t} = \left[ \frac{\mu_{j,t-\bar{q}+1}}{\check{\mu}_{j,j+1,t}} \right], \quad \check{C}_{j,j,t} = \left[ \frac{\check{\sigma}_{j,t}^2}{\check{\Sigma}_{j,j+1,t}} \middle| \frac{\check{\Sigma}'_{j,j+1,t}}{\check{C}_{j,j+1,t}} \right],$$

donde las medias son  $\check{\mu}_{j,i,t} \equiv E(\check{y}_{i,t} | y_{(j-1)}, \tilde{z}_t, \mathcal{D}_t)$  y  $\mu_{j,t-\bar{q}+1} \equiv E(\tilde{y}_{j,t} | y_{(j-1)}, \tilde{z}_t, \mathcal{D}_t)$ , mientras que las varianzas son  $\check{\sigma}_{j,t}^2 \equiv \text{Var}(\check{y}_{j,t} | y_{(j-1)}, \tilde{z}_t, \mathcal{D}_t)$ ,  $\check{C}_{j,i,t} \equiv \text{Var}(\check{y}_{i,t} | y_{(j-1)}, \tilde{z}_t, \mathcal{D}_t)$  y  $\check{\Sigma}_{j,i,t} \equiv \text{Cov}(\check{y}_{i,t}, \check{y}_{j,t} | y_{(j-1)}, \tilde{z}_t, \mathcal{D}_t)$ . Las particiones implican que  $y_{j,t-\bar{q}+1}$ ,  $\mu_{j,t-\bar{q}+1}$  y  $\check{\sigma}_{j,t}^2$  son escalares. Por lo tanto, siga los pasos descritos a continuación:

- a) Dado que  $\check{y}_{1,t} = y_{t-\bar{q}+1}$ , nótese que  $\check{\mu}_{1,1,t} = \tilde{\mu}_{t-\bar{q}+1}$  y  $\check{C}_{1,1,t} = \tilde{C}_{t-\bar{q}+1}$ .
- b) Por cada  $j \in \{1, \dots, m\}$ , considere el  $j$ -ésimo vector  $(m \times 1)$ ,  $\Phi_{j,\bar{q}}$  en la partición  $\Phi_{\bar{q}} \equiv [\Phi_{1,\bar{q}}, \dots, \Phi_{j,\bar{q}}, \dots, \Phi_{m,\bar{q}}]$ , y muestree  $y_{j,t-\bar{q}+1}$  a partir de:<sup>9</sup>

<sup>9</sup> Para fines de codificación, aún podría haber alguna inestabilidad numérica en casos en los que  $\check{\sigma}_{j,t}^2$  es positivo, si bien muy poca.

$$\left( y_{j,t-\bar{q}+1} \mid y_{(j-1)}, z_{t+1}, D_t \right) \sim N \left( \check{\mu}_{j,t-\bar{q}+1}^y, \left( \check{\sigma}_{j,t}^y \right)^2 \right)$$

$$\check{\mu}_{j,t-\bar{q}+1}^y = \begin{cases} \left( \frac{\Phi'_{j,\bar{q}} \Sigma_e^{-1} \check{\epsilon}_{t+1} + \left( \check{\sigma}_{j,t}^2 \right)^{-1} \mu_{j,t-\bar{q}+1}}{\Phi'_{j,\bar{q}} \Sigma_e^{-1} \Phi_{j,\bar{q}} + \left( \check{\sigma}_{j,t}^2 \right)^{-1}} \right), & \text{si } \check{\sigma}_{j,t}^2 > 0 \\ \mu_{j,t-\bar{q}+1}, & \text{si } \check{\sigma}_{j,t}^2 = 0 \end{cases}$$

$$\left( \check{\sigma}_{j,t}^y \right)^2 = \begin{cases} \frac{1}{\Phi'_{j,\bar{q}} \Sigma_e^{-1} \Phi_{j,\bar{q}} + \left( \check{\sigma}_{j,t}^2 \right)^{-1}}, & \text{si } \check{\sigma}_{j,t}^2 > 0 \\ 0, & \text{si } \check{\sigma}_{j,t}^2 = 0 \end{cases}$$

donde

$$\check{\mu}_{(j+1),(j+1),t}^y = \begin{cases} \check{\mu}_{j,j+1,t}^y + \frac{1}{\check{\sigma}_{j,t}^2} \check{\Sigma}_{j,j+1,t} \left( y_{j,t-\bar{q}+1} + \mu_{j,t-\bar{q}+1} \right), & \text{si } \check{\sigma}_{j,t}^2 > 0 \\ \check{\mu}_{j,j+1,t}^y, & \text{si } \check{\sigma}_{j,t}^2 = 0 \end{cases}$$

$$\check{C}_{(j+1),(j+1),t} = \begin{cases} \check{C}_{j,j+1,t} - \frac{1}{\check{\sigma}_{j,t}^2} \check{\Sigma}_{j,j+1,t} \check{\Sigma}'_{j,j+1,t} & \text{si } \check{\sigma}_{j,t}^2 > 0 \\ \check{C}_{j,j+1,t} & \text{si } \check{\sigma}_{j,t}^2 = 0 \end{cases}$$

5) El vector muestreado  $z_t$  se obtiene al concatenar

$$z_t = \left[ \check{z}'_t, y'_{t-\bar{q}+1} \right]'$$

*Suavizado.* Por lo que respecta a la distribución suavizada, su recuperación también está eventualmente sujeta a la inversión de la matriz de covarianzas  $R_t$  cuando se emplea el método estándar. Sin embargo, proponemos modificar el algoritmo FPMR para obtener la distribución, con lo cual se evitará lo anterior. En el sentido del algoritmo, desde  $t = (T-1)$  hasta  $t = 1$ , mantenemos las medias y las

---

En este caso, la restricción  $\check{\sigma}_{j,t}^2 > 0$  podría convertirse en algo como, por ejemplo,  $\check{\sigma}_{j,t}^2 > Toler \cdot \mu_{j,t-\bar{q}+1}$ , donde *Toler* es un parámetro de tolerancia relativa.



varianzas, en vez de muestrear los estados, y ajustamos los vectores y matrices en consecuencia. Al final, habremos recuperado la distribución suavizada. Nótese que el procedimiento no es lo mismo que mantener las medias y las varianzas en  $t = \tau$  al emplear el algoritmo FPMR para muestrear los estados en  $t = \tau + 1$ , pues las matrices y los vectores auxiliares no serán los mismos.

*Pronósticos, pronósticos retrospectivos y pronósticos inmediatos.* Al realizar pronósticos y pronósticos retrospectivos, todo lo que el algoritmo debe hacer es considerar la muestra ampliada para la cual no hay ninguna observación. Por lo tanto, el algoritmo FPMR también muestreará los valores faltantes en los ejercicios de pronóstico y pronóstico retrospectivo. El único requisito es que el paso del VAR, descrito en la sección 3.2, sólo debe considerar el periodo de muestra realmente disponible al muestrear parámetros condicionados sobre los valores observables y faltantes.

No se necesita poner atención adicional en los ejercicios de pronóstico inmediato, ya que la definición de esta requiere que haya al menos un valor observado. Por lo tanto, los pronósticos inmediatos se llevan a cabo en el periodo de muestra realmente disponible.

### 3.2 VAR bayesiano

De acuerdo con Canova (2007, cap. 10), el sistema lineal puede representarse en dos formas compañeras. La primera se define de la siguiente manera:

$$\mathbf{Y} = \mathbf{X}\Phi + \mathcal{E},$$

donde  $\mathbf{Y} \equiv [y_1, \dots, y_T]'$  es una matriz de variables ( $T \times m$ ),  $\Phi \equiv [\Phi_1, \dots, \Phi_q, \bar{\Phi}_c]'$  denota una matriz de coeficientes ( $k \times m$ ),  $\mathbf{X} \equiv [x_1, \dots, x_T]'$  es una matriz de regresores ( $T \times k$ ),  $x_t \equiv [y'_{t-1}, \dots, y'_{t-q}, \bar{x}'_{ct}]'$  es un vector de regresores ( $k \times 1$ ), y  $\mathcal{E} \equiv [e_1, \dots, e_T]'$  es una matriz de términos de error ( $T \times m$ ).<sup>10</sup>

<sup>10</sup> Nótese que la ecuación VAR para un periodo  $t$  es  $y_t = \Phi'x_t + e_t$ .

Nótese que cada columna de  $\mathbf{Y}$ ,  $\Phi$ , y  $\mathcal{E}$  corresponde a ecuaciones que describen la dinámica de una secuencia variable única  $\{y_{it}\}$  para  $i \in \{1, \dots, m\}$ :

$$\mathbf{x} = \begin{bmatrix} y_{1,1-1} & \cdots & y_{m,1-1} & \cdots & y_{1,1-q} & \cdots & y_{m,1-q} & \bar{x}_{c,11} & \cdots & \bar{x}_{c,m_c 1} \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ y_{1,t-1} & \cdots & y_{m,t-1} & \cdots & y_{1,t-q} & \cdots & y_{m,t-q} & \bar{x}_{c,1t} & \cdots & \bar{x}_{c,m_c t} \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ y_{1,T-1} & \cdots & y_{m,T-1} & \cdots & y_{1,T-q} & \cdots & y_{m,T-q} & \bar{x}_{c,1T} & \cdots & \bar{x}_{c,m_c T} \end{bmatrix},$$

$$\mathbf{Y} = \begin{bmatrix} y_{1,1} & \cdots & y_{i,1} & \cdots & y_{m,1} \\ \vdots & & \vdots & & \vdots \\ y_{1,T} & \cdots & y_{i,T} & \cdots & y_{m,T} \end{bmatrix},$$

$$\mathcal{E} = \begin{bmatrix} e_{11} & \cdots & e_{i1} & \cdots & e_{m1} \\ \vdots & & \vdots & & \vdots \\ e_{1T} & \cdots & e_{iT} & \cdots & e_{mT} \end{bmatrix},$$

$$\Phi = \begin{bmatrix} \Phi_{1;11} & \cdots & \Phi_{1;i1} & \cdots & \Phi_{1;m1} \\ \vdots & & \vdots & & \vdots \\ \Phi_{1;1m} & \cdots & \Phi_{1;im} & \cdots & \Phi_{1;mm} \\ \vdots & & \vdots & & \vdots \\ \Phi_{q;11} & \cdots & \Phi_{q;i1} & \cdots & \Phi_{q;m1} \\ \vdots & & \vdots & & \vdots \\ \Phi_{q;1m} & \cdots & \Phi_{q;im} & \cdots & \Phi_{q;mm} \\ \bar{\Phi}_{c;11} & \cdots & \bar{\Phi}_{c;i1} & \cdots & \bar{\Phi}_{c;m1} \\ \vdots & & \vdots & & \vdots \\ \bar{\Phi}_{c;1m_c} & \cdots & \bar{\Phi}_{c;im_c} & \cdots & \bar{\Phi}_{c;mm_c} \end{bmatrix}.$$

Al aplicar el operador  $\text{vec}(\cdot)$  en ambos lados de la ecuación 6, obtenemos la segunda forma compañera:<sup>11</sup>

<sup>11</sup> Ver el anexo para algunos resultados sobre el álgebra matricial y el operador  $\text{vec}(\cdot)$ .

$$\mathbf{y} = \mathbf{x}\phi + \varepsilon,$$

donde  $\mathbf{y} \equiv \text{vec}(\mathbf{Y})$  es un vector de variables endógenas ( $Tm \times 1$ ),  $\mathbf{x} \equiv (I_m \otimes \mathbf{X})$  es un vector de regresores ( $Tm \times km$ ),  $\phi \equiv \text{vec}(\Phi)$  es un vector de coeficientes ( $km \times 1$ ), y  $\varepsilon \equiv \text{vec}(\mathcal{E})$  es una matriz de términos de errores ( $Tm \times 1$ ).

Dados el modelo y un mapeo adecuado entre el conjunto de variables observables ( $\mathbf{Y}^{ob}$ ) y variables endógenas ( $\mathbf{Y}$ ) proporcionadas anteriormente, las siguientes secciones describen el procedimiento para obtener la distribución conjunta de  $(\mathbf{Y}, \Phi | \mathbf{Y}^{ob})$  mediante una simulación de Monte Carlo.

### 3.2.1 Cómo calcular la función de verosimilitud

Suponiendo que  $\varepsilon \sim N_{Tm}(0, \Sigma_\varepsilon)$ , para  $\Sigma_\varepsilon \equiv (\Sigma_\varepsilon \otimes I_T)$ , la función de verosimilitud es

$$p(\mathbf{y} | \phi, \Sigma_\varepsilon) = |2\pi\Sigma_\varepsilon|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{x}\phi)' \Sigma_\varepsilon^{-1} (\mathbf{y} - \mathbf{x}\phi)\right].$$

Sean  $\phi_{ml}$  y  $\Sigma_{ml}$  que denotan el estimador de máxima verosimilitud de  $\phi$  y su matriz de covarianzas:

$$\phi_{ml} \equiv (I_m \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}')\mathbf{y}; \quad \Sigma_{ml} \equiv (\Sigma_\varepsilon \otimes (\mathbf{X}'\mathbf{X})^{-1}).$$

Usando las definiciones de  $\mathbf{x}$  y  $\Sigma_\varepsilon$  nótese que la función de verosimilitud puede escribirse<sup>12</sup> de la siguiente manera:

$$\begin{aligned} \mathbf{8} \quad p(\mathbf{y} | \phi, \Sigma_\varepsilon) &= |2\pi|^{-\frac{Tm}{2}} |\Sigma_\varepsilon|^{-\frac{(T-k-m-1)+m+1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\bar{\Lambda}_s \Sigma_\varepsilon^{-1})\right) \\ &\quad \times |\Sigma_\varepsilon|^{-\frac{k}{2}} \exp\left(-\frac{1}{2} (\phi - \phi_{ml})' \Sigma_{ml}^{-1} (\phi - \phi_{ml})\right), \end{aligned}$$

donde

$$\mathbf{9} \quad \Phi_{ml} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}; \quad \bar{\Lambda}_s \equiv (\mathbf{Y} - \mathbf{X}\Phi_{ml})' (\mathbf{Y} - \mathbf{X}\Phi_{ml})$$

<sup>12</sup> Ver el anexo para más detalles.

### 3.2.2 Distribuciones conjuntas a priori y a posteriori

Dada la forma de la función de verosimilitud, consideramos que una distribución conjunta a priori natural conjugada para  $(\phi, \Sigma_e)$  es una generalización de lo que Gelman *et al.* (2003) llaman la distribución normal Wishart inversa. De manera que la distribución conjunta a priori se especifica de acuerdo con la siguiente estructura jerárquica:

$$\Sigma_e \sim W_m^{-1}(\delta_0, \Lambda_0^{-1}); \quad (\phi | \Sigma_e) \sim N_{km}(\phi_0, \Sigma_0); \quad \Sigma_0 \equiv \frac{1}{\sigma_0}(\Sigma_e \otimes \tilde{I}_k),$$

donde  $\delta_0 > (m+1)$  son los grados de libertad de la distribución de Wishart inversa (que puede entenderse como el tamaño de muestra a priori de  $\Sigma_e$ ),  $E\Sigma_e = \Lambda_0 / (\delta_0 - m - 1)$  es la estimación a priori para  $\Sigma_e$ ,  $\sigma_0$  es un parámetro de escala para la distribución condicional normal (que puede entenderse como el tamaño de muestra a priori de  $\phi$ ),  $\phi_0$  es la estimación a priori para  $\phi$ , e  $\tilde{I}_k$  es una matriz diagonal ( $k \times k$ ) cuyos términos diagonales satisfacen  $\tilde{I}_{\mathfrak{R}\mathfrak{R}} \in (0, 1]$  para  $\mathfrak{R} \in \{1, \dots, k\}$ .

Nótese que la matriz de identidad  $I_k$  es un caso específico de  $\tilde{I}_k$ . Además, esta definición permite el uso de un prior tipo Minnesota (ver, por ejemplo, Doan *et al.* 1984, y Litterman, 1986, para obtener más información sobre el prior de Minnesota original).

Esto implica que la FDP (función de densidad de probabilidad) conjunta a priori  $p(\phi, \Sigma_e) = p(\Sigma_e)p(\phi | \Sigma_e)$  es proporcional a:

$$\begin{aligned} 10 \quad p(\phi, \Sigma_e) &\propto |\Sigma_e|^{-\frac{(\delta_0+m+1)}{2}} \exp\left(-\frac{1}{2} \text{tr}(\Lambda_0 \Sigma_e^{-1})\right) \\ &\times |2\pi\Sigma_0|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\phi - \phi_0)' \Sigma_0^{-1}(\phi - \phi_0)\right]. \end{aligned}$$

Por lo tanto, la FDP conjunta a posteriori  $p(\phi, \Sigma_e | \mathbf{y}) \propto p(\phi, \Sigma_e) p(\mathbf{y} | \phi, \Sigma_e)$  es proporcional a:<sup>13</sup>

<sup>13</sup> Los detalles se muestran en el anexo.

11

$$p(\phi, \Sigma_e | \mathbf{y}) \propto |\Sigma_e|^{-\frac{(\delta_0 + T + m + 1)}{2}} \exp\left\{-\frac{1}{2} \text{tr}\left[(\Lambda_0 + \bar{\Lambda}_s + \Lambda_T) \Sigma_e^{-1}\right]\right\} \\ \times |2\pi \Sigma_T|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (\phi - \phi_T)' \Sigma_T^{-1} (\phi - \phi_T)\right],$$

donde

$$\Sigma_T^{-1} \equiv (\Sigma_0^{-1} + \Sigma_{\text{ml}}^{-1}) = (\Sigma_e^{-1} \otimes (\sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X}))$$

$$\phi_T \equiv \Sigma_T (\Sigma_0^{-1} \phi_0 + \Sigma_{\text{ml}}^{-1} \phi_{\text{ml}})$$

$$\Phi_T \equiv \text{vec}^{-1}(\phi_T)$$

$$= (\sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X})^{-1} (\sigma_0 \tilde{I}_k^{-1}) \Phi_0 + (\sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X}) \Phi_{\text{ml}}$$

$$\Lambda_T \equiv \Phi_0' (\sigma_0 \tilde{I}_k^{-1}) \Phi_0 + \Phi_{\text{ml}}' (\mathbf{X}'\mathbf{X}) \Phi_{\text{ml}} - \Phi_T' (\sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X}) \Phi_T.$$

Esto significa que la distribución conjunta a posteriori de  $(\phi, \Sigma_e | \mathbf{y})$  se especifica de acuerdo con la siguiente estructura jerárquica:

$$(\Sigma_e | \mathbf{y}) \sim W_m^{-1}(\delta_0 + T, (\Lambda_0 + \bar{\Lambda}_s + \Lambda_T)^{-1}); \quad (\phi | \Sigma_e, \mathbf{y}) \sim N_{km}(\phi_T, \Sigma_T).$$

### 3.2.3 Cómo especificar la distribución conjunta a priori

Por lo que respecta a la distribución conjunta a priori, proponemos la siguiente variante del prior de Minnesota:

12

$$\Phi_{0\ell;ij} = \begin{cases} 1 & \text{si } \ell = 1 \text{ e } i = j \\ 0 & \text{de lo contrario} \end{cases}, \quad \bar{\Phi}_{0c} = 0_{m \times m_c}$$

$$\tilde{I}_{\tilde{\mathcal{R}}\tilde{\mathcal{R}}} = \begin{cases} \tilde{\mathcal{R}}^{-\alpha} \text{ (decaimiento arm nico)} & \text{si } 1 \leq \tilde{\mathcal{R}} \leq (k - m_c) \\ \alpha^{1-\tilde{\mathcal{R}}} \text{ (decaimiento geom trico)} & \\ 1 & \text{de lo contrario} \end{cases},$$

donde  $\tilde{\mathfrak{K}} \equiv \text{int}\left(\frac{\mathfrak{K}-1}{m}\right) + 1$ ,  $\alpha > 0$ ,  $\mathfrak{K} \in \{1, \dots, k\}$ ,  $\ell \in \{1, \dots, q\}$ ,  $i \in \{1, \dots, m\}$ ,  $j \in \{1, \dots, m\}$ , y  $\bar{m}_c \in \{1, \dots, m_c\}$ , y de nuevo  $k \equiv mq + m_c$ .

Por congruencia, suponemos que  $E\Sigma_e = \bar{\Sigma}_e$ , donde  $E\Sigma_e = \frac{\Lambda_0}{(\delta_0 - m - 1)}$  y  $\bar{\Sigma}_e \equiv \frac{\bar{\Lambda}_s}{(T-1)}$ . Este supuesto implica que

$$\Lambda_0 = \frac{(\delta_0 - m - 1)}{(T-1)} \bar{\Lambda}_s.$$

Un prior difuso, de ser esa la elección, requiere bajos niveles para  $\delta_0 > (m+1)$  y  $\sigma_0 > 0$ . En el caso limitante  $\delta_0 \rightarrow (m+1)$  y  $\sigma_0 \rightarrow 0$  la distribución a posteriori implica los estimadores de máxima verosimilitud:<sup>14</sup>

$$\begin{aligned} (\Sigma_e | \mathbf{y}) &\sim W_m^{-1}\left(m+1+T, (\bar{\Lambda}_s)^{-1}\right); (\phi | \Sigma_e, \mathbf{y}) \sim N_{km}\left(\phi_{\text{ml}}, \Sigma_{\text{ml}}^{-1}\right) \text{ para el cual} \\ \Sigma_{\text{ml}}^{-1} &= E\Sigma_e = \frac{\bar{\Lambda}_s}{T}, \quad \Sigma_{\text{ml}} \equiv \left(\Sigma_e \otimes (\mathbf{X}'\mathbf{X})^{-1}\right), \quad \phi_{\text{ml}} \equiv \left(I_m \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\right) \mathbf{y}, \\ \text{y } \bar{\Lambda}_s &\equiv (\mathbf{Y} - \mathbf{X}\Phi_{\text{ml}})' (\mathbf{Y} - \mathbf{X}\Phi_{\text{ml}}). \end{aligned}$$

<sup>14</sup> Si  $\sigma_0 = 0$  y  $\delta_0 = (m+1)$ , concluimos que  $\Phi_T = \Phi_{\text{ml}}$  (o  $\phi_T = \phi_{\text{ml}}$ ),  $\Lambda_T = \Lambda_0 = \Sigma_0^{-1} = 0$ , y  $\Sigma_T = \Sigma_{\text{ml}}$ .





# 4

## Propiedades del VAR de frecuencia mixta: simulaciones





El marco más común para el cálculo de frecuencia mixta es cuando el investigador dispone de un conjunto de información con información regularmente espaciada en un panel completo. A manera de ejemplo, uno podría estar interesado en estimar un modelo con inflación y tasas de interés (medidas por lo general con frecuencia mensual) y el producto (de las cuentas nacionales con frecuencia trimestral). En lugar de agregar información sobre la inflación y las tasas de interés con frecuencia trimestral, para hacer coincidir los datos del producto, nuestro marco permite el cálculo conjunto del modelo.

Para el primer ejercicio, supóngase que el conjunto de datos tiene cuatro series de tiempo que abarcan una muestra  $T$  de 200 periodos con frecuencia mensual. En este ejercicio, las series artificiales  $N < 4$  tendrán vacíos igualmente distribuidos a lo largo de las series de tiempo ( $N$  cambiará con las simulaciones). Las estimaciones de las observaciones faltantes se obtendrán con el modelo de VAR y se compararán con los valores verdaderos del conjunto de datos artificial simulado inicialmente. Las series de tiempo para el ejercicio se simulan a partir del siguiente VAR con frecuencia mensual:

$$Y_t = AY_{t-1} + Se_t,$$

donde

$$A = \begin{bmatrix} 0.900 & 0.010 & -0.020 & 0.050 \\ 0.000 & 0.900 & -0.113 & -0.010 \\ 0.000 & 0.195 & 0.800 & 0.000 \\ -0.269 & 0.000 & 0.000 & 0.700 \end{bmatrix}$$

$$S = \begin{bmatrix} 0.01 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.01 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.01 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.01 \end{bmatrix}$$

Los valores de la matriz  $A$  fueron elegidos para asegurar la misma variación a largo plazo para cada variable observada en el vector  $Y_t$ , dada la falta de correlación condicional contemporánea proporcionada por la matriz  $S$ .

Se analizaron valores alternativos para la matriz  $S$ , imponiendo alguna correlación en todas las variables observadas. Los resultados fueron cualitativamente similares.

El cuadro 1 muestra la mediana de la raíz del error cuadrático medio (RECM) en 1,000 series de tiempo simuladas, medidas como una proporción de la desviación estándar no condicional de las series en el VAR verdadero. Cada línea muestra un total de seis simulaciones. Primero, cada grupo de dos columnas calcula la RECM utilizando la media, mediana y moda de las simulaciones en el muestreo de Gibbs como el pronóstico para la observación faltante. Para cada grupo de columnas, dos simulaciones intentan esclarecer la incertidumbre de la estimación de parámetros en el VAR al simular las observaciones faltantes: las columnas marcadas como *Valores verdaderos* reportan la RECM del VAR estimado frente a los valores reales de las observaciones faltantes; las columnas marcadas como *VAR verdadero* usan los valores verdaderos de los parámetros que describen el VAR para simular la observación faltante. El cuadro 1 también divide el análisis en función del número de series de tiempo observadas con valores faltantes en el VAR ( $N$ ) y la frecuencia de las observaciones faltantes ( $Q$ ,  $S$  y  $Y$  corresponden a la información en cada trimestre, semestre y año, respectivamente), mientras que se mantiene fijo el tamaño de la muestra,  $T = 200$ .

El cuadro 1 muestra tres resultados interesantes. Primero, como se esperaba, la RECM de las series de tiempo simuladas del VAR de frecuencia mixta depende de la frecuencia de las observaciones faltantes. Para todos los ejercicios que se muestran en el cuadro, al cambiar la frecuencia de la información faltante de datos

**Cuadro 1**

## MEDIANA DE LA RECM Y SIMULACIÓN DE VALORES FALTANTES

	<i>RECM-media</i>		<i>RECM-mediana</i>		<i>RECM-moda</i>	
	<i>Valores verdaderos</i>	<i>VAR verdadero</i>	<i>Valores verdaderos</i>	<i>VAR verdadero</i>	<i>Valores verdaderos</i>	<i>VAR verdadero</i>
	<i>T = 200; N = 1</i>					
F = Q	0.4181	0.1074	0.4182	0.1068	0.4531	0.2073
F = S	0.6592	0.4186	0.6572	0.4137	0.7966	0.6155
F = Y	0.9520	0.7655	0.9461	0.7592	1.1432	0.9899
	<i>T = 200; N = 2</i>					
F = Q	0.4143	0.0976	0.4143	0.0972	0.4414	0.1813
F = S	0.6706	0.4357	0.6634	0.4264	0.8381	0.6622
F = Y	1.1254	0.9445	1.0264	0.8336	1.8651	1.7677
	<i>T = 200; N = 3</i>					
F = Q	0.4301	0.0912	0.4293	0.0904	0.4555	0.1718
F = S	0.6677	0.3866	0.6630	0.3788	0.8505	0.6517
F = Y	1.6257	1.4690	1.1518	0.9336	6.6475	6.6008

trimestrales a anuales, se incrementó la RECM de las simulaciones, sin tomar en cuenta la estadística que resume el pronóstico de valores faltantes. A pesar de un gran aumento en la RECM al usar el modo de simulaciones en el muestreo de Gibbs con información anual en tres series de tiempo del VAR, todas las simulaciones mostraron variaciones similares en la RECM al cambiar la frecuencia de información faltante en el modelo.

En segundo lugar, algo que también se esperaba, la RECM de las series de tiempo simuladas depende del número de series de tiempo con observaciones faltantes en el VAR. Si bien este resultado no parece significativo cuando se dispone de información con una frecuencia trimestral, las simulaciones con datos semestrales y, sobre todo, anuales muestran que el aumento del número de series de tiempo con observaciones faltantes incrementa la RECM de las simulaciones, lo cual perjudica el rendimiento del algoritmo.

**Cuadro 2**

MEDIANA DE LA RECM Y SIMULACIÓN DE VALORES FALTANTES, T=400

	<i>RECM-media</i>		<i>RECM-mediana</i>		<i>RECM-moda</i>	
	<i>Valores verdaderos</i>	<i>VAR verdadero</i>	<i>Valores verdaderos</i>	<i>VAR verdadero</i>	<i>Valores verdaderos</i>	<i>VAR verdadero</i>
<i>T=400; N=1</i>						
F=Q	0.4119	0.0817	0.4120	0.0819	0.4366	0.1616
F=S	0.6107	0.3431	0.6100	0.3424	0.6915	0.4753
F=Y	0.8119	0.5677	0.8081	0.5619	0.9963	0.8011
<i>T= 400; N=2</i>						
F=Q	0.4103	0.0750	0.4103	0.0750	0.4245	0.1343
F=S	0.6083	0.3340	0.6067	0.3324	0.6802	0.4543
F=Y	0.8692	0.6334	0.8516	0.6090	1.1470	0.9759
<i>T= 400; N=3</i>						
F=Q	0.4185	0.0616	0.4184	0.0615	0.4310	0.1231
F=S	0.6144	0.3049	0.6140	0.3049	0.6870	0.4327
F=Y	1.0238	0.7834	0.9577	0.6968	1.7854	1.6559

Por último, hay una ventaja significativa al conocer los valores verdaderos de los parámetros en el sistema, lo que significa que los problemas de estimación deben abordarse con cuidado en aplicaciones empíricas. Al comparar los resultados de las columnas marcadas como *Valores verdaderos* con aquellos que utilizan el conjunto de parámetros verdaderos del VAR, marcado como *VAR verdadero*, las ventajas en función de la RECM al conocer exactamente los parámetros del modelo son significativas, sin tomar en cuenta la estadística que resume las simulaciones en el muestreo de Gibbs y la frecuencia de la información faltante.

El cuadro 2 considera el caso en el que el tamaño de la muestra disponible,  $T$ , con frecuencia mensual es el doble comparado con el del ejercicio de referencia. Los resultados muestran que el desempeño del VAR con datos faltantes con frecuencia trimestral no muestra una mejora significativa, a menos que se conozcan

todos los parámetros del sistema de ecuaciones. El coeficiente entre la RECM y la volatilidad de las series de tiempo simuladas permanece casi igual en ambos ejercicios con valores faltantes con frecuencia trimestral. Por otro lado, las simulaciones con variables observables en cada semestre o año muestran una mejora significativa, incluso teniendo en cuenta la incertidumbre relacionada con la estimación de los parámetros del VAR.





# 5

## Análisis empírico





**E**n esta sección, dos ejercicios muestran la capacidad que tiene el VAR de frecuencia mixta (VAR-FM) para simular series de tiempo de indicadores económicos que podrían ser útiles para un análisis económico. Los dos ejercicios exploran algunas de las propiedades analizadas en la sección anterior y, a la vez, muestran la flexibilidad del procedimiento cuando se describen conjuntos de datos. El primer ejercicio proporciona estimaciones de la serie de tiempo del PIB real de Brasil con frecuencia mensual, utilizando un amplio conjunto de variables observables que permiten construir un indicador de actividad económica de mayor frecuencia. En este ejercicio, las estimaciones muestran las consecuencias de especificar adecuadamente la matriz de transformación  $\Lambda_z$  en aplicaciones empíricas.

El segundo ejercicio intenta recuperar información de un amplio conjunto de variables observables para componer una serie de tiempo de tasas de desempleo mensuales con base en la última edición de la encuesta de desempleo de Brasil denominada “Pesquisa Nacional por Amostra de Domicílios Contínua” (Encuesta Nacional por Muestreo de Hogares Continua, PNAD-C). Los primeros resultados de dicha encuesta abarcan el periodo comprendido entre marzo de 2012 y marzo de 2015. Mediante un conjunto de modelos de VAR-FM con diferentes variables observables, incluyendo información de una versión anual de la PNAD, esta sección estudia la capacidad que tienen los modelos de generar estimaciones congruentes de la tasa de desempleo si los datos faltantes se encuentran irregularmente espaciados.

## 5.1 Cómo calcular el PIB real con frecuencia mensual

Una aplicación interesante de los VAR-FM es la obtención de estimaciones de alta frecuencia del producto interno bruto real utilizando indicadores de la actividad económica. En este caso, utilizamos un grupo de indicadores económicos para desarrollar un VAR-FM de cinco variables y analizar las propiedades de la combinación de pronósticos de estos VAR-FM. Estos indicadores incluyen la producción industrial total, la producción industrial de bienes de capital, las ventas minoristas, las exportaciones totales, las importaciones totales, el consumo de energía, el consumo de energía de la industria manufacturera, el coeficiente del saldo de la balanza comercial respecto al flujo comercial, la producción de petróleo y la producción de acero. Todos estos indicadores están disponibles con frecuencia mensual. Los VAR-FM incluyen variables ficticias estacionales y cambian el supuesto sobre las tendencias deterministas entre sin tendencia, tendencia lineal o una tendencia cuadrática.

En aplicaciones empíricas del VAR-FM, es importante caracterizar correctamente la configuración de la matriz  $\Lambda_z$  en la ecuación 2, relacionando variables observables a alta y baja frecuencia. La gráfica 1 a continuación muestra la trayectoria simulada de uno de los VAR-FM<sup>15</sup> correspondiente a las estimaciones de alta frecuencia de las tasas de crecimiento trimestral del PIB real (en logaritmos)<sup>16</sup> bajo dos supuestos para la matriz  $\Lambda_z$ : la línea verde supone una configuración simple, en la que el PIB real observado está directamente correlacionado con una observación de alta frecuencia en el último mes del trimestre; la línea naranja establece un marco adecuado,

---

<sup>15</sup> El VAR-FM presenta tres rezagos e incluye, además del logaritmo del PIB con frecuencia trimestral, factores estacionales y una tendencia lineal, el logaritmo de producción de bienes de capital, ventas minoristas, exportaciones y consumo de energía.

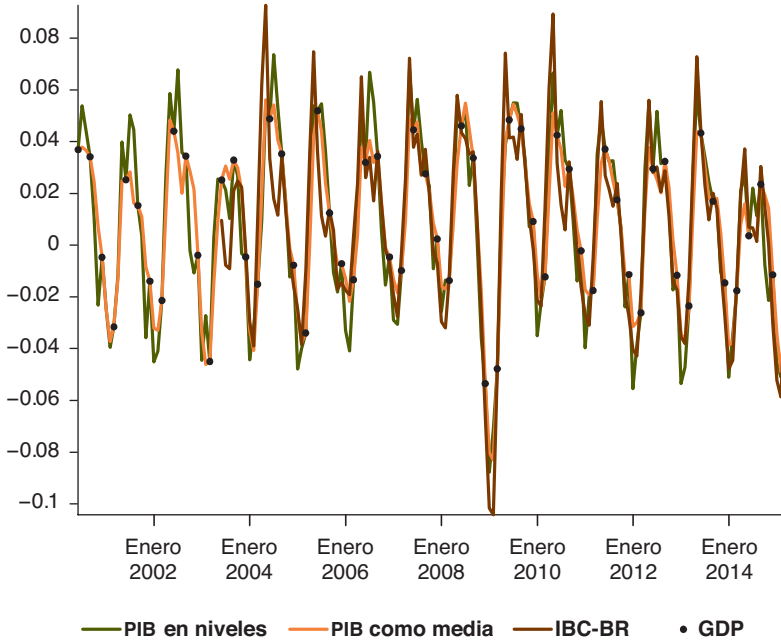
<sup>16</sup> Los datos sobre el PIB real utilizados en este ejercicio fueron publicados por el IBGE a finales de mayo de 2015, con la primera recolección de datos para el primer trimestre de 2015. Este ejercicio no maneja *pronósticos en tiempo real*, ya que un cambio metodológico en el cálculo del PIB impide una clara comparación entre recolecciones de datos recientes y anteriores.

---

### Gráfica 1

#### VARIACIÓN DEL PIB TRIMESTRAL EN ALTA FRECUENCIA

De junio de 2000 a marzo de 2015



en el que el PIB real observado corresponde a la suma del PIB real simulado en los tres meses del trimestre. La línea marrón muestra un indicador de estimaciones de alta frecuencia del PIB real estimado por el Banco Central do Brasil, el llamado IBC-BR. De la gráfica se desprende claramente que una configuración adecuada de la matriz  $\Lambda_z$  permite una mayor correspondencia entre la trayectoria simulada del VAR-FM y las estimaciones proporcionadas por el IBC-BR, ya que el PIB real en niveles tiende a sobreestimar sistemáticamente los movimientos estacionales al principio de cada año, y omite, al mismo tiempo, la ubicación de los picos a mediados de cada año. La configuración con el PIB real como media trimestral proporciona una mejor coincidencia con el IBC-BR en ambos puntos.

Para el ejercicio, se compara un pronóstico combinado de varios VAR-FM con los pronósticos de un proceso AR (1) simple para la tasa de crecimiento del PIB real (o de su logaritmo) con variables ficticias estacionales y un conjunto de VAR bayesianos con frecuencia trimestral con los mismos observables utilizados en los VAR-FM. El objetivo principal de utilizar los VAR bayesianos para comparar con los VAR-FM es medir el efecto que tiene el uso de información de alta frecuencia para desarrollar pronósticos, en lugar de simplemente agregar información a frecuencias más bajas. La comparación con un proceso AR(1) proporciona un punto de referencia en términos de capacidad de pronóstico del VAR-FM. Los VAR-FM se establecen arbitrariamente con tres rezagos para caracterizar correctamente la dinámica entre los datos de frecuencias mensuales y trimestrales. Los VAR bayesianos con frecuencia trimestral, por otro lado, debido al pequeño tamaño de la muestra, se estiman con sólo un rezago.

**Tabla 3**

PRONÓSTICO DEL PIB CON VAR-FM: RECM,  
DE DICIEMBRE DE 2010 A MARZO DE 2015

	<i>Observaciones</i>	<i>RECM</i>			<i>RECM relativo</i>	
		<i>VAR-FM</i>	<i>VAR</i>	<i>AR (1)</i>	<i>VAR</i>	<i>AR (1)</i>
t=0	18	0.0083	0.0200	0.0109	2.40	1.31
t=1	17	0.0170	0.0298	0.0157	1.76	0.92
t=2	16	0.0256	0.0291	0.0189	1.08	0.74
t=3	15	0.0298	0.0291	0.0256	0.98	0.86
t=4	14	0.0341	0.0415	0.0277	1.22	0.81

El cuadro 3 compara el resultado de los pronósticos en los tres modelos. En el cuadro,  $t=0$  muestra las propiedades de lo que se conoce como *pronóstico inmediato* en el VAR-FM, que es el pronóstico del PIB real en un momento dado  $t$ , dado que toda la información para otras variables, excepto el PIB, está disponible para el pronóstico. Cabe señalar que la información adicional con que se desarrolla el *pronóstico inmediato* sólo afecta a los pronósticos de los VAR-FM y del VAR bayesiano; por lo que respecta al proceso AR(1), el *pronóstico inmediato* es equivalente a un pronóstico a un paso adelante.

El cuadro 3 muestra las ventajas de utilizar información de alta frecuencia, en lugar de agregar datos a frecuencias más bajas. El desempeño del VAR-FM es, en el peor de los escenarios, similar al del VAR bayesiano, cuando se pronostican tres trimestres adelante. En el resto de los horizontes considerados, el desempeño del VAR-FM se impone claramente sobre el del VAR bayesiano. Sin embargo, a pesar de brindar un buen desempeño en términos de los *pronósticos inmediatos*, tanto el VAR-FM como el VAR bayesiano no son superiores a un sistema univariado estimado con frecuencia trimestral. Así, la principal ventaja de utilizar el VAR-FM, parece estar asociada con la capacidad de utilizar información de alta frecuencia para hacer inferencias sobre el periodo actual.

## 5.2 Cómo estimar la tasa de desempleo con datos irregularmente espaciados

Uno de los posibles usos de los VAR-FM es la simulación de información faltante en la serie de tiempo de una variable cuando existen otras variables sustitutas para aproximar la dinámica de la variable con información faltante. Aquí presentamos un conjunto de VAR-FM diseñados para estimar los valores faltantes de la tasa de desempleo en Brasil medida por la PNAD-C, una encuesta realizada por el Instituto Brasileiro de Geografia e Estatística (Instituto Brasileño de Geografía y Estadística, IBGE). Dicha encuesta muestra estimaciones de la tasa de desempleo en el trimestre anterior con una frecuencia mensual, abarcando casi cuatro veces el total de la población encuestada en la antigua encuesta realizada por el IBGE, la Pesquisa Mensual de Emprego, (Encuesta Mensual

**Cuadro 4**

**ENCUESTAS SOBRE DESEMPLEO EN BRASIL**

PME (IBGE) PED (DIEESE) PNAD (IBGE) PNAD-C (IBGE)

Frecuencia Mensual Mensual Anual Mensual

Cobertura Seis áreas metropolitanas Cuatro áreas metropolitanas Brasil Brasil

Índice de población (2012) 0.25 0.15 1.00 1.00

Periodo 1999m05-2015m03 1999m05-2015m03 1999-2013 2012m03-2015m03

Observación Se realizó una revisión importante en 2001 Limitada a las mismas áreas metropolitanas No se realizó en años de censo (2000 y 2010) Información sobre el último trimestre

<sup>1</sup> La información se refiere al promedio móvil del trimestre finalizado en un mes dado. Fuentes: IBGE y DIEESE (Departamento Intersindical de Estadística e Estudios Socioeconómicos/Departamento Intersindical de Estadística y Estudios Socioeconómicos)

de Empleo, PME). También hay una encuesta anual del IBGE llamada “Pesquisa Nacional por Amostra de Domicílios” (Encuesta Nacional por Muestreo de Hogares, PNAD), la cual también recaba información sobre el empleo, con una cobertura similar a la de la PNAD-C. La información de la situación laboral en Brasil es muy irregular. Sin embargo, dado el papel que desempeñan los mercados laborales en el análisis económico, generar una serie de tiempo de desempleo basada en la PNAD-C representa un dato valioso. El cuadro 4 resume las principales características de las cuatro encuestas sobre empleo en Brasil.

Para simular las series de tiempo del desempleo en Brasil con base en la PNAD-C, primero asumimos que las tasas de desempleo de la PNAD-C y la PNAD difieren sólo por un factor multiplicativo,<sup>17</sup> el cual se recupera al tomar en cuenta el único mes para el cual tenemos observaciones de ambas series, es decir, septiembre de 2012.<sup>18</sup> Posteriormente, ajustamos el nivel de las observaciones anuales de la PNAD y las combinamos con las observaciones mensuales de la PNAD-C. Dada una relación lineal entre la información superpuesta entre la PNAD y la PNAD-C, establecemos al menos un punto de datos para el periodo comprendido entre 1999 y 2011, el periodo previo a la publicación de los resultados de la PNAD-C. A continuación, se estimó un conjunto de VAR-FM de cuatro variables utilizando información de las encuestas PME y PED, con mediciones del desempleo y tasas de participación, datos de empleo formal del Catastro Geral de Empregados y Desempregados (registro general de empleados y desempleados, CAGED) y datos sobre el empleo en el sector manufacturero de la Confederação Nacional da Indústria (confederación nacional de la industria, CNI). Los VAR-FM incluyeron variables ficticias estacionales y se estimaron con tendencias deterministas lineales y cuadráticas y sin ellas. No se realizaron estimaciones que incluyeran un mayor número de variables observables debido al pequeño tamaño de la muestra y los grandes vacíos

---

<sup>17</sup> Los resultados no varían mucho cuando asumimos un factor aditivo en su lugar.

<sup>18</sup> Las cifras del PNAD evalúan la situación laboral el último día de septiembre de cada año.



de información de la PNAD-C. Además, en términos de la restricción de la selección de variables en los VAR, al menos una de las series de tiempo de desempleo se incluyó en la estimación, junto con los datos de la PNAD-C. Se utilizaron un total de 36 modelos para simular las series de tiempo de desempleo.

Las estimaciones finales basadas en la media y la mediana de las series de tiempo de la PNAD-C se muestran en la gráfica 2, junto con bandas de confianza basadas en la distribución de simulaciones en todos los modelos. De acuerdo con lo esperado, las bandas de confianza son mayores cuando el vacío entre la información disponible en la PNAD-C es también mayor. También de la manera esperada, las bandas de confianza se incrementan en los puntos de inflexión de las series de tiempo simuladas. Desde luego, como lo sugiere el análisis de la sección 4, a medida que haya más datos disponibles para la PNAD-C, debería ser posible obtener mejores estimaciones de la dinámica de las series de tiempo durante dichos puntos de inflexión. En consecuencia, esperamos que estos intervalos de confianza se hagan más pequeños con el incremento de la cantidad de información de la PNAD-C. También realizamos un ejercicio de pronóstico dentro de la muestra durante el último año de la misma, con el fin de establecer pesos en diferentes modelos basados en el error cuadrático medio de los modelos. La última serie de tiempo para el desempleo generada a partir de este ejercicio, la cual no se muestra en la gráfica 2, está muy cerca de la mediana de las simulaciones en todos los modelos.

La serie de tiempo simulada sigue una trayectoria cercana al compararla con las series de tiempo reales de las encuestas PME y PED. La principal diferencia entre estas dos encuestas y la trayectoria simulada de la PNAD-C que aquí se presenta está relacionada con la brecha entre el mínimo y el máximo de cada serie de tiempo: mientras que la PME y la PED, en promedio, mostraron tasas de desempleo superiores al 12% en el periodo comprendido entre 1999 y 2004, alcanzando incluso el 19% en una de las mediciones proporcionadas por la PED, el promedio de la PNAD-C en la muestra quedó en 10.5%. Por otro lado, para la última parte de la muestra, mientras que la PME alcanzó mínimos históricos del 4.3% en diciembre de 2013 y diciembre de 2014, el mínimo de la PNAD-C, que en

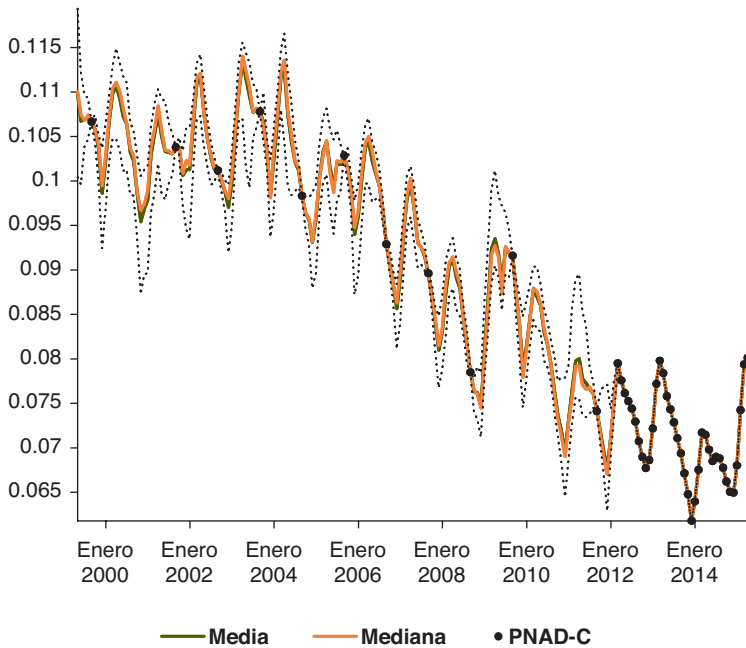
la muestra es una observación y no una simulación, se ubicó en el 6.2% en diciembre de 2013. El principal motivo de esta discrepancia es el uso de información anual de la PNAD para el periodo comprendido entre 1999 y 2011: sin el respaldo de dicha información, las trayectorias simuladas de la PNAD-C estarían más cerca de la media entre la PME y la PED para la primera parte de la muestra.

---

### Gráfica 2

#### SERIES DE TIEMPO SIMULADAS: PNAD-C

De Mayo de 1999 a Abril de 2015







# 6

## Conclusión



Este estudio presentó un nuevo y eficiente algoritmo para extraer una inferencia a partir de la distribución conjunta de un conjunto de datos con información faltante. El algoritmo aquí presentado es suficientemente flexible para adaptarse a la mayoría de los problemas que se enfrentan al tratar con conjuntos de datos irregularmente espaciados. También permite una configuración sencilla de las restricciones lineales en las variables observables, sin importar de la presencia de información faltante en dichas series de tiempo. Desde esta perspectiva, la implementación del modelo aprovecha la mayoría de los hallazgos en bibliografía más temprana publicada sobre el tema, generalizando algunos resultados y aplicaciones para el marco de la estimación de frecuencia mixta de modelos VAR.

El análisis con conjuntos de datos tanto simulados como reales muestra la flexibilidad del marco para manejar adecuadamente conjuntos de datos que presentan el tamaño habitual observado en las aplicaciones del mundo real. A pesar de que los resultados con datos simulados sugieren que es necesario realizar un análisis cuidadoso de la convergencia del muestreo de Gibbs, con el fin de garantizar que las series de tiempo simuladas se extraigan de valores de parámetros apropiados, la poca volatilidad de las simulaciones muestra que el algoritmo es una herramienta muy útil para manejar problemas asociados con información faltante en grandes conjuntos de datos.

En cuanto a las aplicaciones empíricas con datos reales, los principales resultados muestran la importancia de utilizar información

de alta frecuencia para hacer inferencias en un conjunto de datos macroeconómicos con valores faltantes. Las aplicaciones muestran la ventaja de utilizar información de alta frecuencia en dos contextos diferentes: primero, en términos de pronóstico, los VAR-FM que utilizan el marco aquí propuesto tuvieron un mejor desempeño que los VAR-FM bayesianos tradicionales estimados utilizando datos a frecuencias más bajas; segundo, el algoritmo aquí propuesto calculó trayectorias pasadas de variables con información faltante combinando un amplio conjunto de información, lo cual generó simulaciones con una dispersión relativamente pequeña en torno a su media.



# Anexos





## Anexo A. Algunos resultados

### *Álgebra matricial*

Recuerde que, para cualesquiera matrices  $A_{k \times \ell}$ ,  $B_{\ell \times m}$ , y  $C_{m \times n}$  el operador  $\text{vec}(\cdot)$  satisface las siguientes propiedades:

$$\text{tr}(A'B) = (\text{vec}(B))' \text{vec}(A) = (\text{vec}(A))' \text{vec}(B)$$

$$\text{vec}(ABC) = (C' \otimes \mathbf{A}) \text{vec}(B)$$

$$\text{vec}(ABC) = \text{vec}(AB \cdot C \cdot I_n) = (I_n \otimes AB) \text{vec}(C)$$

$$\text{vec}(ABC) = \text{vec}(I_k \cdot A \cdot BC) = (C'B' \otimes I_k) \text{vec}(A)$$

$$\text{vec}(AB) = \text{vec}(A \cdot B \cdot I_m) = (I_m \otimes \mathbf{A}) \text{vec}(B)$$

$$\text{vec}(AB) = \text{vec}(I_k \cdot A \cdot B) = (B' \otimes I_k) \text{vec}(A)$$

Recuerde que, para cualesquiera matrices  $A$ ,  $B$ ,  $C$  y  $D$ , el producto de Kronecker satisface (cuando tiene sentido) las siguientes propiedades:

$$(A \otimes B)(C \otimes D) = (AC \otimes BD)$$

$$(A \otimes B)^{-1} = (A^{-1} \otimes B^{-1})$$

$$\text{tr}(A \otimes B) = \text{tr}(A) \cdot \text{tr}(B)$$

$$(A \otimes B)' = (A' \otimes B')$$

$$|A \otimes B| = |A|^\ell |B|^k \text{ si } A \text{ es } (k \times k), \text{ y } B \text{ es } (\ell \times \ell).$$

***Estimador de máxima verosimilitud-distribución  
gaussiana multivariado***

El estimador proviene de  $0 = -\left(\Sigma_\varepsilon^{-\frac{1}{2}} \mathbf{x}\right)' \left[\Sigma_\varepsilon^{-\frac{1}{2}} (\mathbf{y} - \mathbf{x}\phi_{ml})\right]^{-1}$ , lo cual implica:

$$\begin{aligned} \phi_{ml} &= \left[ \left(\Sigma_\varepsilon^{-\frac{1}{2}} \mathbf{x}\right)' \left(\Sigma_\varepsilon^{-\frac{1}{2}} \mathbf{x}\right) \right]^{-1} \left(\Sigma_\varepsilon^{-\frac{1}{2}} \mathbf{x}\right)' \left(\Sigma_\varepsilon^{-\frac{1}{2}} \mathbf{y}\right) \\ &= \left[ \left(\left(\Sigma_\varepsilon \otimes I_T\right)^{-\frac{1}{2}} (I_m \otimes \mathbf{X})\right)' \left(\left(\Sigma_\varepsilon \otimes I_T\right)^{-\frac{1}{2}} (I_m \otimes \mathbf{X})\right) \right]^{-1} \left(\left(\Sigma_\varepsilon \otimes I_T\right)^{-\frac{1}{2}} (I_m \otimes \mathbf{X})\right)' \left(\left(\Sigma_\varepsilon \otimes I_T\right)^{-\frac{1}{2}} \mathbf{y}\right) \\ &= \left[ \left(\left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T\right) (I_m \otimes \mathbf{X})\right)' \left(\left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T\right) (I_m \otimes \mathbf{X})\right) \right]^{-1} \left(\left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T\right) (I_m \otimes \mathbf{X})\right)' \left(\left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T\right) \mathbf{y}\right) \\ &= \left[ \left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X}\right)' \left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X}\right) \right]^{-1} \left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X}\right)' \left(\left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T\right) \mathbf{y}\right) \end{aligned}$$

$$\begin{aligned}
&= \left[ \left( \Sigma_e^{-\frac{1}{2}} \otimes \mathbf{X}' \right) \left( \Sigma_e^{-\frac{1}{2}} \otimes \mathbf{X} \right) \right]^{-1} \left( \Sigma_e^{-\frac{1}{2}} \otimes \mathbf{X}' \right) \left( \Sigma_e^{-\frac{1}{2}} \otimes I_T \right) \mathbf{y} \\
&= \left( \Sigma_e^{-1} \otimes \mathbf{X}' \mathbf{X} \right)^{-1} \left( \Sigma_e^{-1} \otimes \mathbf{X}' \right) \mathbf{y} = \left( \Sigma_e \otimes (\mathbf{X}' \mathbf{X})^{-1} \right) \left( \Sigma_e^{-1} \otimes \mathbf{X}' \right) \mathbf{y} = \left( I_m \otimes (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \right) \mathbf{y}.
\end{aligned}$$

### ***Reescritura de la función de verosimilitud***

Tenga en cuenta que:

$$\begin{aligned}
p(\mathbf{y} | \phi, \Sigma_e) &= |2\pi \Sigma_e|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{y} - \mathbf{x}\phi)' \Sigma_e^{-1} (\mathbf{y} - \mathbf{x}\phi) \right] \\
&= |2\pi \Sigma_e|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \Sigma_e^{-\frac{1}{2}} (\mathbf{y} - \mathbf{x}\phi) \right]' \left[ \Sigma_e^{-\frac{1}{2}} (\mathbf{y} - \mathbf{x}\phi) \right] \right\}.
\end{aligned}$$

Y, por consiguiente, reemplazando  $(\phi_{ml} + \phi - \phi_{ml})$  con  $\phi$ , obtenemos:

$$\begin{aligned}
\left[ \Sigma_e^{-\frac{1}{2}} (\mathbf{y} - \mathbf{x}\phi) \right]' \left[ \Sigma_e^{-\frac{1}{2}} (\mathbf{y} - \mathbf{x}\phi) \right] &= \left[ \left( \Sigma_e^{-\frac{1}{2}} \otimes I_T \right) (\mathbf{y} - (I_m \otimes \mathbf{X})\phi) \right]' \left[ \left( \Sigma_e^{-\frac{1}{2}} \otimes I_T \right) (\mathbf{y} - (I_m \otimes \mathbf{X})\phi) \right] \\
&= \left[ \left( \Sigma_e^{-\frac{1}{2}} \otimes I_T \right) \mathbf{y} - \left( \Sigma_e^{-\frac{1}{2}} \otimes \mathbf{X} \right) \phi \right]' \left[ \left( \Sigma_e^{-\frac{1}{2}} \otimes I_T \right) \mathbf{y} - \left( \Sigma_e^{-\frac{1}{2}} \otimes \mathbf{X} \right) \phi \right] \\
&= \left[ \left( \Sigma_e^{-\frac{1}{2}} \otimes I_T \right) \mathbf{y} - \left( \Sigma_e^{-\frac{1}{2}} \otimes \mathbf{X} \right) \phi_{ml} \right]' \left[ \left( \Sigma_e^{-\frac{1}{2}} \otimes I_T \right) \mathbf{y} - \left( \Sigma_e^{-\frac{1}{2}} \otimes \mathbf{X} \right) \phi_{ml} \right] + \\
&\quad + (\phi - \phi_{ml})' \left( \Sigma_e^{-\frac{1}{2}} \otimes \mathbf{X} \right)' \left( \Sigma_e^{-\frac{1}{2}} \otimes \mathbf{X} \right) (\phi - \phi_{ml}) - 2(\phi - \phi_{ml})' \chi_e,
\end{aligned}$$

donde

$$\begin{aligned}
\chi_e &\equiv \left( \Sigma_e^{-\frac{1}{2}} \otimes \mathbf{X} \right)' \left[ \left( \Sigma_e^{-\frac{1}{2}} \otimes I_T \right) \mathbf{y} - \left( \Sigma_e^{-\frac{1}{2}} \otimes \mathbf{X} \right) \phi_{ml} \right] \\
&= \left( \Sigma_e^{-\frac{1}{2}} \otimes \mathbf{X} \right)' \left[ \left( \Sigma_e^{-\frac{1}{2}} \otimes I_T \right) - \left( \Sigma_e^{-\frac{1}{2}} \otimes \mathbf{X} \right) \left( I_m \otimes (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \right) \right] \mathbf{y} \\
&= \left[ \left( \Sigma_e^{-1} \otimes \mathbf{X}' \right) - \left( \Sigma_e^{-1} \otimes \mathbf{X}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \right) \right] \mathbf{y} \\
&= \left[ \left( \Sigma_e^{-1} \otimes \mathbf{X}' \right) - \left( \Sigma_e^{-1} \otimes \mathbf{X}' \right) \right] \mathbf{y} = 0.
\end{aligned}$$

Por lo tanto:

$$p(\mathbf{y}|\phi, \Sigma_e) = |2\pi \Sigma_e|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\phi_{ml})' (\Sigma_e \otimes I_T)^{-1} (\mathbf{y} - \mathbf{X}\phi_{ml}) - \frac{1}{2}(\phi - \phi_{ml})' (\Sigma_e^{-1} \otimes \mathbf{X}'\mathbf{X})(\phi - \phi_{ml})\right],$$

donde

$$\begin{aligned} (\mathbf{y} - \mathbf{X}\phi_{ml})' (\Sigma_e \otimes I_T)^{-1} (\mathbf{y} - \mathbf{X}\phi_{ml}) &= (\text{vec}(\mathbf{Y} - \mathbf{X}\Phi_{ml}))' (\Sigma_e^{-1} \otimes I_T) \text{vec}(\mathbf{Y} - \mathbf{X}\Phi_{ml}) \\ &= (\text{vec}(\mathbf{Y} - \mathbf{X}\Phi_{ml}))' \text{vec}(I_T (\mathbf{Y} - \mathbf{X}\Phi_{ml}) \Sigma_e^{-1}) \\ &= \text{tr}\left[(\mathbf{Y} - \mathbf{X}\Phi_{ml})' (\mathbf{Y} - \mathbf{X}\Phi_{ml}) \Sigma_e^{-1}\right]. \end{aligned}$$

Nótese también que

$$\text{vec}(\Phi_{ml}) = \phi_{ml} = (I_m \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}') \mathbf{y} = (I_m \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}') \text{vec}(\mathbf{Y}) = \text{vec}\left((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}\right),$$

$$\begin{aligned} \text{vec}(\mathbf{X}\Phi_{ml}) &= \mathbf{x}\phi_{ml} = (I_m \otimes \mathbf{X}) (I_m \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}') \mathbf{y} = (I_m \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}') \mathbf{y} \\ &= (I_m \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}') \text{vec}(\mathbf{Y}) = \text{vec}\left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}\right). \end{aligned}$$

### *La FDP conjunta a posteriori del VAR bayesiano*

Tenemos que calcular los siguientes términos sumatorios:

$$\begin{aligned} S_1 &= (\phi - \phi_0)' \Sigma_0^{-1} (\phi - \phi_0) + (\phi - \phi_{ml})' \Sigma_{ml}^{-1} (\phi - \phi_{ml}) \\ &= (\phi - \phi_T)' \Sigma_T^{-1} (\phi - \phi_T) + \phi_0' \Sigma_0^{-1} \phi_0 + \phi_{ml}' \Sigma_{ml}^{-1} \phi_{ml} - \phi_T' \Sigma_T^{-1} \phi_T, \end{aligned}$$

$$\begin{aligned} S_2 &= \phi_0' \Sigma_0^{-1} \phi_0 + \phi_{ml}' \Sigma_{ml}^{-1} \phi_{ml} - \phi_T' \Sigma_T^{-1} \phi_T \\ &= \text{vec}(\Phi_0)' (\Sigma_e^{-1} \otimes \sigma_0 \tilde{I}_k) \text{vec}(\Phi_0) + \\ &\quad + \text{vec}(\Phi_{ml})' (\Sigma_e^{-1} \otimes (\mathbf{X}'\mathbf{X})) \text{vec}(\Phi_{ml}) - \text{vec}(\Phi_T)' (\Sigma_e^{-1} \otimes (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})) \text{vec}(\Phi_T) \\ &= \text{vec}(\Phi_0)' \text{vec}\left((\sigma_0 \tilde{I}_k) \Phi_0 \Sigma_e^{-1}\right) + \\ &\quad + \text{vec}(\Phi_{ml})' \text{vec}\left((\mathbf{X}'\mathbf{X}) \Phi_{ml} \Sigma_e^{-1}\right) - \text{vec}(\Phi_T)' \text{vec}\left((\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X}) \Phi_T \Sigma_e^{-1}\right) \\ &= \text{tr}\left[\left(\Phi_0' (\sigma_0 \tilde{I}_k) \Phi_0 + \Phi_{ml}' (\mathbf{X}'\mathbf{X}) \Phi_{ml} - \Phi_T' (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X}) \Phi_T\right) \Sigma_e^{-1}\right] = \text{tr}\left[\Lambda_T \Sigma_e^{-1}\right], \end{aligned}$$

donde

$$\Sigma_T^{-1} \equiv (\Sigma_0^{-1} + \Sigma_{ml}^{-1}) = \Sigma_e^{-1} \otimes (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X}),$$

$$\begin{aligned} \phi_T &\equiv \Sigma_T (\Sigma_0^{-1} \phi_0 + \Sigma_{ml}^{-1} \Phi_{ml}) \\ &= \left( I_m \otimes (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\sigma_0 \tilde{I}_k) \right) \phi_0 + \left( I_m \otimes (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X}) \right) \phi_{ml} \end{aligned}$$

$$\Phi_T \equiv \text{vec}^{-1}(\phi_T) = (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\sigma_0 \tilde{I}_k) \Phi_0 + (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X}) \Phi_{ml},$$

$$\Lambda_T \equiv \Phi_0' (\sigma_0 \tilde{I}_k) \Phi_0 + \Phi_{ml}' (\mathbf{X}'\mathbf{X}) \Phi_{ml} - \Phi_T' (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X}) \Phi_T.$$

La derivación de la expresión correspondiente a  $\Phi_T$  se realiza de la siguiente manera:

$$\begin{aligned} \text{vec}(\Phi_T) &\equiv \left( I_m \otimes (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\sigma_0 \tilde{I}_k) \right) \text{vec}(\Phi_0) + \left( I_m \otimes (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X}) \right) \text{vec}(\Phi_{ml}) \\ &= \text{vec} \left( (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\sigma_0 \tilde{I}_k) \Phi_0 + (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X}) \Phi_{ml} \right). \end{aligned}$$

Nótese también que

$$\begin{aligned} \phi_T &\equiv \Sigma_T (\Sigma_0^{-1} \phi_0 + \Sigma_{ml}^{-1} \phi_{ml}) = (\Sigma_0^{-1} + \Sigma_{ml}^{-1})^{-1} (\Sigma_0^{-1} \phi_0 + \Sigma_{ml}^{-1} \phi_{ml}) \\ &= \left( \sigma_0 (\Sigma_0^{-1} \otimes \tilde{I}_k^{-1}) + (\Sigma_e^{-1} \otimes (\mathbf{X}'\mathbf{X})) \right)^{-1} \left( \sigma_0 (\Sigma_0^{-1} \otimes \tilde{I}_k^{-1}) \phi_0 + (\Sigma_e^{-1} \otimes (\mathbf{X}'\mathbf{X})) \phi_{ml} \right) \\ &= \left( \Sigma_e^{-1} \otimes (\sigma_0 \tilde{I}_k^{-1} + (\mathbf{X}'\mathbf{X})) \right)^{-1} \left( (\Sigma_e^{-1} \otimes \sigma_0 \tilde{I}_k^{-1}) \phi_0 + (\Sigma_e^{-1} \otimes (\mathbf{X}'\mathbf{X})) \phi_{ml} \right) \\ &= \left( \Sigma_e \otimes (\sigma_0 \tilde{I}_k^{-1} + (\mathbf{X}'\mathbf{X})) \right)^{-1} \left( (\Sigma_e^{-1} \otimes \sigma_0 \tilde{I}_k^{-1}) \phi_0 + (\Sigma_e^{-1} \otimes (\mathbf{X}'\mathbf{X})) \phi_{ml} \right) \\ &= \left( I_m \otimes (\sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X}) \right)^{-1} (\sigma_0 \tilde{I}_k^{-1}) \phi_0 + \left( I_m \otimes (\sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X}) \right)^{-1} (\mathbf{X}'\mathbf{X}) \phi_{ml}. \end{aligned}$$

## Anexo B. Pruebas

### Proposición 1

La estructura de Markov de la ecuación de evolución de cualquier MLD implica

$$\Pr(Z_T | \mathcal{D}_T) = \Pr(z_T | \mathcal{D}_T) \Pr(z_{T-1} | z_T, \mathcal{D}_{T-1}) \dots \Pr(z_1 | z_2, \mathcal{D}_1).$$

## Prueba

Nótese que  $\Pr(Z_T | \mathcal{D}_T)$  puede ampliarse de la siguiente manera:

$$\begin{aligned}
 \Pr(Z_T | \mathcal{D}_T) &= \Pr(z_T, Z_{T-1} | \mathcal{D}_T) = \Pr(z_T | \mathcal{D}_T) \Pr(Z_{T-1} | z_T, \mathcal{D}_T) \\
 &= \Pr(z_T | \mathcal{D}_T) \Pr(z_{T-1}, Z_{T-2} | z_T, \mathcal{D}_T) \\
 &= \Pr(z_T | \mathcal{D}_T) \Pr(z_{T-1} | z_T, \mathcal{D}_T) \Pr(Z_{T-2} | z_{T-1}, z_T, \mathcal{D}_T) \\
 &= \Pr(Z_T | \mathcal{D}_T) \Pr(z_{T-1} | z_T, \mathcal{D}_T) \Pr(z_{T-2}, Z_{T-3} | z_{T-1}, z_T, \mathcal{D}_T) \\
 &= \Pr(z_T | \mathcal{D}_T) \dots \underbrace{\Pr(z_{T-2} | z_{T-1}, z_T, \mathcal{D}_T)}_{\Pr(z_{T-2} | z_{t > (T-2)}, \mathcal{D}_T)} \Pr(Z_{T-3} | z_{T-2}, z_{T-1}, z_T, \mathcal{D}_T) \\
 &\quad \vdots \\
 &= \Pr(z_T | \mathcal{D}_T) \Pr(z_{T-1} | z_{t > (T-1)}, \mathcal{D}_T) \Pr(z_{T-2} | z_{t > (T-2)}, \mathcal{D}_T) \dots \Pr(z_1 | z_{t > 1}, \mathcal{D}_T).
 \end{aligned}$$

Sean entonces  $y_{t > \bar{t}}^{ob} \equiv [y_{\bar{t}+1}^{ob}, y_T^{ob}]$  y  $z_{t > \bar{t}} \equiv [z_{\bar{t}+1}, z_T]$  que denotan los conjuntos de todas las variables observables y de estado para los periodos comprendidos entre  $\bar{t} + 1$  y  $T$ . Recuerde que la ecuación de observación implica que, supeditada a  $z_t$ ,  $y_t^{ob}$  no depende de valores pasados de la variable de estado. Como se muestra a continuación, la propiedad markoviana, al enfrentar este hecho, implica que nos permite mostrar que  $\Pr(z_{T-j} | z_{t > T-j}, \mathcal{D}_T) = \Pr(z_{T-j} | z_{T-j+1}, \mathcal{D}_{T-j})$ . De hecho:

Utilizando nuestro resultado anterior, concluimos que

$$\begin{aligned}
 \Pr(z_{T-j} | z_{t > T-j}, \mathcal{D}_T) &= \Pr(z_{T-j} | z_{t > T-j}, y_{t > T-j}^{ob}, \mathcal{D}_{T-j}) \\
 &= \frac{\Pr(z_{T-j}, y_{t > T-j}^{ob} | z_{t > T-j}, \mathcal{D}_{T-j})}{\Pr(y_{t > T-j}^{ob} | z_{t > T-j}, \mathcal{D}_{T-j})}, \\
 &= \frac{\Pr(y_{t > T-j}^{ob} | z_{T-j}, z_{t > T-j}, \mathcal{D}_{T-j}) \Pr(z_{T-j} | z_{t > T-j}, \mathcal{D}_{T-j})}{\Pr(y_{t > T-j}^{ob} | z_{t > T-j}, \mathcal{D}_{T-j})}
 \end{aligned}$$

Debido a la ecuación observada

$$\begin{aligned}
&= \frac{\Pr\left(y_{t>T-j}^{ob} \mid z_{t>T-j}, \mathcal{D}_{T-j}\right) \Pr\left(z_{T-j} \mid z_{t>T-j}, \mathcal{D}_{T-j}\right)}{\Pr\left(y_{t>T-j}^{ob} \mid z_{t>T-j}, \mathcal{D}_{T-j}\right)} \\
&= \Pr\left(z_{T-j} \mid z_{t>T-j}, \mathcal{D}_{T-j}\right) \\
&= \Pr\left(z_{T-j} \mid z_{T-j+1}, z_{t>T-j+1}, \mathcal{D}_{T-j}\right) \\
&= \frac{\Pr\left(z_{T-j}, z_{t>T-j+1} \mid z_{T-j+1}, \mathcal{D}_{T-j}\right)}{\Pr\left(z_{t>T-j+1} \mid z_{T-j+1}, \mathcal{D}_{T-j}\right)} \\
&= \frac{\Pr\left(z_{t>T-j+1} \mid z_{T-j}, z_{T-j+1}, \mathcal{D}_{T-j}\right) \Pr\left(z_{T-j} \mid z_{T-j+1}, \mathcal{D}_{T-j}\right)}{\Pr\left(z_{t>T-j+1} \mid z_{T-j+1}, \mathcal{D}_{T-j}\right)}
\end{aligned}$$

Propiedad markoviana

$$\begin{aligned}
&= \frac{\Pr\left(z_{t>T-j+1} \mid z_{T-j+1}, \mathcal{D}_{T-j}\right) \Pr\left(z_{T-j} \mid z_{T-j+1}, \mathcal{D}_{T-j}\right)}{\Pr\left(z_{t>T-j+1} \mid z_{T-j+1}, \mathcal{D}_{T-j}\right)} \\
&= \Pr\left(z_{T-j} \mid z_{T-j+1}, \mathcal{D}_{T-j}\right).
\end{aligned}$$

$$\Pr\left(Z_T \mid \mathcal{D}_T\right) = \Pr\left(z_T \mid \mathcal{D}_T\right) \Pr\left(z_{T-1} \mid z_T, \mathcal{D}_{T-1}\right) \dots \Pr\left(z_1 \mid z_2, \mathcal{D}_1\right).$$

■



## Proposición 2

La densidad condicional de  $(y_{t-q+1} | z_{t+1}, \mathcal{D}_t)$  es proporcional a

$$\Pr(y_{t-q+1} | z_{t+1}, \mathcal{D}_t) \propto \Pr(\tilde{e}_{t+1} | \tilde{z}_{t+1}, y_{t-q+1}, \mathcal{D}_t) \Pr(y_{t-q+1} | \tilde{z}_{t+1}, \mathcal{D}_t).$$

### Prueba

$$\begin{aligned} \Pr(y_{t-q+1} | z_{t+1}, \mathcal{D}_t) &= \Pr(y_{t-q+1} | y_{t+1}, \tilde{z}_t, \mathcal{D}_t) \\ &= \frac{\Pr(y_{t+1}, y_{t-q+1} | \tilde{z}_t, \mathcal{D}_t)}{\Pr(y_{t+1} | \tilde{z}_t, \mathcal{D}_t)} = \frac{\Pr(y_{t+1} | \tilde{z}_t, y_{t-q+1}, \mathcal{D}_t) \Pr(y_{t-q+1} | \tilde{z}_t, \mathcal{D}_t)}{\Pr(y_{t+1} | \tilde{z}_t, \mathcal{D}_t)} \end{aligned}$$

Condicionando el primer término también contenido en  $\tilde{e}_{t+1}$

$$\begin{aligned} &= \frac{\Pr(y_{t+1}, \tilde{e}_{t+1} | \tilde{z}_t, y_{t-q+1}, \mathcal{D}_t)}{\Pr(\tilde{e}_{t+1} | y_{t+1}, \tilde{z}_t, y_{t-q+1}, \mathcal{D}_t)} \frac{\Pr(y_{t-q+1} | \tilde{z}_t, \mathcal{D}_t)}{\Pr(y_{t+1} | \tilde{z}_t, \mathcal{D}_t)} \\ &= \frac{\Pr(y_{t+1} | \tilde{z}_t, y_{t-q+1}, \tilde{e}_{t+1}, \mathcal{D}_t)}{\Pr(\tilde{e}_{t+1} | y_{t+1}, \tilde{z}_t, y_{t-q+1}, \mathcal{D}_t)} \frac{\Pr(\tilde{e}_{t+1} | \tilde{z}_t, y_{t-q+1}, \mathcal{D}_t) \Pr(y_{t-q+1} | \tilde{z}_t, \mathcal{D}_t)}{\Pr(y_{t+1} | \tilde{z}_t, \mathcal{D}_t)} \end{aligned}$$

Condicionado en  $\tilde{e}_{t+1}$ , el primer término no depende de  $y_{t-q+1}$

$$= \frac{\Pr(y_{t+1} | \tilde{z}_t, \tilde{e}_{t+1}, \mathcal{D}_t)}{\Pr(\tilde{e}_{t+1} | y_{t+1}, \tilde{z}_t, \mathcal{D}_t)} \frac{\Pr(\tilde{e}_{t+1} | \tilde{z}_t, y_{t-q+1}, \mathcal{D}_t) \Pr(y_{t-q+1} | \tilde{z}_t, \mathcal{D}_t)}{\Pr(y_{t+1} | \tilde{z}_t, \mathcal{D}_t)}.$$

Debido a que los términos  $\Pr(y_{t+1} | \tilde{z}_t, \tilde{e}_{t+1}, \mathcal{D}_t)$ ,  $\Pr(\tilde{e}_{t+1} | y_{t+1}, \tilde{z}_t, \mathcal{D}_t)$  y  $\Pr(y_{t+1} | \tilde{z}_t, \mathcal{D}_t)$  no son funciones de  $y_{t-q+1}$ , estos son constantes con respecto a  $\Pr(y_{t-q+1} | z_{t+1}, \mathcal{D}_t)$ .

Por lo tanto, obtenemos:

$$\Pr(y_{t-q+1} | z_{t+1}, \mathcal{D}_t) \propto \Pr(\tilde{e}_{t+1} | \tilde{z}_{t+1}, y_{t-q+1}, \mathcal{D}_t) \Pr(y_{t-q+1} | \tilde{z}_{t+1}, \mathcal{D}_t).$$

■



# Bibliografía



- Basdevant, Olivier (2003), *On Applications of State-space Modeling in Macroeconomics*, Reserve Bank of New Zealand Discussion Paper Series, núm. DP2003/02, Reserve Bank of New Zealand.
- Canova, Fabio (2007), *Methods for Applied Macroeconomic Research*, Princeton University Press, Princeton.
- Carter, Christopher K., y Robert Kohn (1994), "On Gibbs Sampling for State Space Models," *Biometrika*, vol. 81, núm. 3, agosto, pp. 541-553.
- Doan, Thomas, Robert B. Litterman, y Christopher A. Sims (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, vol. 3, núm. 1, enero, pp. 1-100.
- Durbin, James, y Siem Jan Koopman (2012), *Time Series Analysis by State Space Methods*, 2da. ed., núm. 38, Oxford Statistical Science, Oxford University Press, agosto.
- Eraker, Bjørn, Ching Wai (Jeremy) Chiu, Andrew T. Foerster, Tae Bong Kim, y Hernán D. Seoane (2015), "Bayesian Mixed Frequency VARs," *Journal of Financial Econometrics*, vol. 13, núm. 3, pp. 698-721.
- Frühwirth-Schnatter, Sylvia (1994), "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis*, vol. 15, núm. 2, marzo, pp. 183-202.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, y Donald B. Rubin (2003), *Bayesian Data Analysis, Second Edition*, Chapman and Hall/CRC, Boca Raton, Fla., julio.
- Hamilton, James D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, N. J.

- Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. 82, núm. 1, pp. 35-45.
- Litterman, Robert B. (1986), "Forecasting with Bayesian Vector Autoregressions-five Years of Experience," *Journal of Business & Economic Statistics*, vol. 4, núm. 1, pp. 25-38.
- Prado, Raquel, y Mike West (2010), *Time Series*, Chapman & Hall, Londres.
- Schorfheide, Frank, y Dongho Song (2015), "Real-time Forecasting with a Mixed-frequency VAR," *Journal of Business & Economic Statistics*, vol. 33, núm. 3, pp. 366-380.
- West, Mike (1996), "Bayesian Time Series: Models and Computations for the Analysis of Time Series in the Physical Sciences," en K. M. Hanson y R. N. Silver (eds.), *Maximum Entropy and Bayesian Methods*, núm. 79, Fundamental Theories of Physics, Springer Netherlands, pp. 23-34.
- West, Mike (1997), "Time Series Decomposition," *Biometrika*, vol. 84, núm. 2, enero, pp. 489-494.
- West, Mike, y Jeff Harrison (1997), *Bayesian Forecasting and Dynamic Models*, Springer, Nueva York.







# Not Just Another Mixed Frequency Paper







Sergio Alfonso Lago Alves  
Angelo Marsiglia Fasolo

*Not Just Another Mixed  
Frequency Paper*

CENTRAL BANK AWARD RODRIGO GÓMEZ, 2016

---

CENTER FOR LATIN AMERICAN MONETARY STUDIES

Mexico City, 2019



## PREFACE

This study presents a new algorithm, based on a two-part Gibbs sampler with a forward-filtering-backward-sampling (FFBS) method, to recover the joint distribution of missing observations in a mixed-frequency dataset. The new algorithm relaxes most of the constraints usually presented in the literature, namely: 1) it does not require at least one time series to be observed every period; 2) it provides an easy way to add linear restrictions based on the state space representation of the VAR; 3) it does not require regularly-spaced time series at lower frequencies; and, 4) it avoids degeneration problems arising when states, or linear combination of states, are actually observed. In addition, the algorithm is well suited for embedding high-frequency real-time information for improving nowcasts and forecasts of lower frequency time series. We evaluate the properties of the algorithm using simulated data. Moreover, as empirical applications, we simulate monthly Brazilian GDP, comparing our results to the Brazilian IBC-BR, and recover what would historical PNAD-C unemployment rates look like prior to 2012.

Keywords: Gibbs sampling, forward-filtering-backward-sampling (FFBS), mixed frequency data, Bayesian methods, real-time data, macroeconomic forecasting, vector autoregressions, PNAD.

JEL classification: C11, C32, C53, E27.

## ABOUT THE AUTHORS

**S**ergio Afonso Lago Alves is currently the Deputy Head of the Research Department, Banco Central do Brasil. He has a BA degree in Mechanical Engineering from the Instituto Tecnológico de Aeronáutica (1996), a MSc degree in Air Transportation from the Instituto Tecnológico de Aeronáutica (2000), a MSc degree in Economics from the Universidade de Brasília (2005), and a PhD degree in Economics from the University of California-Santa Cruz (2011). His research has focused on monetary economics, trend inflation, unemployment, zero lower bound on nominal interest rates, Bayesian inference, DSGE modelling, and quantitative methods. <sergio.lago@bcb.gov.br>.

**A**ngelo Marsiglia Fasolo is currently a senior advisor at the Research Department of the Banco Central do Brasil. He has a MSc degree in Economics from Universidade Federal do Rio Grande do Sul (2003) and a PhD degree in Economics from Duke University (2010). He has experience in quantitative methods and macroeconomic modelling, with publications about the solution and estimation of DSGE models, international economics and monetary and fiscal policy coordination. <angelo.fasolo@bcb.gov.br>.

## ACKNOWLEDGEMENTS

**T**he authors would like to thank Hernán D. Seoane and Andrew Foerster for early comments on the main ideas of this research, and the participants of seminars at the Banco Central do Brasil. The views expressed here are of the authors and do not represent those of the Banco Central do Brasil.



# 1

# Introduction



Many economic time series of developing and underdeveloped countries suffer from issues preventing their information content to be fully explored. These common issues usually arise from 1) *short length*; 2) measurement only available in *low frequencies*, such as quarterly or annual; 3) discontinuity due to changes in measurement methodology into another; 4) measurement or technical issues leading to *randomly spaced missing observations* in time series; and 5) *frequency switch* when the frequency of observations changes along the sample, for example, series starts as annual but then change to quarterly after a couple of years.

The short length is a common issue that might affect developed countries as well, but is more often present in developing countries. In a nutshell, the problem arises when a time series either ceases to be measured or has only recently started to be measured. Understanding the inference limitations caused by low-frequency time series is straightforward due to the reduced number of observations available. The discontinuity problem arises when a time series has its measurement method changed while not updating the method for past observations. In this case, data properties might be significantly affected due to the new measurement method, in such a way that observations obtained from the old and new method must be considered as two different time series.

Randomly spaced missing observations may arise by either measurement flaws when collecting and treating raw data, or by



deliberately excluding observations that are considered outliers or not consistent with the effect of hypotheses one wants to measure and test. Finally, series with frequency switches may be seen as a special case of series with missing observations, except that spacing between observations is set from a discretionary choice. For simplicity, we treat this case just like we treat randomly spaced missing observations.<sup>1</sup> In the context of multivariate analysis, a problematic situation occurs when a combination of the issues listed above leads to what is known as *unequally spaced mixed-frequency* data sets, where the extreme case shows periods with no available observations at all; that is, all elements from the time series might be missing in given a time.

Under mixed-frequency, multivariate framework, usual choices might compromise the statistical inference, leading to wrong conclusions. For instance, the analyst might either fill the gaps in time series with ad hoc values, that is, mechanical approaches without formal statistical support, such as interpolations or trimming out observations in periods in which one or more time series have missing observations. Alternatively, one might consider reducing the time series frequency of the whole dataset in order to avoid dealing with mixed-frequency issues. Both trimming out periods or reducing frequencies might lead to the loss of relevant information, though. Hence, a problem arises in finding ways to simultaneously explore the full information content available in the mixed-frequency dataset, while avoiding the use of ad hoc methods to fill information gaps.

In economics, the literature has evolved into the development of Bayesian Gibbs samplers to recover the entire joint distribution of the missing observations, rather than pin-pointing single values to fill the gaps. In this study, we contribute to the literature by bringing an efficient and intuitive way to recover this joint distribution using a two-part Gibbs sampler, built mostly upon recent important contributions to the literature on mixed-frequency filtering by Schorfheide and Song (2015) and Eraker et al. (2015), and, from the statistics

---

<sup>1</sup> See Prado and West (2010, Section 4.3.3) for a better discussion on series with discretionary spaced missing observations.

literature, on the efficient algorithm described in West (1996, 1997) and West and Harrison (1997, Section 15.3.2).

We highlight the point that the literature in statistics has considered the problem of recovering the entire joint distribution of missing observations way before the literature in economics, with efficient algorithms using Gibbs samplers, like the one developed by West and detailed further on, developed during the 1990s. Therefore, the main point of this paper is to adapt and bring some of the early developments from the statistics literature to applications in economics. That been said, our view is that the method proposed by West (1996, 1997) is more efficient, simpler and more intuitive than the ones proposed by Schorfheide and Song (2015) and Eraker et al. (2015), even though generating about the same results.

As in Schorfheide and Song (2015) and Eraker et al. (2015), we assume that a VAR with unknown parameters is able to describe the dynamics of the multivariate time series if their missing values were all observed. The task of the Gibbs sampler is to recover the joint distribution of the missing values and the unknown VAR parameters. As in West (1996, 1997) and West and Harrison (1997, Section 15.3.2), our strategy for sampling time series is using a data augmentation approach based on Carter and Kohn (1994) and Frühwirth-Schnatter (1994). Frühwirth-Schnatter calls this approach *forward-filtering-backward-sampling* (FFBS) algorithm, also used in Schorfheide and Song (2015). This is one step of the Gibbs sampler, which takes sampled parameters and covariance matrix from a previous step of the Gibbs sampler, based on a Bayesian VAR with a modified Minnesota-type prior (see Doan et al., 1984, and Litterman, 1986, for the original Minnesota prior).

In its general form, the FFBS algorithm is straightforward to implement. An issue arises, however, in models with autoregressive (AR) components, or any other model in which consecutive state vectors contain common components. In this case, linear combinations of elements from the current state are actually observed, and the basic FFBS algorithm degenerates. From this point, we depart from Schorfheide and Song (2015) by proposing an algorithm intended to perform FFBS for a VAR case, avoiding degeneration issues. The algorithm is based on West's method on efficiently sampling

state vectors in AR models, adapted to our VAR case. We also depart from Schorfheide and Song (2015) by allowing for irregularities in the frequency of observations, generalizing the shape of the selection matrix introduced by the authors for regularly-spaced mixed-frequency time series.

Eraker et al. (2015) also work on unevenly-spaced mixed-frequency time-series, but do not benefit from using a state-space specification, which allows for additional linear restrictions on the relation between observed time series and latent variables. Normally, dealing with missing values in a state-space framework, like the one used in our Gibbs step for sampling missing observations conditional on the parameter set, is relatively easy (see, e.g., Prado and West, 2010, Section 4.3.3, and Durbin and Koopman, 2012, Section 4.10). The approach is flexible enough for not requiring the existence of at least one time series to be observed every period. Indeed, it works even when there are no observations at particular periods. Embedding this approach, we also depart from Schorfheide and Song (2015) and Eraker et al. (2015), for they require at least one completely observed time series. Therefore, our approach benefits from the best features of both Schorfheide and Song (2015) and Eraker et al. (2015) procedures.

The flexibility of not requiring the existence of at least one time series to be observed every period comes extremely in handy when performing forecasts and backcasts, as the algorithm only has to consider these exercises as an extended sample for which there is no observation at all. In other words, the FFBS algorithm samples missing values in forecast and backcast exercises. Therefore, our approach is efficient, intuitive, and well suited for dealing with all four common issues previously described: short length, low frequencies, discontinuity, randomly spaced missing observations, and multivariate time series with periods in which no observations are available.

The algorithm is also well suited for dealing with high-frequency real-time information, in order to improve nowcasts and forecasts of lower frequency time series. An empirical exercise in Section 5 shows the improvement in quarterly Brazilian GDP's nowcasts as real-time high-frequency information arrives, and compares the results to the Brazilian IBC-BR. The exercise also highlights the role of a

proper setup of the state-space representation of the model, in order to establish the relation between monthly and quarterly observations. In another exercise, we use the algorithm to recover the joint distribution of historical PNAD-C Brazilian unemployment rates prior to 2012, using information from other surveys such as annual PNAD and PME. The importance of the latter exercise is due to the fact that PNAD-C is a new monthly time series, whose values are available only from 2012 on.

The remainder of the study is organized as follows. The model is described in Section 2. The Gibbs sampler, and details on how to recover the full joint distribution of the missing observations are described in Section 3. Details on the efficient FFBS algorithm and Bayesian VAR are described in sections 3.1 and 3.2. We test the performance of our approach with simulated data sets in Section 4. Two empirical exercises are shown in Section 5, based on information of Brazilian GDP and unemployment. The empirical exercises explore both the issue of the regularity of missing information and the use of linear restrictions in the model. Section 6 summarizes the paper's conclusions.





# 2

## The Model



We assume the existence of a sample  $\mathbf{Y}^{ob} \equiv \{y_1^{ob}, \dots, y_T^{ob}\}$  of size  $T$ . At each period,  $y_t^{ob}$  is a  $(m_t^{ob} \times 1)$  vector of observed endogenous variables, whose dimension  $m_t^{ob}$  changes over time due to irregularly-spaced mixed-frequency. This definition of  $m_t^{ob}$  does not exclude even the possibility that, at certain periods, no observations are available, that is,  $m_t^{ob} = 0$ .

In this context, define first  $y_t$  as a  $(m \times 1)$  vector of endogenous variables of interest, whose dimension  $m \geq m_t^{ob}$  ( $m > m_t^{ob}$ , if  $m_t^{ob} = 0$ ) is time-invariant. Notice that it is possible that the set of endogenous variables of interest might not be directly observed at all in the sample. For instance,  $y_t^{ob}$  might contain moving averages of  $y_t$ , which are affine transformations of  $y_t$  and its lagged values. In this case,  $y_t^{ob}$  is not even a subset of  $y_t$ . The relation between  $y_t^{ob}$  and  $y_t$  is formally addressed in Section 3.1.

Our objective is then to infer the joint distribution of the  $(T \times m)$  sequence  $\mathbf{Y} \equiv [y_1, \dots, y_T]'$  of endogenous variables of interest. To fulfill this task, assume that the dynamics of  $y_t$  can be represented by a VAR $_m(q)$ :

$$\mathbf{1} \quad y_t = \Phi_{1y_{t-1}} + \dots + \Phi_q y_{t-q} + \bar{\Phi}_c \bar{x}_{ct} + e_t, \quad e_t \stackrel{iid}{\sim} N(0, \Sigma_e),$$

where  $\bar{x}_{ct}$  is a  $(m_c \times 1)$  matrix of deterministic variables, such as the constant, trend and seasonal dummies,  $\Phi_\ell$  is a  $(m \times m)$  matrix of coefficients for  $\ell \in \{1, \dots, q\}$ ,  $\bar{\Phi}_c$  is a  $(m \times m_c)$  matrix of coefficients,



and  $\Sigma_e$  is a  $(m \times m)$  positive definite covariance matrix. Each equation has  $k \equiv mq + m_c$  regressors, the VAR has  $mk$  coefficients, and the whole system has  $mk + m^2$  parameters.

$$y_t = [y_{1t} \dots y_{mt}]'$$

$$e_t = [e_{1t} \dots e_{mt}]'$$

$$\bar{x}_{ct} = [\bar{x}_{c;1t} \dots \bar{x}_{c;m_c t}]'$$

$$\Phi_\ell = \begin{pmatrix} \Phi_{\ell;11} & \dots & \Phi_{\ell;m1} \\ \vdots & & \vdots \\ \Phi_{\ell;1m} & \dots & \Phi_{\ell;mm} \end{pmatrix}$$

$$\bar{\Phi}_c = \begin{pmatrix} \bar{\Phi}_{c;11} & \dots & \bar{\Phi}_{c;m1} \\ \vdots & & \vdots \\ \bar{\Phi}_{c;1m_c} & \dots & \bar{\Phi}_{c;m m_c} \end{pmatrix}$$

For analytical convenience, each part of the Gibbs sampler, described in Section 3, considers different companion forms when dealing with the linear system. More details are provided in Sections 3.1 and 3.2.



# 3

# Gibbs Sampling



This section describes the steps of a Gibbs algorithm to draw a sample of size  $S$ , given the information set described by  $\mathbf{Y}^{ob}$ , from the joint distribution of  $(\mathbf{Y}, \Phi | \mathbf{Y}^{ob})$ . Each step of the sampler consists of two parts: 1) sample draws of  $\mathbf{Y}$  from the conditional distribution of  $(\mathbf{Y} | \mathbf{Y}^{ob}, \Phi)$ , using Kalman filtering and a modified form of the *forward-filtering-backward-sampling* (FFBS) algorithm, as described in Section 3.1; and 2) sample draws of  $\Phi$  from the conditional distribution of  $(\Phi | \mathbf{Y}^{ob}, \mathbf{Y})$ , using a Bayesian VAR, as described in Section 3.2.

The sampler is initialized, in part 1) of step 0, by setting  $\Phi_{init}$  to imply a random walk dynamics to  $\mathbf{Y}$ , that is,  $\Phi_\ell = \frac{1}{q} \mathbf{I}_{(m \times m)}$ , for  $\ell \in \{1, \dots, q\}$ , and  $\bar{\Phi}_\ell = 0$ . The FFBS algorithm of part 1) generates a smoothed path of  $\mathbf{Y}_0$  from  $(\mathbf{Y} | \mathbf{Y}^{ob}, \Phi_{init})$  to be used in part 2), where a Bayesian VAR generates the initial parameter set  $\Phi_0$  from  $(\Phi | \mathbf{Y}^{ob}, \mathbf{Y}_0)$ , still at step 0. From step  $s=1$  to  $s=S$ , the Gibbs algorithm evolves as usual: 1) draw  $\mathbf{Y}_s$  from  $(\mathbf{Y} | \mathbf{Y}^{ob}, \Phi_{s-1})$ , and 2) draw  $\Phi_s$  from  $(\Phi | \mathbf{Y}^{ob}, \mathbf{Y}_s)$ .

### 3.1 The Dynamic Linear Model

In what follows, we characterize the  $VAR_m(q)$  system 1) as a dynamic linear model (DLM), as in Schorfheide and Song (2015), and describe the sequential updating, that is, Kalman (1960) filtering, equations.<sup>2</sup> A companion form of the  $VAR_m(q)$  system is:

---

<sup>2</sup> Great references on dynamic linear models (DLM) and inference using the Kalman filter are Hamilton (1994), Prado and West (2010)

$$\begin{aligned}
 & \text{Observation} \\
 & \text{equation} \\
 y_t^{ob} &= F_t z_t; \quad F_t \equiv M_t \Lambda_z \\
 & \text{Transition} \\
 & \text{equation} \\
 z_t &= G_1 z_{t-1} + \omega_t; \quad \omega_t \sim N^{ind}(G_2 \bar{x}_{ct}, \Omega_e)
 \end{aligned}$$

where  $y_t^{ob}$  is a  $(m_t^{ob} \times 1)$  vector of observed endogenous variables, whose dimension  $m_t^{ob}$  changes due to mixed-frequency or irregularly-spaced observations,  $M_t$  and  $\Lambda_z$  are generalizations of what was done in Schorfheide and Song (2015), that is,  $M_t$  is a  $(m_t^{ob} \times m)$  selection matrix and  $\Lambda_z$  is a  $(m \times m\bar{q})$  transformation matrix,<sup>3</sup> useful when the observed variable is a known affine transformation of the states (for example, three-month moving averages or any other linear combinations),  $z_t \equiv [y_t', \dots, y_{t-\bar{q}+1}']'$  is a  $(m\bar{q} \times 1)$  vector of states,  $G_1$  is a  $(m\bar{q} \times m\bar{q})$  matrix of coefficients for endogenous variables,  $G_2$  is a  $(m\bar{q} \times m_c)$  matrix of coefficients for exogenous variables, and  $\Omega_e$  is a  $(m\bar{q} \times m\bar{q})$  positive semi-definite covariance matrix:

$$G_1 = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{\bar{q}-1} & \Phi_{\bar{q}} \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \quad G_2 = \begin{bmatrix} \bar{\Phi}_c \\ 0m \times m_c \\ 0m \times m_c \\ \vdots \\ 0m \times m_c \end{bmatrix} \quad \Omega_e = \begin{bmatrix} \Sigma_e & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Note that  $\bar{q}$  does not necessarily match  $q$ , the number of lags in the  $VAR_m(q)$  describing the dynamics of  $y_t$ . For instance, suppose

---

and West and Harrison (1997). For DLMS applications in macroeconomic models, see Basdevant (2003).

<sup>3</sup> In this regard,  $\Lambda_z$  should be designed with care in order to avoid *aliasing* issues when we only observe averages, or any other linear combination, of latent variables of interest. For instance, suppose that we observe three quarterly time series  $y_t = [y_{1t}, y_{2t}, y_{3t}]'$ , which actually aggregates or averages latent monthly variables during each quarter, such as sectoral GDP's. In this case, there are an infinite set of possible monthly sectoral GDP time series that are consistent with the observed variables. Those sets include monthly time series with unreasonable extremely large variances. In this case, the algorithm will not converge to a unique stationary distribution

that  $y_t^{ob}$  represents quarterly averages ( $\bar{q} = 3$ ) of monthly nonobserved values of  $y_t$ , whose dynamics is described by a VAR with four lags ( $q = 4$ ). Conversely, if the dynamics are better described by a  $VAR_m(1)$ , a similar issue arises. Those cases are easily addressed as follows: 1) if  $\bar{q} < q$ , set  $\bar{q} = q$  and extend  $\Lambda_z$  with zero matrices, that is,  $\Lambda_z = [\Lambda_z, 0_{m \times m(q-\bar{q})}]$ ; 2) if  $\bar{q} > q$ , consider additional zero matrices  $\Phi_{\bar{\ell}} = 0_{m \times m}$ , for  $\bar{\ell} \in \{q+1, \dots, \bar{q}\}$ .

In this context,  $M_t$  is a time-varying selection matrix where the number of rows is adjusted to match the number of absent observations in each period. From this perspective, starting from an identity matrix  $M$  with size  $(m \times m)$ ,  $M_t$  is set in each period by eliminating  $m - m_t^{ob}$  rows of  $M$ , corresponding to nonobserved variables of vector  $y_t^{ob}$ .

Let  $\mathcal{D}_t \equiv \{\mathcal{D}_{t-1}, y_t^{ob}\}$  denote the information set at each period. If there are no observed values at the period  $t$ , that is,  $y_t^{ob}$  is an empty vector, then  $M_t$  is an empty matrix and  $\mathcal{D}_t = \mathcal{D}_{t-1}$ .<sup>4</sup>

Consider the following initial prior Gaussian density for  $z_0$  is  $(z_0 | \mathcal{D}_0) \sim N(\mu_0, C_0)$ . Therefore, the Kalman filtering (KF) equations are:

a) Prior density for  $z_t$  with information up to period  $(t-1)$ :

$$(z_t | \mathcal{D}_{t-1}) \sim N(a_t, R_t)$$

$$a_t \equiv G_1 \mu_{t-1} + G_2 \bar{x}_{ct}; R_t \equiv G_1 C_{t-1} G_1' + \Omega_e$$

b) One-step forecast of  $y_t^{ob}$  with information up to period  $(t-1)$ :

$$(y_t^{ob} | \mathcal{D}_{t-1}) \sim N(f_t, Q_t)$$

$$f_t \equiv F_t a_t; Q_t \equiv F_t R_t F_t'$$

c) Posterior density for  $z_t$  with information up to period  $t$ :

---

<sup>4</sup> A better discussion on models with missing and unequally spaced data can be found in, for example, Prado and West (2010, Section 4.3.3) and Durbin and Koopman (2012, Section 4.10).

$$(z_t | \mathcal{D}_t) \sim N(\mu_t, C_t).$$

If there are observations at period  $t$ :

$$\mu_t \equiv a_t + A_t v_t; \quad C_t \equiv R_t - A_t Q_t A_t'; \quad v_t \equiv y_t^{ob} - f_t; \quad A_t \equiv R_t F_t' Q_t^{-1}.$$

Otherwise:

$$\mathcal{D}_t = \mathcal{D}_{t-1}; \quad (z_t | \mathcal{D}_t) \equiv (z_t | \mathcal{D}_{t-1}); \quad \mu_t \equiv a_t; \quad C_t \equiv R_t.$$

Our strategy for completing the series with mixed-frequency or irregularly-spaced observations is to use a data augmentation approach based on Carter and Kohn (1994) and Frühwirth-Schnatter (1994) to sample complete sequences of state variables. We also follow Frühwirth-Schnatter (1994) by calling this approach as *forward-filtering-backward-sampling* (FFBS) algorithm. Let  $Z_T \equiv [z_T, z_{T-1}, \dots, z_1]$  denote the whole sequence of state vectors. The nature of the FFBS algorithm comes from exploring the Markov structure of the evolution equation of any DLM, which allows us to write

$$\Pr(Z_T | \mathcal{D}_T) = \Pr(z_1 | z_2, \mathcal{D}_1) \dots \Pr(z_t | z_{t+1}, \mathcal{D}_t) \dots \Pr(z_{T-1} | z_T, \mathcal{D}_{T-1}) \Pr(z_T | \mathcal{D}_T).$$

The proof is shown in the Annex.

Therefore, the FFBS algorithm consists of sampling  $z_T$  from  $\Pr(z_T | \mathcal{D}_T)$ , and sequentially sampling  $z_t$  from  $\Pr(z_t | z_{t+1}, \mathcal{D}_t)$  using equations directly derived from the Kalman filter.

In a paper, similar to ours, Schorfheide and Song (2015) also use a data augmentation approach based on Carter and Kohn (1994). We depart from Schorfheide and Song (2015) by allowing for irregularities in the frequency of observations, instead of simple, regularly-spaced mixed-frequency time-series. This extension, also carried out in Eraker et al. (2015), is obtained by generalizing the shape of  $M_t$  matrix. Eraker et al. (2015), on the other hand, do not benefit from using a state-space specification, which allows for additional linear restrictions on the relation between observables and states, captured by matrix  $\Lambda_z$ . Therefore, our approach benefits from the best features of both procedures.





Adapting West's modification of the FFBS algorithm for sampling  $z_t$  in our case is straightforward. First, let  $\tilde{z}_t \equiv [y'_t, \dots, y'_{t-\bar{q}+2}]'$  denote the already sampled  $(m(\bar{q}-1) \times 1)$  vector of common elements of  $z_{t+1}$  and  $z_t$ , that is, we partition both vectors as  $z_{t+1} = [y'_{t+1}, \tilde{z}'_t]'$  and  $z_t = [\tilde{z}'_t, y'_{t-\bar{q}+1}]'$ . In this context, let  $\tilde{z}_{j,t}$  denote each singular element of  $\tilde{z}_t$  for  $j \in \{1, \dots, m(\bar{q}-1)\}$ .

For analytical simplicity, consider the subvector  $\delta_{j,t} \equiv [\tilde{z}'_{j:m(\bar{q}-1),t}, y'_{t-\bar{q}+1}]'$  where  $\tilde{z}_{j_1:j_2,t} \equiv [\tilde{z}_{j_1,t}, \dots, \tilde{z}_{j_2,t}]'$  for  $j_2 < j_1$ ,<sup>6</sup> and  $\tilde{z}_{j_1:j_2,t} \equiv \text{empty}$  for  $j_2 < j_1$  and the following partitions, for  $j \leq i$ :

$$\delta_{j,t} = \begin{bmatrix} \tilde{z}_{j,t} \\ \delta_{j+1,t} \end{bmatrix}, \quad \mu_{j,j,t} = \begin{bmatrix} \tilde{\mu}_{j,t} \\ \mu_{j,j+1,t} \end{bmatrix}, \quad C_{j,j,t} = \begin{bmatrix} \tilde{\sigma}_{j,t}^2 & \Sigma'_{j,j+1,t} \\ \Sigma_{j,j+1,t} & C_{j,j+1,t} \end{bmatrix},$$

where  $\mu_{j,i,t} \equiv E(\delta_{i,t} | \tilde{z}_{1:(j-1),t}, \mathcal{D}_t)$ ,  $C_{j,i,t} \equiv \text{Var}(\delta_{i,t} | \tilde{z}_{1:(j-1),t}, \mathcal{D}_t)$ ,  $\tilde{\mu}_{j,t} \equiv E(\tilde{z}_{j,t} | \tilde{z}_{1:(j-1),t}, \mathcal{D}_t)$ ,  $\tilde{\sigma}_{j,t}^2 \equiv \text{Var}(\tilde{z}_{j,t} | \tilde{z}_{1:(j-1),t}, \mathcal{D}_t)$  and  $\Sigma_{j,i,t} \equiv \text{Cov}(\delta_{i,t}, \tilde{z}_{j,t} | \tilde{z}_{1:(j-1),t}, \mathcal{D}_t)$ . The partitions imply that  $\tilde{z}_{j,t}, \tilde{\mu}_{j,t}$  and  $\tilde{\sigma}_{j,t}^2$  are scalars.

The following steps describe the sampling strategy:

- 1) Sample a value of  $z_T$  from  $(z_T | \mathcal{D}_T) \sim N(\mu_T, C_T)$ ;
- 2) For each  $t \in \{(T-1), (T-2), \dots, 1\}$ , sequentially sample values from the conditional distribution of  $(y_{t-\bar{q}+1} | z_{t+1}, \mathcal{D}_t)$  and complete vector  $z_t$  using the relevant elements from the vector  $z_{t+1}$ . This is achieved by obtaining the distributions of  $(y_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t)$  and  $(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$ . Also note, as shown in the Annex, that  $\Pr(y_{t-\bar{q}+1} | z_{t+1}, \mathcal{D}_t) \propto \Pr(y_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t) \Pr(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$ . Since  $z_t = [\tilde{z}'_t, y'_{t-\bar{q}+1}]'$ , the first choice to retrieve  $\Pr(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$  would

<sup>6</sup> Note that  $\delta_{1,t} = z_t$  and  $\delta_{(m(\bar{q}-1)+1),t} = y_{t-\bar{q}+1}$ .

be using the KF distribution  $(z_t | \mathcal{D}_t) \sim N(\mu_t, C_t)$  to directly obtain the conditional distribution of  $(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$ . This method, however, involves inverting  $Var(\tilde{z}_t | \mathcal{D}_t)$  which is often singular.<sup>7</sup> Therefore, based on West's method, we efficiently obtain the conditional distribution  $(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$  by sequentially conditioning on each scalar element  $\tilde{z}_{j,t}$  of  $\tilde{z}_t$ , reducing the dimension of the distribution by one at each stage. The steps are described below.

a) Since  $\hat{z}_{1,t} = z_t$ , note that  $\mu_{1,1,t} = \mu_t$  and  $C_{1,1,t} = C_t$ .

b) For  $j \in \{1, \dots, m(\bar{q}-1)\}$ , compute the distribution of  $(\hat{z}_{(j+1),t} | \tilde{z}_{1:j,t}, \mathcal{D}_t)$ :<sup>8</sup>

$$\begin{aligned} & \left( \hat{z}_{(j+1),t} | \tilde{z}_{1:j,t}, \mathcal{D}_t \right) \sim N \left( \mu_{(j+1),(j+1),t}, C_{(j+1),(j+1),t} \right) \\ \mu_{(j+1),(j+1),t} &= \begin{cases} \mu_{j,j+1,t} + \frac{1}{\sigma_{j,t}^2} \Sigma_{j,j+1,t} (\tilde{z}_{j,t} - \tilde{\mu}_{j,t}), & \text{if } \tilde{\sigma}_{j,t}^2 > 0 \\ \mu_{j,j+1,t}, & \text{if } \tilde{\sigma}_{j,t}^2 = 0 \end{cases}, \\ C_{(j+1),(j+1),t} &= \begin{cases} C_{j,j+1,t} - \frac{1}{\tilde{\sigma}_{j,t}^2} \Sigma_{j,j+1,t} \Sigma'_{j,j+1,t}, & \text{if } \tilde{\sigma}_{j,t}^2 > 0 \\ C_{j,j+1,t} & \text{if } \tilde{\sigma}_{j,t}^2 = 0 \end{cases}. \end{aligned}$$

<sup>7</sup> The same reasons why  $R_t$  is eventually singular might cause  $Var(\tilde{z}_t | \mathcal{D}_t)$  to be singular as well.

<sup>8</sup> If  $\tilde{\sigma}_{j,t}^2 = 0$ , then  $\Sigma_{j,j+1,t} = 0$  and  $\tilde{z}_{j,t} = \tilde{\mu}_{j,t}$ . In this case, the limiting value of  $\frac{1}{\tilde{\sigma}_{j,t}^2} \Sigma_{j,j+1,t} (\tilde{z}_{j,t} - \tilde{\mu}_{j,t})$  and  $\frac{1}{\tilde{\sigma}_{j,t}^2} \Sigma_{j,j+1,t} \Sigma'_{j,j+1,t}$  are both zero. For coding purposes, there still might be some numerical instability in cases where  $\tilde{\sigma}_{j,t}^2$  is positive, but very small. In this case, the restriction  $\tilde{\sigma}_{j,t}^2 > 0$  might be translated into something as, for instance,  $\tilde{\sigma}_{j,t}^2 > Toler \cdot \tilde{\mu}_{j,t}$ , where *Toler* is a relative tolerance parameter.

c) Since  $\mathfrak{z}_{(m(\bar{q}-1)+1),t} = y_{t-\bar{q}+1}, (y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$  is distributed as  $(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t) \sim N(\tilde{\mu}_{t-\bar{q}+1}, \tilde{C}_{t-\bar{q}+1})$ , where

$$\begin{aligned}\tilde{\mu}_{t-\bar{q}+1} &= \mu_{(m(\bar{q}-1)+1), (m(\bar{q}-1)+1), t}, \\ \tilde{C}_{t-\bar{q}+1} &= C_{(m(\bar{q}-1)+1), (m(\bar{q}-1)+1), t}.\end{aligned}$$

3) Based on the already sampled vector  $\tilde{z}_{t+1}$  and exogenous vector  $\bar{x}_{ct+1}$ , compute the partial residual

$$\tilde{e}_{t+1} \equiv y_{t+1} - (\Phi_1 y_t + \dots + \Phi_{\bar{q}-1} y_{t-\bar{q}+2} + \bar{\Phi}_c \bar{x}_{ct+1}).$$

Note that  $\Pr(\tilde{e}_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t) = \Pr(y_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t)$ , whose conditional distribution is  $(y_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t) \sim N(\Phi_{\bar{q}} y_{t-\bar{q}+1}, \Sigma_e)$ .

4) Note that  $\Pr(y_{t-\bar{q}+1} | z_{t+1}, \mathcal{D}_t) \propto \Pr(y_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t) \Pr(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$ , as the Annex shows. If  $\tilde{C}_{t-\bar{q}+1}$  is not singular, all we have to do is to sample  $y_{t-q+1}$  from

$$4 \quad (y_{t-\bar{q}+1} | z_{t+1}, \mathcal{D}_t) \sim N(\mu_{t-q+1}^y, C_{t-q+1}^y)$$

where

$$\begin{aligned}\mu_{t-q+1}^y &\equiv \left( \Phi_{\bar{q}}' \Sigma_e^{-1} \Phi_{\bar{q}} + (\tilde{C}_{t-\bar{q}+1})^{-1} \right)^{-1} \left( \Phi_{\bar{q}}' \Sigma_e^{-1} \tilde{e}_{t+1} + (\tilde{C}_{t-\bar{q}+1})^{-1} \tilde{\mu}_{t-\bar{q}+1} \right), \\ C_{t-q+1}^y &\equiv \left( \Phi_{\bar{q}}' \Sigma_e^{-1} \Phi_{\bar{q}} + (\tilde{C}_{t-\bar{q}+1})^{-1} \right)^{-1}.\end{aligned}$$

Up to this point, we closely followed West's suggestions. There might be; however, an issue when sampling  $(y_{t-\bar{q}+1} | z_{t+1}, \mathcal{D}_t)$  based on the last result, that is, there might be cases in which  $\tilde{C}_{t-\bar{q}+1}$  is singular. In this case, we propose a procedure similar to what was done in 2.b. Our suggestion is to sequentially sample each scalar element  $y_{j,t-q+1}$  of  $y_{t-q+1}$ , for  $j \in \{1, \dots, m\}$ , reducing the dimension of the

distribution  $(y_{t-\bar{q}+1} | \tilde{z}_t, \mathcal{D}_t)$  by one at each stage. For that, consider the sub-vectors  $\check{y}_{j,t} \equiv y_{j:m,t-q+1}$  and  $\hat{y}_{j,t} \equiv y_{1:j,t-q+1}$  the conditional distribution  $\Pr(\check{y}_{j,t} | \hat{y}_{(j-1)}, z_{t+1}, \mathcal{D}_t) \propto \Pr(\tilde{e}_{t+1} | \tilde{z}_t, y_{t-\bar{q}+1}, \mathcal{D}_t) \Pr(\check{y}_{j,t} | \hat{y}_{(j-1)}, \tilde{z}_t, \mathcal{D}_t)$ , where  $\tilde{y}_{j,t} \equiv y_{j,t-q+1}$  is a simplifying notation, and the following partitions

$$\check{y}_{j,t} = \begin{bmatrix} y_{j,t-q+1} \\ \check{y}_{(j+1),t} \end{bmatrix}, \quad \check{\mu}_{j,j,t} = \begin{bmatrix} \mu_{j,t-q+1} \\ \check{\mu}_{j,j+1,t} \end{bmatrix}, \quad \check{C}_{j,j,t} = \begin{bmatrix} \check{\sigma}_{j,t}^2 & \check{\Sigma}'_{j,j+1,t} \\ \check{\Sigma}_{j,j+1,t} & \check{C}_{j,j+1,t} \end{bmatrix},$$

where the means are

$$\check{\mu}_{j,i,t} \equiv E(\check{y}_{i,t} | y_{(j-1)}, \tilde{z}_t, \mathcal{D}_t) \quad \text{and} \quad \mu_{j,t-q+1} \equiv E(\tilde{y}_{j,t} | \hat{y}_{(j-1)}, \tilde{z}_t, \mathcal{D}_t),$$

while the variances are

$$\check{\sigma}_{j,t}^2 \equiv \text{Var}(\check{y}_{j,t} | \hat{y}_{(j-1)}, \tilde{z}_t, \mathcal{D}_t), \quad \check{C}_{j,i,t} \equiv \text{Var}(\check{y}_{i,t} | \hat{y}_{(j-1)}, \tilde{z}_t, \mathcal{D}_t)$$

$$\text{and } \check{\Sigma}_{j,i,t} \equiv \text{Cov}(\check{y}_{i,t}, \check{y}_{j,t} | y_{(j-1)}, \tilde{z}_t, \mathcal{D}_t).$$

The partitions imply that  $y_{j,t-q+1}$ ,  $\mu_{j,t-q+1}$  and  $\check{\sigma}_{j,t}^2$  are scalars. Therefore, take the steps described below:

- a) Since  $\check{y}_{1,t} = y_{t-q+1}$ , note that  $\check{\mu}_{1,1,t} = \tilde{\mu}_{t-\bar{q}+1}$  and  $\check{C}_{1,1,t} = \tilde{C}_{t-\bar{q}+1}$ .
- b) For each  $j \in \{1, \dots, m\}$ , consider the  $j$ th ( $m \times 1$ ) vector  $\Phi_{j,\bar{q}}$  in the partition  $\Phi_{\bar{q}} \equiv [\Phi_{1,\bar{q}}, \dots, \Phi_{j,\bar{q}}, \dots, \Phi_{m,\bar{q}}]$ , and sample  $y_{j,t-\bar{q}+1}$  from:<sup>9</sup>

$$(y_{j,t-\bar{q}+1} | \hat{y}_{(j-1)}, z_{t+1}, \mathcal{D}_t) \sim N\left(\check{\mu}_{j,t-q+1}^y, (\check{\sigma}_{j,t}^y)^2\right)$$

<sup>9</sup> For coding purposes, there still might be some numerical instability in cases where  $\check{\sigma}_{j,t}^2$  is positive, but very small. In this case, the restriction  $\check{\sigma}_{j,t}^2 > 0$  might be translated into something as, for instance,  $\check{\sigma}_{j,t}^2 > \text{Toler} \cdot \mu_{j,t-q+1}$ , where *Toler* is a relative tolerance parameter.

$$\check{\mu}_{j,t-q+1}^y = \begin{cases} \left( \frac{\Phi'_{j,\bar{q}} \Sigma_e^{-1} \check{\ell}_{t+1} + (\check{\sigma}_{j,t}^2)^{-1} \mu_{j,t-q+1}}{\Phi'_{j,\bar{q}} \Sigma_e^{-1} \Phi_{j,\bar{q}} + (\check{\sigma}_{j,t}^2)^{-1}} \right), & \text{if } \check{\sigma}_{j,t}^2 > 0 \\ \mu_{j,t-q+1}, & \text{if } \check{\sigma}_{j,t}^2 = 0 \end{cases}$$

$$(\check{\sigma}_{j,t}^y)^2 = \begin{cases} \frac{1}{\Phi'_{j,\bar{q}} \Sigma_e^{-1} \Phi_{j,\bar{q}} + (\check{\sigma}_{j,t}^2)^{-1}}, & \text{if } \check{\sigma}_{j,t}^2 > 0 \\ 0, & \text{if } \check{\sigma}_{j,t}^2 = 0 \end{cases}$$

where

$$\check{\mu}_{(j+1),(j+1),t} = \begin{cases} \check{\mu}_{j,j+1,t} + \frac{1}{\check{\sigma}_{j,t}^2} \check{\Sigma}_{j,j+1,t} (y_{j,t-\bar{q}+1} + \mu_{j,t-q+1}), & \text{if } \check{\sigma}_{j,t}^2 > 0 \\ \check{\mu}_{j,j+1,t}, & \text{if } \check{\sigma}_{j,t}^2 = 0 \end{cases},$$

$$\check{C}_{(j+1),(j+1),t} = \begin{cases} \check{C}_{j,j+1,t} - \frac{1}{\check{\sigma}_{j,t}^2} \check{\Sigma}_{j,j+1,t} \check{\Sigma}'_{j,j+1,t} & \text{if } \check{\sigma}_{j,t}^2 > 0 \\ \check{C}_{j,j+1,t} & \text{if } \check{\sigma}_{j,t}^2 = 0 \end{cases}.$$

5) Sampled vector  $z_t$  is obtained by concatenating

$$z_t = \left[ \check{z}'_t, y'_{t-\bar{q}+1} \right]'$$

*Smoothing.* As for the smoothed distribution, retrieving it is also eventually subject to inverting the covariance matrix  $R_t$  when using the standard method. However, we propose modifying the FFBS algorithm to obtain the distribution, avoiding the issue. In the spirit of the algorithm, all the way from  $t=(T-1)$  to  $t=1$ , we keep the means and variances, instead of sampling the states, and adjust vectors and matrices accordingly. By the end, we have retrieved the smoothed distribution. Note that the procedure is not the same as keeping the means and variances at  $t = \tau$  when performing the FFBS algorithm for sampling the states at  $t = \tau + 1$ , for the auxiliary matrices and vectors will not be the same.

*Forecasts, backcasts, and nowcasts.* When performing forecasts and backcasts, all the algorithm has to do is to consider the extended sample for which there is no observation at all. Therefore, the FFBS algorithm will also sample missing values in forecast and backcast exercises. The only requirement is that the VAR step, described in Section 3.2, must only consider the actually available sample period when sampling parameters conditional on observables and missing values.

No additional care is needed in nowcast exercises, for the definition of nowcasts requires at least one observed value to be available. Therefore, nowcasts are carried out on the actually available sample period.

### 3.2 Bayesian VAR

Following Canova (2007, chap. 10), the linear system can be represented in two companion forms. The first one is defined as follows:

$$\mathbf{Y} = \mathbf{X}\Phi + \mathcal{E},$$

where  $\mathbf{Y} \equiv [y_1, \dots, y_T]'$  is a  $(T \times m)$  matrix of variables  $\Phi \equiv [\Phi_1, \dots, \Phi_q, \bar{\Phi}_c]'$  denotes a  $(k \times m)$  matrix of coefficients,  $\mathbf{X} \equiv [x_1, \dots, x_T]'$  is a  $(T \times k)$  matrix of regressors,  $x_t \equiv [y'_{t-1}, \dots, y'_{t-q}, \bar{x}'_{ct}]'$  is a  $(k \times 1)$  vector of regressors, and  $\mathcal{E} \equiv [e_1, \dots, e_T]'$  is a  $(T \times m)$  matrix of error terms.<sup>10</sup> Note that each column of  $\mathbf{Y}$ ,  $\Phi$ , and  $\mathcal{E}$  corresponds to equations describing the dynamics of a unique variable sequence  $\{y_{it}\}$ , for  $i \in \{1, \dots, m\}$ :

$$\mathbf{x} = \begin{bmatrix} y_{1,1-1} & \cdots & y_{m,1-1} & \cdots & y_{1,1-q} & \cdots & y_{m,1-q} & \bar{x}_{c,11} & \cdots & \bar{x}_{c,m_c1} \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ y_{1,t-1} & \cdots & y_{m,t-1} & \cdots & y_{1,t-q} & \cdots & y_{m,t-q} & \bar{x}_{c,1t} & \cdots & \bar{x}_{c,m_c t} \\ \vdots & & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ y_{1,T-1} & \cdots & y_{m,T-1} & \cdots & y_{1,T-q} & \cdots & y_{m,T-q} & \bar{x}_{c,1T} & \cdots & \bar{x}_{c,m_c T} \end{bmatrix},$$

<sup>10</sup> Note that the VAR equation for a period  $t$  is  $y_t = \Phi' x_t + e_t$ .

$$\mathbf{Y} = \begin{bmatrix} y_{11} & \vdots & y_{i,1} & \cdots & y_{m,1} \\ \vdots & & \vdots & & \vdots \\ y_{1T} & \vdots & y_{i,T} & \cdots & y_{m,T} \end{bmatrix} \quad \mathbf{E} = \begin{bmatrix} e_{11} & \vdots & e_{i1} & \cdots & e_{m1} \\ \vdots & & \vdots & & \vdots \\ e_{1T} & \vdots & e_{iT} & \cdots & e_{mT} \end{bmatrix} \quad \Phi = \begin{array}{c|c|c} \Phi_{1;11} & \cdots & \Phi_{1;1} & \cdots & \Phi_{1;m1} \\ \vdots & & \vdots & & \vdots \\ \Phi_{1;1m} & \cdots & \Phi_{1;im} & \cdots & \Phi_{1:mm} \\ \vdots & & \vdots & & \vdots \\ \Phi_{q;11} & \cdots & \Phi_{q;1} & \cdots & \Phi_{q;m1} \\ \vdots & & \vdots & & \vdots \\ \Phi_{q;1m} & \cdots & \Phi_{q;im} & \cdots & \Phi_{q:mm} \\ \hline \bar{\Phi}_{c;11} & \cdots & \bar{\Phi}_{c;1} & \cdots & \bar{\Phi}_{c;m1} \\ \vdots & & \vdots & & \vdots \\ \bar{\Phi}_{c;1m_c} & \cdots & \bar{\Phi}_{c;im_c} & \cdots & \bar{\Phi}_{c:mm_c} \end{array}$$

Applying the  $vec(\cdot)$  operator to both sides of Equation 6, we obtain the second companion form:<sup>11</sup>

**7**  $\mathbf{y} = \mathbf{x}\phi + \varepsilon,$

where  $\mathbf{y} \equiv vec(\mathbf{Y})$  is a  $(Tm \times 1)$  vector of endogenous variables,  $\mathbf{x} \equiv (I_m \otimes \mathbf{X})$  is a  $(Tm \times km)$  vector of regressors,  $\phi \equiv vec(\Phi)$  is a  $(km \times 1)$  vector of coefficients, and  $\varepsilon \equiv vec(\mathbf{E})$  is a  $(Tm \times 1)$  matrix of error terms.

Given the model and a proper mapping between the set of observable variables ( $\mathbf{Y}^{ob}$ ) and endogenous variables ( $\mathbf{Y}$ ) provided above, the next sections describe the procedure to obtain the joint distribution of  $(\mathbf{Y}, \Phi | \mathbf{Y}^{ob})$  by means of a Monte Carlo simulation.

### 3.2.1 Computing the Likelihood Function

Assuming that  $\varepsilon \sim N_{Tm}(0, \Sigma_\varepsilon)$ , for  $\Sigma_\varepsilon \equiv (\Sigma_e \otimes I_T)$ , the likelihood function is

$$p(\mathbf{y} | \phi, \Sigma_\varepsilon) = |2\pi\Sigma_\varepsilon|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{x}\phi)' \Sigma_\varepsilon^{-1} (\mathbf{y} - \mathbf{x}\phi)\right].$$

<sup>11</sup> See the Annex for some results on matrix algebra and the  $vec(\cdot)$  operator.

Let  $\phi_{\text{ml}}$  and  $\Sigma_{\text{ml}}$  denote the maximum likelihood estimator of  $\phi$  and its covariance matrix:

$$\phi_{\text{ml}} \equiv \left( I_m \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \mathbf{y}; \quad \Sigma_{\text{ml}} \equiv \left( \Sigma_e \otimes (\mathbf{X}'\mathbf{X})^{-1} \right).$$

Using the definitions of  $\mathbf{x}$  and  $\Sigma_e$ , note that the likelihood function can be written<sup>12</sup> as follows:

$$\begin{aligned} \mathbf{8} \quad p(\mathbf{y}|\phi, \Sigma_e) &= |2\pi|^{-\frac{Tm}{2}} |\Sigma_e|^{-\frac{(T-k-m-1)+m+1}{2}} \exp\left(-\frac{1}{2} \text{tr}(\bar{\Lambda}_s \Sigma_e^{-1})\right) \\ &\quad \times |\Sigma_e|^{-\frac{k}{2}} \exp\left(-\frac{1}{2} (\phi - \phi_{\text{ml}})' \Sigma_{\text{ml}}^{-1} (\phi - \phi_{\text{ml}})\right), \end{aligned}$$

where

$$\mathbf{9} \quad \Phi_{\text{ml}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}; \quad \bar{\Lambda}_s \equiv (\mathbf{Y} - \mathbf{X}\Phi_{\text{ml}})' (\mathbf{Y} - \mathbf{X}\Phi_{\text{ml}})$$

### 3.2.2 Joint Prior and Posterior

Given the form of the likelihood function, we consider that a natural conjugate joint prior distribution for  $(\phi, \Sigma_e)$  is a generalization of what Gelman et al. (2003) call the normal-inverse-Wishart distribution. Thus, the joint prior distribution is specified according to the following hierarchical structure:

$$\Sigma_e \sim W_m^{-1}(\tilde{\delta}_0, \Lambda_0^{-1}); \quad (\phi|\Sigma_e) \sim N_{km}(\phi_0, \Sigma_0); \quad \Sigma_0 \equiv \frac{1}{\sigma_0} (\Sigma_e \otimes \tilde{I}_k),$$

where  $\tilde{\delta}_0 > (m+1)$  is the degrees of freedom of the inverse-Wishart distribution (which can be understood as the prior sample size for  $\Sigma_e$ ),  $E\Sigma_e = \Lambda_0 / (\tilde{\delta}_0 - m - 1)$  is the prior estimate for  $\Sigma_e, \sigma_0$ , is a scale parameter for the conditional normal distribution (which can be understood as the prior sample size for  $\phi$ ),  $\phi_0$  is the prior estimate for  $\phi$ , and  $\tilde{I}_k$  is a  $(k \times k)$  diagonal matrix whose diagonal terms satisfy  $\tilde{I}_{\mathfrak{R}\mathfrak{R}} \in (0, 1]$ , and  $\mathfrak{R} \in \{1, \dots, k\}$ .

<sup>12</sup> See the Annex for more details.



Note that the identity matrix  $I_k$  is a particular case of  $\tilde{I}_k$ . Moreover, this definition allows for using a Minnesota-type prior (see, for example, Doan et al., 1984, and Litterman, 1986, for the original Minnesota prior).

It implies that the joint prior probability density function (PDF)  $p(\phi, \Sigma_e) = p(\Sigma_e)p(\phi|\Sigma_e)$  is proportional to:

$$\begin{aligned} \mathbf{10} \quad p(\phi, \Sigma_e) &\propto |\Sigma_e|^{-\frac{(\delta_0+m+1)}{2}} \exp\left(-\frac{1}{2}tr(\Lambda_0 \Sigma_e^{-1})\right) \\ &\times |2\pi\Sigma_0|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\phi - \phi_0)' \Sigma_0^{-1} (\phi - \phi_0)\right]. \end{aligned}$$

Therefore, the joint posterior PDF  $p(\phi, \Sigma_e | \mathbf{y}) \propto p(\phi, \Sigma_e)p(\mathbf{y}|\phi, \Sigma_e)$  is proportional to:<sup>13</sup>

$$\begin{aligned} \mathbf{11} \quad p(\phi, \Sigma_e | \mathbf{y}) &\propto |\Sigma_e|^{-\frac{(\delta_0+T+m+1)}{2}} \exp\left\{-\frac{1}{2}tr\left[(\Lambda_0 + \bar{\Lambda}_s + \Lambda_T)\Sigma_e^{-1}\right]\right\} \\ &\times |2\pi\Sigma_T|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\phi - \phi_T)' \Sigma_T^{-1} (\phi - \phi_T)\right], \end{aligned}$$

where

$$\begin{aligned} \Sigma_T^{-1} &\equiv (\Sigma_0^{-1} + \Sigma_{ml}^{-1}) = (\Sigma_e^{-1} \otimes (\sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X})) \\ \phi_T &\equiv \Sigma_T (\Sigma_0^{-1} \phi_0 + \Sigma_{ml}^{-1} \phi_{ml}) \\ \Phi_T &\equiv vec^{-1}(\phi_T) \\ &= (\sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X})^{-1} (\sigma_0 \tilde{I}_k^{-1}) \Phi_0 + (\sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X}) \Phi_{ml} \\ \Lambda_T &\equiv \Phi_0' (\sigma_0 \tilde{I}_k^{-1}) \Phi_0 + \Phi_{ml}' (\mathbf{X}'\mathbf{X}) \Phi_{ml} - \Phi_T' (\sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X}) \Phi_T. \end{aligned}$$

This means that the joint posterior distribution of  $(\phi, \Sigma_e | \mathbf{y})$  is specified according to the following hierarchical structure:

---

<sup>13</sup> Details are shown in the Annex.

$$(\Sigma_e | \mathbf{y}) \sim W_m^{-1} \left( \delta_0 + T, (\Lambda_0 + \bar{\Lambda}_s + \Lambda_T)^{-1} \right); \quad (\phi | \Sigma_e, \mathbf{y}) \sim N_{km}(\phi_T, \Sigma_T).$$

### 3.2.3 Specifying the Joint Prior Distribution

As for the joint prior distribution, we propose the following variation of the Minnesota prior:

$$\begin{aligned} \mathbf{12} \quad \Phi_{0\ell;ij} &= \begin{cases} 1 & \text{if } \ell=1 \text{ and } i=j \\ 0 & \text{otherwise} \end{cases}, \\ \bar{\Phi}_{0c} &= 0_{m \times m_c} \\ \tilde{I}_{\tilde{\mathfrak{R}}\tilde{\mathfrak{R}}} &= \begin{cases} \tilde{\mathfrak{R}}^{-\alpha} \text{ (harmonic decay) or} & \text{if } 1 \leq \tilde{\mathfrak{R}} \leq (k - m_c), \\ \alpha^{1-\tilde{\mathfrak{R}}} \text{ (geometric decay)} & \\ 1 & \text{otherwise} \end{cases}, \end{aligned}$$

where  $\tilde{\mathfrak{R}} \equiv \text{int} \left( \frac{\mathfrak{R}-1}{m} \right) + 1$ ,  $\alpha > 0$ ,  $\mathfrak{R} \in \{1, \dots, k\}$ ,  $\ell \in \{1, \dots, q\}$ ,  $i \in \{1, \dots, m\}$ ,  $j \in \{1, \dots, m\}$ , and  $\bar{m}_c \in \{1, \dots, m_c\}$ , and again  $k \equiv mq + m_c$ .

For consistency, we assume that  $E\Sigma_e = \bar{\Sigma}_e$ , where  $E\Sigma_e = \frac{\Lambda_0}{(\delta_0 - m - 1)}$  and  $\bar{\Sigma}_e \equiv \frac{\bar{\Lambda}_s}{(T-1)}$ . This assumption implies that

$$\mathbf{13} \quad \Lambda_0 = \frac{(\delta_0 - m - 1)}{(T-1)} \bar{\Lambda}_s.$$

A diffuse prior, if chosen, requires low levels for  $\delta_0 > (m+1)$  and  $\sigma_0 > 0$ . In the limiting case  $\delta_0 \rightarrow (m+1)$  and  $\sigma_0 \rightarrow 0$  the posterior distribution implies the maximum likelihood estimators:<sup>14</sup>

<sup>14</sup> If  $\sigma_0 = 0$  and  $\delta_0 = (m+1)$ , we conclude that  $\Phi_T = \Phi_{\text{ml}}$  (or  $\phi_T = \phi_{\text{ml}}$ ),  $\Lambda_T = \Lambda_0 = \Sigma_0^{-1} = 0$ , and  $\Sigma_T = \Sigma_{\text{ml}}$ .

$$\begin{aligned}
& (\Sigma_e | \mathbf{y}) \sim W_m^{-1} \left( m+1+T, (\bar{\Lambda}_s)^{-1} \right); (\phi | \Sigma_e, \mathbf{y}) \sim N_{km} \left( \phi_{ml}, \Sigma_{ml}^{-1} \right) \text{ for which} \\
& \Sigma_{ml}^{-1} = E \Sigma_e = \frac{\bar{\Lambda}_s}{T}, \quad \Sigma_{ml} \equiv \left( \Sigma_e \otimes (\mathbf{X}'\mathbf{X})^{-1} \right), \quad \phi_{ml} \equiv \left( I_m \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \mathbf{y}, \text{ and} \\
& \bar{\Lambda}_s \equiv (\mathbf{Y} - \mathbf{X}\Phi_{ml})' (\mathbf{Y} - \mathbf{X}\Phi_{ml}).
\end{aligned}$$



# 4

## Properties of Mixed-frequency VAR: Simulations



The most common framework for mixed frequency estimation is when the researcher has available an information set with regularly spaced information in a complete panel. As an example, one might be interested in estimating a model with inflation and interest rates (usually measured at a monthly frequency) and output (from the National Accounts at a quarterly frequency). Instead of aggregating information on inflation and interest rates to quarterly frequency, in order to match output data, our framework allows joint estimation of the model.

For the first exercise, assume that the dataset has four-time series covering a sample  $T$  of 200 periods at a monthly frequency. In this exercise,  $N < 4$  artificial series will have equally distributed gaps along the time series ( $N$  will change with simulations). The VAR model will obtain estimates of the missing observations and compare with the true value of artificial dataset initially simulated. Time series for the exercise are simulated from the following VAR at monthly frequency:

$$Y_t = AY_{t-1} + Se_t,$$

where

$$A = \begin{bmatrix} 0.900 & 0.010 & -0.020 & 0.050 \\ 0.000 & 0.900 & -0.113 & -0.010 \\ 0.000 & 0.195 & 0.800 & 0.000 \\ -0.269 & 0.000 & 0.000 & 0.700 \end{bmatrix} \quad S = \begin{bmatrix} 0.01 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.01 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.01 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.01 \end{bmatrix}$$

Values on matrix  $A$  were selected in order to ensure the same long-run variance for each observed variable in vector  $Y_t$ , given the lack of contemporaneous conditional correlation given by matrix  $S$ . Alternative values for matrix  $S$  were tested, imposing some correlation across observed variables. Results were qualitatively similar.

Table 1 shows the median of the RMSE across 1,000 simulated time series, measured as a proportion of the unconditional standard deviation of the series in the true VAR. A total of six simulations are shown in each line. First, each group of two columns computes the RMSE using the mean, the median and the mode of the simulations in the Gibbs sampling as the forecast for the missing observation. For each group of columns, two simulations try to disentangle the uncertainty from parameter estimation in the VAR when simulating the missing observations: the columns labeled *True values* report the RMSE of the estimated VAR against the true values of the missing observations; the columns labeled *True VAR* use the true values of parameters describing the VAR to simulate the missing observation. Table 1 also split the analysis in terms of number of observed time series with missing values in the VAR ( $N$ ) and the frequency of missing observations ( $Q$ ,  $S$ , and  $Y$  for information in each quarter, semester and year, respectively), while keeping the sample size,  $T = 200$ , fixed.

Table 1 shows three interesting results. First, as expected, the RMSE of simulated time series of the mixed-frequency VAR does depend on the frequency of missing observations. For all exercises shown in the table, changing the frequency of missing informations from quarterly to yearly data increased the RMSE of the simulations, irrespective of the statistic summarizing the forecast of missing values. Despite a huge increase in the RMSE when using the mode of simulations in the Gibbs sampling with yearly information in three time series of the VAR, all simulations showed similar variations in the RMSE when changing the frequency of missing information in the model.

Second, also expected, the RMSE of simulated time series does depend on the number of time series with missing observations in the VAR. While this result does not look significant when information is available at a quarterly frequency, simulations with semiannual and, especially, annual data show that increasing the number of time series

**Table 1**

## MEDIAN RMSE AND SIMULATION OF MISSING VALUES

	<i>RMSE - mean</i>		<i>RMSE - median</i>		<i>RMSE - mode</i>	
	<i>True values</i>	<i>True VAR</i>	<i>True values</i>	<i>True VAR</i>	<i>True values</i>	<i>True VAR</i>
<i>T = 200; N = 1</i>						
F = Q	0.4181	0.1074	0.4182	0.1068	0.4531	0.2073
F = S	0.6592	0.4186	0.6572	0.4137	0.7966	0.6155
F = Y	0.9520	0.7655	0.9461	0.7592	1.1432	0.9899
<i>T = 200; N = 2</i>						
F = Q	0.4143	0.0976	0.4143	0.0972	0.4414	0.1813
F = S	0.6706	0.4357	0.6634	0.4264	0.8381	0.6622
F = Y	1.1254	0.9445	1.0264	0.8336	1.8651	1.7677
<i>T = 200; N = 3</i>						
F = Q	0.4301	0.0912	0.4293	0.0904	0.4555	0.1718
F = S	0.6677	0.3866	0.6630	0.3788	0.8505	0.6517
F = Y	1.6257	1.4690	1.1518	0.9336	6.6475	6.6008

**Table 2**

## MEDIAN RMSE AND SIMULATION OF MISSING VALUES, T = 400

	<i>RMSE - mean</i>		<i>RMSE - median</i>		<i>RMSE - mode</i>	
	<i>True values</i>	<i>True VAR</i>	<i>True values</i>	<i>True VAR</i>	<i>True values</i>	<i>True VAR</i>
<i>T = 400; N = 1</i>						
F = Q	0.4119	0.0817	0.4120	0.0819	0.4366	0.1616
F = S	0.6107	0.3431	0.6100	0.3424	0.6915	0.4753
F = Y	0.8119	0.5677	0.8081	0.5619	0.9963	0.8011
<i>T = 400; N = 2</i>						
F = Q	0.4103	0.0750	0.4103	0.0750	0.4245	0.1343
F = S	0.6083	0.3340	0.6067	0.3324	0.6802	0.4543
F = Y	0.8692	0.6334	0.8516	0.6090	1.1470	0.9759
<i>T = 400; N = 3</i>						
F = Q	0.4185	0.0616	0.4184	0.0615	0.4310	0.1231
F = S	0.6144	0.3049	0.6140	0.3049	0.6870	0.4327
F = Y	1.0238	0.7834	0.9577	0.6968	1.7854	1.6559



with missing observations increases the RMSE of simulations, hurting the performance of the algorithm.

Finally, there is a significant gain in knowing the true values of parameters in the system, meaning that estimation problems should be carefully addressed in empirical applications. Comparing results from columns labeled *True values* with those using the true set of parameters of the VAR, labeled *True VAR*, the gains in terms of RMSE in knowing exactly the parameters of the model are significant, irrespective of the statistic summarizing the simulations in the Gibbs sampler and the frequency of missing information.

Table 2 considers the case where the sample available,  $T$ , at the monthly frequency is twice the size of the benchmark exercise. Results show that the performance of the VAR with missing data at quarterly frequency does not show significant improvement, unless all parameters of the system of equations are known. The ratio between the RMSE and the volatility of the simulated time series remains almost the same in both exercises with missing values at a quarterly frequency. On the other hand, simulations with observable variables in each semester or in each year show a significant improvement, even considering the uncertainty related to parameter estimation of the VAR.



# 5

# Empirical Analysis



In this section, two exercises show the ability of MF-VAR to simulate time series of economic indicators that might be useful for economic analysis. The two exercises explore some of the properties discussed in the previous section, while also showing the flexibility of the procedure when describing the datasets. The first exercise provides estimates of time series of real GDP for Brazil at a monthly frequency, using a large set of observable variables to build an indicator of economic activity at a higher frequency. In this exercise, estimates show the consequences of properly specifying the transformation matrix  $\Lambda_z$  in empirical applications.

The second exercise tries to retrieve information from a large set of observable variables to build a time series for monthly unemployment rates based on the latest release of the unemployment survey for Brazil—the so-called “Pesquisa Nacional por Amostra de Domicílios Contínua” (Continuous National Household Sample Survey, PNAD-C)—. The first results of this survey cover the period between March 2012 and March 2015. Using a set of MF-VAR models with different observable variables, including information from an annual version of PNAD, the section explores the ability of the models to generate consistent estimates of unemployment rate if missing data are irregularly spaced.

## 5.1 Estimating Real GDP at a Monthly Frequency

One interesting application of mixed-frequency VARs (MF-VARs) is to build high-frequency estimations of real gross domestic product using indicators of economic activity. Here, we use a group of economic indicators to build five-variable MF-VARs and analyze the properties of the combination of the forecasts of these MF-VARs. These indicators include total industrial production, industrial production of capital goods, retail sales, total exports, total imports, energy consumption, energy consumption of manufacturing industry, the ratio of the trade balance over the trade flow, oil production and steel production. All these indicators are available at a monthly frequency. The MF-VARs include seasonal dummies and change the assumption about deterministic trends between no trend, linear trend or a quadratic trend.

In empirical applications of the MF-VAR, it is important to properly characterize the setup of matrix  $\Lambda_z$  in Equation 2, relating observable variables at high- and low-frequency. Figure 1 below shows the simulated path from one of the MF-VARs<sup>15</sup> for the quarterly growth rates of high-frequency estimates of (log-) real GDP<sup>16</sup> under two assumptions for matrix  $\Lambda_z$ : the green line assumes a simple setup, where observed real GDP is directly correlated with high-frequency observation in the last month of the quarter; the orange line sets a proper framework, with observed real GDP as the sum of simulated real GDP in the three months of the quarter. The brown line shows an indicator of high-frequency estimates of real GDP calculated by the Banco Central do Brasil—the so-called IBC-BR. It is clear from the figure that a proper setup of matrix  $\Lambda_z$  allows a closer match between the simulated path of the MF-VAR and the estimates provided by IBC-BR, as the real GDP in levels tends

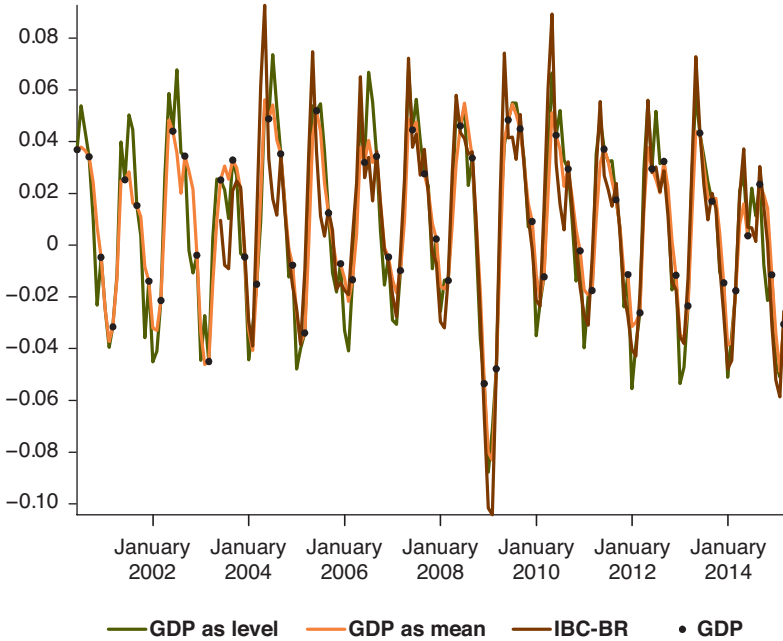
---

<sup>15</sup> The MF-VAR has three lags, and includes, besides the log of GDP at quarterly frequency, seasonal factors, and a linear trend, the log of production of capital goods, retail sales, exports, and energy consumption.

<sup>16</sup> Data on real GDP used in this exercise was published by IBGE in the end of May 2015, with first vintage of data for the first quarter of 2015. This exercise does not handle with *real time forecasting*, as a methodological change in GDP computation prevents a clear comparison between recent and previous vintages of data.

**Figure 1**

**QUARTELY GDP CHANGE IN HIGH-FREQUENCY**  
June-2000 to March-2015



to systematically overestimate seasonal movements early in every year, while also missing the location of peaks in the middle of each year. The setup with real GDP as a quarterly mean provides a better match with IBC-BR at both these points.

For the exercise, a combined forecast of MF-VARs is compared with the forecasts of a simple AR(1) process for the growth rate of (log-) real GDP with seasonal dummies and a set of Bayesian VARs at the quarterly frequency with the same observables used in MF-VARs. The main objective of using Bayesian VARs to compare with MF-VARs is to measure the effect of using high-frequency information to build forecasts, instead of simply aggregating information at lower

**Table 3**

FORECASTING GDP WITH MF-VAR: RMSE, DECEMBER 2010 TO MARCH 2015

	<i>Observations</i>	<i>RMSE</i>			<i>Relative RMSE</i>	
		<i>MF -VAR</i>	<i>VAR</i>	<i>AR (1)</i>	<i>VAR</i>	<i>AR (1)</i>
$t = 0$	18	0.0083	0.0200	0.0109	2.40	1.31
$t = 1$	17	0.0170	0.0298	0.0157	1.76	0.92
$t = 2$	16	0.0256	0.0291	0.0189	1.08	0.74
$t = 3$	15	0.0298	0.0291	0.0256	0.98	0.86
$t = 4$	14	0.0341	0.0415	0.0277	1.22	0.81

frequencies. The comparison with an AR(1) process provides a benchmark in terms of forecasting ability of the MF-VAR. The MF-VARs are arbitrarily set with three lags, in order to properly characterize the dynamics between monthly and quarterly frequency data. Bayesian VARs at a quarterly frequency, on the other hand, due to the small sample size, are estimated with only one lag.

Table 3 compares the outcome of forecasts in the three models. In the table,  $t=0$  shows the properties of the so-called *nowcasting* in the MF-VAR, which is the forecast for real GDP at a given time  $t$ , given that all information for other variables, except GDP, is available for forecasting. It is worth noting that the additional information building the *nowcast* affects only forecasts from the MF-VARs and the Bayesian VARs; for the AR(1) process, the *nowcasting* is equivalent to a one-step-ahead forecast.

Table 3 shows the gains of using high-frequency information, instead of aggregating data to lower frequencies. The performance of the MF-VAR is, in a worst-case scenario, similar to the Bayesian VAR, when forecasting three quarters ahead. In all other horizons considered, the performance of the MF-VAR clearly dominates the Bayesian VAR. However, despite providing good performance in terms of *nowcasting*, both the MF-VAR and the Bayesian VAR are not better than a univariate system estimated at a quarterly frequency. The main advantage of using the MF-VAR, thus, seems to be associated with the ability to use high-frequency information to make inference about the current period.

## 5.2 Estimating the Unemployment Rate with Irregularly Spaced Data

One of the possible uses of MF-VARs is in the simulation of missing information in the time series of a variable when there are other proxy variables to approximate the dynamics of the variable with missing information. Here, we present a set of MF-VARs designed to estimate the missing values of the unemployment rate in Brazil measured by the PNAD-C, a survey of the Instituto Brasileiro de Geografia e Estatística (Brazilian Institute of Geography and Statistics, IBGE). This survey shows estimates of the unemployment rate in the previous quarter at a monthly frequency, covering almost four times the total population surveyed in the old survey conducted by IBGE (Pesquisa Mensal de Emprego –Monthly Employment Survey–, PME). There is also an annual survey from IBGE, called “Pesquisa Nacional por Amostra de Domicílios” (National Household Sample Survey, PNAD), where information on employment is also collected, with similar coverage compared to PNAD-C. Information on employment status in Brazil is very irregular. However, given the role of labor markets for economic analysis, generating a time series of unemployment based on PNAD-C constitutes a valuable piece of information. Table 4 summarizes the main characteristics of four surveys on employment for Brazil.

**Table 4**

UNEMPLOYMENT SURVEYS IN BRAZIL				
	<i>PME (IBGE)</i>	<i>PED (DIEESE)</i>	<i>PNAD (IBGE)</i>	<i>PNAD-C (IBGE)</i>
Frequency	Monthly	Monthly	Annual	Monthly <sup>1</sup>
Coverage	Six metro areas	Four metro areas	Brazil	Brazil
Population ratio (2012)	0.25	0.15	1.00	1.00
Time span	1999M05-2015M03	1999M05-2015M03	1999-2013	2012M03-2015M03
Observation	Major review in 2001	Restricted to same metro areas	Missing censitary years (2000 and 2010)	Information about last quarter

<sup>1</sup> Information refers to the moving average of the quarter finished at a given month. Sources: IBGE and DIEESE (Departamento Intersindical de Estatística e Estudos Socioeconômicos)



In order to simulate time series for unemployment in Brazil based on PNAD-C, we first assume that unemployment rates from PNAD-C and PNAD differ only by a multiplicative factor,<sup>17</sup> which is retrieved by considering the only month for which we have observations from both series, that is, September 2012.<sup>18</sup> After that, we adjust the level of PNAD annual observations and merge them with PNAD-C monthly observations. Given a linear relation between the overlapping information between PNAD and PNAD-C, we set at least one point of data for the period between 1999 and 2011—the period before PNAD-C results was published—. After that, a set of four-variable MF-VARS was estimated using information from PME and PED surveys, with measures of unemployment and participation rates, *Cadastro Geral de Empregados e Desempregados* (general record of employees and unemployees, CAGED) data of formal employment and *Confederação Nacional da Indústria* (national confederation of industry, CNI) data on employment in the manufacturing sector. MF-VARS included seasonal dummies and were estimated with and without deterministic linear and quadratic trends. Estimations including a larger number of observable variables were not performed due to the small sample size combined with larger gaps in the information from PNAD-C. Also, in terms of restriction to variable selection in the VARS, at least one of the unemployment time series was included in the estimation, together with PNAD-C data. A total of 36 models were used to simulate the time series of unemployment.

Final estimates based on the mean and the median for the time series of PNAD-C are shown in Figure 2, together with confidence bands based on the distribution of simulations across models. As expected, confidence bands are larger when the gap between information available at PNAD-C is also larger. Also as expected, confidence bands increase during turning points of the simulated time series. Of course, as the analysis in Section 4 suggests, as more data for PNAD-C become available, it should be possible to obtain better estimates of the

---

<sup>17</sup> Results do not change much when we assume an additive factor instead.

<sup>18</sup> Figures from PNAD evaluate employment situation on the last day of September in each year.

dynamics of the time series during these turning points. As a consequence, we expect these confidence intervals to become smaller with the increase in the amount of information from PNAD-C. We have also run an in-sample forecasting exercise over the last year of the sample, in order to set weights across different models based on the mean square error of the models. The final time series for unemployment generated from this exercise, not shown in Figure 2, is very close to the median of the simulations across models.

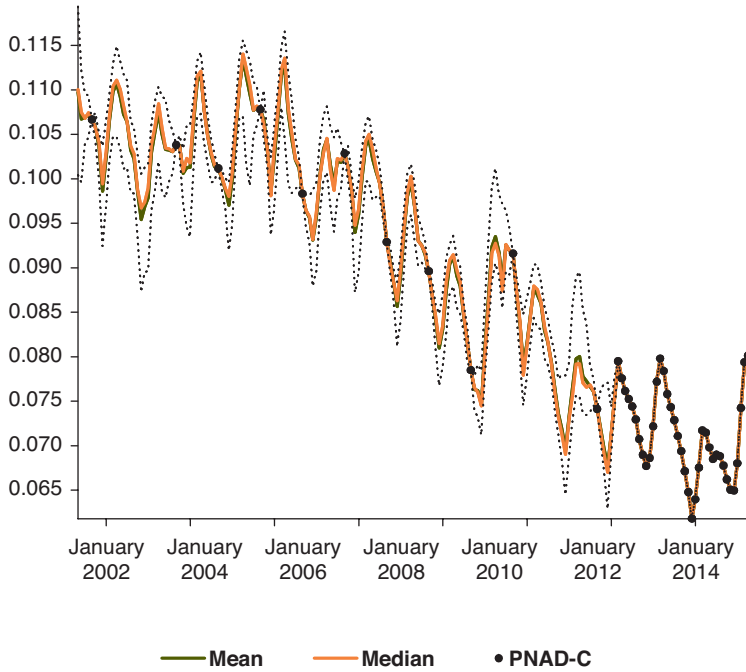
The simulated time series follows a close path when compared to the actual time series of PME and PED surveys. The main difference between these two surveys and the simulated path of PNAD-C presented here is related to the gap between the minimum and the maximum of each time series: While PME and PED, on average, showed unemployment rates above 12% in the period between 1999 and 2004—even reaching 19% in one of the measures provided in PED—the average of PNAD-C in the sample is set at 10.5%. On the other hand, for the late part of the sample, while PME reached historical lows in December 2013 and December 2014 of 4.3%, the minimum of PNAD-C—which, in the sample, is an observable, not a simulation—was found in December 2013, at 6.2%. The main reason for this discrepancy is the use of annual information from PNAD for the period between 1999 and 2011: without the support of such information, simulated paths of PNAD-C would be closer to the mean between PME and PED for the early part of the sample.

---

**Figure 2**

**SIMULATED TIME SERIES: PNAD-C**

May-1999 to April-2015





# 6

## Conclusion



This study presented a new efficient algorithm to draw an inference from the joint distribution of a dataset with missing information. The algorithm presented here is flexible enough to accommodate most of the issues faced when dealing with irregularly spaced datasets. It also allows for easy setup of linear restrictions in observable variables, irrespective of the presence of missing information in those time series. From this perspective, the implementation of the model uses most of the findings in early literature on the subject, generalizing some results and applications for the framework of mixed-frequency estimation of VAR models.

Both the analysis with simulated and real datasets show the flexibility of the framework to properly handle datasets with an usual size observed in the real world's applications. Despite results with simulated data suggest that some careful analysis of the convergence of the Gibbs sampler is necessary to ensure that simulated time series are being drawn from appropriate parameter values, the small volatility of simulations shows that the algorithm is a very useful tool to handle problems associated with missing information in large datasets.

In terms of the empirical applications with real data, the main results show the importance of using high-frequency information to make inference on a macroeconomic dataset with missing values. The applications show the advantage of using high-frequency information in two different contexts: first, in terms of forecasting,

MF-VARs using the framework proposed here had a better performance than traditional Bayesian MF-VARs estimated using data at lower frequencies; second, the algorithm proposed here estimated past trajectories of variables with missing information combining a large set of information, generating simulations with relatively small dispersion around its mean.



# Annex





## Annex A. Some Results

### *Matrix Algebra*

Recall that, for any matrices  $A_{k \times \ell}$ ,  $B_{\ell \times m}$ , and  $C_{m \times n}$  the  $vec(\cdot)$  operator satisfies the following properties:

$$tr(A'B) = (vec(B))' vec(A) = (vec(A))' vec(B)$$

$$vec(ABC) = (C' \otimes \mathbf{A})vec(B)$$

$$vec(ABC) = vec(AB \cdot C \cdot I_n) = (I_n \otimes AB)vec(C)$$

$$vec(ABC) = vec(I_k \cdot A \cdot BC) = (C'B' \otimes I_k)vec(A)$$

$$vec(AB) = vec(A \cdot B \cdot I_m) = (I_m \otimes \mathbf{A})vec(B)$$

$$vec(AB) = vec(I_k \cdot A \cdot B) = (B' \otimes I_k)vec(A)$$

Recall that, for any matrices  $A$ ,  $B$ ,  $C$ , and  $D$ , the Kronecker product satisfies (when it makes sense) the following properties:

$$(A \otimes B)(C \otimes D) = (AC \otimes BD)$$

$$(A \otimes B)^{-1} = (A^{-1} \otimes B^{-1})$$

$$\text{tr}(A \otimes B) = \text{tr}(A) \cdot \text{tr}(B)$$

$$(A \otimes B)' = (A' \otimes B)$$

$$|A \otimes B| = |A|^\ell |B|^k \text{ if } A \text{ is } (k \times k), \text{ and } B \text{ is } (\ell \times \ell).$$

### Maximum Likelihood Estimator–Multivariate Gaussian Distribution

The estimator comes from  $0 = -\left(\Sigma_\varepsilon^{-\frac{1}{2}} \mathbf{X}\right)' \left[\Sigma_\varepsilon^{-\frac{1}{2}} (\mathbf{y} - \mathbf{X}\phi_{\text{ml}})\right]^{-1}$  which implies:

$$\begin{aligned} \phi_{\text{ml}} &= \left[ \left(\Sigma_\varepsilon^{-\frac{1}{2}} \mathbf{X}\right)' \left(\Sigma_\varepsilon^{-\frac{1}{2}} \mathbf{X}\right) \right]^{-1} \left(\Sigma_\varepsilon^{-\frac{1}{2}} \mathbf{X}\right)' \left(\Sigma_\varepsilon^{-\frac{1}{2}} \mathbf{y}\right) \\ &= \left[ \left(\Sigma_\varepsilon \otimes I_T\right)^{-\frac{1}{2}} (I_m \otimes \mathbf{X}) \right]' \left(\Sigma_\varepsilon \otimes I_T\right)^{-\frac{1}{2}} (I_m \otimes \mathbf{X}) \right]^{-1} \left(\Sigma_\varepsilon \otimes I_T\right)^{-\frac{1}{2}} (I_m \otimes \mathbf{X}) \left(\Sigma_\varepsilon \otimes I_T\right)^{-\frac{1}{2}} \mathbf{y} \\ &= \left[ \left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T\right) (I_m \otimes \mathbf{X}) \right]' \left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T\right) (I_m \otimes \mathbf{X}) \right]^{-1} \left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T\right) (I_m \otimes \mathbf{X}) \left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T\right) \mathbf{y} \\ &= \left[ \left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X}\right)' \left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X}\right) \right]^{-1} \left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X}\right)' \left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T\right) \mathbf{y} \\ &= \left[ \left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X}'\right) \left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X}\right) \right]^{-1} \left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X}'\right) \left(\Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T\right) \mathbf{y} \\ &= \left(\Sigma_\varepsilon^{-1} \otimes \mathbf{X}'\mathbf{X}\right)^{-1} \left(\Sigma_\varepsilon^{-1} \otimes \mathbf{X}'\right) \mathbf{y} = \left(\Sigma_\varepsilon \otimes (\mathbf{X}'\mathbf{X})^{-1}\right) \left(\Sigma_\varepsilon^{-1} \otimes \mathbf{X}'\right) \mathbf{y} = \left(I_m \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\right) \mathbf{y}. \end{aligned}$$

### Rewriting the likelihood function

Note that:

$$\begin{aligned} p(\mathbf{y}|\phi, \Sigma_\varepsilon) &= |2\pi \Sigma_\varepsilon|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\phi)' \Sigma_\varepsilon^{-1} (\mathbf{y} - \mathbf{X}\phi) \right] \\ &= |2\pi \Sigma_\varepsilon|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \Sigma_\varepsilon^{-\frac{1}{2}} (\mathbf{y} - \mathbf{X}\phi) \right]' \left[ \Sigma_\varepsilon^{-\frac{1}{2}} (\mathbf{y} - \mathbf{X}\phi) \right] \right\}. \end{aligned}$$

And hence, replacing  $(\phi_{ml} + \phi - \phi_{ml})$  for  $\phi$ , we obtain:

$$\begin{aligned}
\left[ \Sigma_\varepsilon^{-\frac{1}{2}}(\mathbf{y} - \mathbf{x}\phi) \right]' \left[ \Sigma_\varepsilon^{-\frac{1}{2}}(\mathbf{y} - \mathbf{x}\phi) \right] &= \left[ \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T \right) (\mathbf{y} - (I_m \otimes \mathbf{X})\phi) \right]' \left[ \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T \right) (\mathbf{y} - (I_m \otimes \mathbf{X})\phi) \right] \\
&= \left[ \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T \right) \mathbf{y} - \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X} \right) \phi \right]' \left[ \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T \right) \mathbf{y} - \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X} \right) \phi \right] \\
&= \left[ \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T \right) \mathbf{y} - \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X} \right) \phi_{ml} \right]' \left[ \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T \right) \mathbf{y} - \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X} \right) \phi_{ml} \right] + \\
&\quad + (\phi - \phi_{ml})' \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X} \right)' \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X} \right) (\phi - \phi_{ml}) - 2(\phi - \phi_{ml})' \chi_\varepsilon,
\end{aligned}$$

where

$$\begin{aligned}
\chi_\varepsilon &\equiv \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X} \right)' \left[ \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T \right) \mathbf{y} - \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X} \right) \phi_{ml} \right] \\
&= \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X} \right)' \left[ \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes I_T \right) - \left( \Sigma_\varepsilon^{-\frac{1}{2}} \otimes \mathbf{X} \right) \left( I_m \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \right] \mathbf{y} \\
&= \left[ \left( \Sigma_\varepsilon^{-1} \otimes \mathbf{X}' \right) - \left( \Sigma_\varepsilon^{-1} \otimes \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \right] \mathbf{y} \\
&= \left[ \left( \Sigma_\varepsilon^{-1} \otimes \mathbf{X}' \right) - \left( \Sigma_\varepsilon^{-1} \otimes \mathbf{X}' \right) \right] \mathbf{y} = 0.
\end{aligned}$$

Therefore

$$p(\mathbf{y}|\phi, \Sigma_\varepsilon) = |2\pi \Sigma_\varepsilon|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2}(\mathbf{y} - \mathbf{x}\phi_{ml})' (\Sigma_\varepsilon \otimes I_T)^{-1} (\mathbf{y} - \mathbf{x}\phi_{ml}) - \frac{1}{2}(\phi - \phi_{ml})' (\Sigma_\varepsilon^{-1} \otimes \mathbf{X}'\mathbf{X}) (\phi - \phi_{ml}) \right],$$

where

$$\begin{aligned}
(\mathbf{y} - \mathbf{x}\phi_{ml})' (\Sigma_\varepsilon \otimes I_T)^{-1} (\mathbf{y} - \mathbf{x}\phi_{ml}) &= (\text{vec}(\mathbf{Y} - \mathbf{X}\Phi_{ml}))' (\Sigma_\varepsilon^{-1} \otimes I_T) \text{vec}(\mathbf{Y} - \mathbf{X}\Phi_{ml}) \\
&= (\text{vec}(\mathbf{Y} - \mathbf{X}\Phi_{ml}))' \text{vec}(I_T (\mathbf{Y} - \mathbf{X}\Phi_{ml}) \Sigma_\varepsilon^{-1}) \\
&= \text{tr} \left[ (\mathbf{Y} - \mathbf{X}\Phi_{ml})' (\mathbf{Y} - \mathbf{X}\Phi_{ml}) \Sigma_\varepsilon^{-1} \right].
\end{aligned}$$

Note also that

$$\text{vec}(\Phi_{\text{ml}}) = \phi_{\text{ml}} = \left( I_m \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \mathbf{y} = \left( I_m \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \text{vec}(\mathbf{Y}) = \text{vec} \left( (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{Y} \right),$$

$$\begin{aligned} \text{vec}(\mathbf{X}\Phi_{\text{ml}}) &= \mathbf{x}\phi_{\text{ml}} = (I_m \otimes \mathbf{X}) \left( I_m \otimes (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \mathbf{y} = \left( I_m \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \mathbf{y} \\ &= \left( I_m \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \text{vec}(\mathbf{Y}) = \text{vec} \left( \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{Y} \right). \end{aligned}$$

Joint posterior PDF of the Bayesian VAR. We need to compute the following summation terms:

$$\begin{aligned} S_1 &= (\phi - \phi_0)' \Sigma_0^{-1} (\phi - \phi_0) + (\phi - \phi_{\text{ml}})' \Sigma_{\text{ml}}^{-1} (\phi - \phi_{\text{ml}}) \\ &= (\phi - \phi_T)' \Sigma_T^{-1} (\phi - \phi_T) + \phi_0' \Sigma_0^{-1} \phi_0 + \phi_{\text{ml}}' \Sigma_{\text{ml}}^{-1} \phi_{\text{ml}} - \phi_T' \Sigma_T^{-1} \phi_T, \end{aligned}$$

$$\begin{aligned} S_2 &= \phi_0' \Sigma_0^{-1} \phi_0 + \phi_{\text{ml}}' \Sigma_{\text{ml}}^{-1} \phi_{\text{ml}} - \phi_T' \Sigma_T^{-1} \phi_T \\ &= \text{vec}(\Phi_0)' \left( \Sigma_e^{-1} \otimes \sigma_0 \tilde{I}_k \right) \text{vec}(\Phi_0) + \\ &\quad + \text{vec}(\Phi_{\text{ml}})' \left( \Sigma_e^{-1} \otimes (\mathbf{X}'\mathbf{X}) \right) \text{vec}(\Phi_{\text{ml}}) - \text{vec}(\Phi_T)' \left( \Sigma_e^{-1} \otimes (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X}) \right) \text{vec}(\Phi_T) \\ &= \text{vec}(\Phi_0)' \text{vec} \left( (\sigma_0 \tilde{I}_k) \Phi_0 \Sigma_e^{-1} \right) + \\ &\quad + \text{vec}(\Phi_{\text{ml}})' \text{vec} \left( (\mathbf{X}'\mathbf{X}) \Phi_{\text{ml}} \Sigma_e^{-1} \right) - \text{vec}(\Phi_T)' \text{vec} \left( (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X}) \Phi_T \Sigma_e^{-1} \right) \\ &= \text{tr} \left[ \left( \Phi_0' (\sigma_0 \tilde{I}_k) \Phi_0 + \Phi_{\text{ml}}' (\mathbf{X}'\mathbf{X}) \Phi_{\text{ml}} - \Phi_T' (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X}) \Phi_T \right) \Sigma_e^{-1} \right] = \text{tr} \left[ \Lambda_T \Sigma_e^{-1} \right], \end{aligned}$$

where

$$\Sigma_T^{-1} \equiv \left( \Sigma_0^{-1} + \Sigma_{\text{ml}}^{-1} \right) = \Sigma_e^{-1} \otimes \left( \sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X} \right),$$

$$\begin{aligned} \phi_T &\equiv \Sigma_T \left( \Sigma_0^{-1} \phi_0 + \Sigma_{\text{ml}}^{-1} \phi_{\text{ml}} \right) \\ &= \left( I_m \otimes (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\sigma_0 \tilde{I}_k) \right) \phi_0 + \left( I_m \otimes (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X}) \right) \phi_{\text{ml}}, \end{aligned}$$

$$\Phi_T \equiv \text{vec}^{-1}(\phi_T) = (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\sigma_0 \tilde{I}_k) \Phi_0 + (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X}) \Phi_{\text{ml}},$$

$$\Lambda_T \equiv \Phi_0' (\sigma_0 \tilde{I}_k) \Phi_0 + \Phi_{\text{ml}}' (\mathbf{X}'\mathbf{X}) \Phi_{\text{ml}} - \Phi_T' (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X}) \Phi_T.$$

Deriving the expression for  $\Phi_T$  goes as follows:

$$\begin{aligned} \text{vec}(\Phi_T) &\equiv \left( I_m \otimes (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\sigma_0 \tilde{I}_k) \right) \text{vec}(\Phi_0) + \left( I_m \otimes (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X}) \right) \text{vec}(\Phi_{\text{ml}}) \\ &= \text{vec} \left( (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\sigma_0 \tilde{I}_k) \Phi_0 + (\sigma_0 \tilde{I}_k + \mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X}) \Phi_{\text{ml}} \right). \end{aligned}$$

Note also that

$$\begin{aligned}
\phi_T &\equiv \Sigma_T \left( \Sigma_0^{-1} \phi_0 + \Sigma_{\text{ml}}^{-1} \phi_{\text{ml}} \right) = \left( \Sigma_0^{-1} + \Sigma_{\text{ml}}^{-1} \right)^{-1} \left( \Sigma_0^{-1} \phi_0 + \Sigma_{\text{ml}}^{-1} \phi_{\text{ml}} \right) \\
&= \left( \sigma_0 \left( \Sigma_0^{-1} \otimes \tilde{I}_k^{-1} \right) + \left( \Sigma_e^{-1} \otimes (\mathbf{X}'\mathbf{X}) \right) \right)^{-1} \left( \sigma_0 \left( \Sigma_0^{-1} \otimes \tilde{I}_k^{-1} \right) \phi_0 + \left( \Sigma_e^{-1} \otimes (\mathbf{X}'\mathbf{X}) \right) \phi_{\text{ml}} \right) \\
&= \left( \Sigma_e^{-1} \otimes \left( \sigma_0 \tilde{I}_k^{-1} + (\mathbf{X}'\mathbf{X}) \right) \right)^{-1} \left( \left( \Sigma_e^{-1} \otimes \sigma_0 \tilde{I}_k^{-1} \right) \phi_0 + \left( \Sigma_e^{-1} \otimes (\mathbf{X}'\mathbf{X}) \right) \phi_{\text{ml}} \right) \\
&= \left( \Sigma_e \otimes \left( \sigma_0 \tilde{I}_k + (\mathbf{X}'\mathbf{X}) \right) \right)^{-1} \left( \left( \Sigma_e^{-1} \otimes \sigma_0 \tilde{I}_k^{-1} \right) \phi_0 + \left( \Sigma_e^{-1} \otimes (\mathbf{X}'\mathbf{X}) \right) \phi_{\text{ml}} \right) \\
&= \left( I_m \otimes \left( \sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X} \right)^{-1} \left( \sigma_0 \tilde{I}_k^{-1} \right) \right) \phi_0 + \left( I_m \otimes \left( \sigma_0 \tilde{I}_k^{-1} + \mathbf{X}'\mathbf{X} \right)^{-1} (\mathbf{X}'\mathbf{X}) \right) \phi_{\text{ml}}.
\end{aligned}$$

## Annex B. Proofs

### *Proposition 1*

The Markov structure of the evolution equation of any DLM implies

$$\Pr(Z_T | \mathcal{D}_T) = \Pr(z_T | \mathcal{D}_T) \Pr(z_{T-1} | z_T, \mathcal{D}_{T-1}) \dots \Pr(z_1 | z_2, \mathcal{D}_1)$$

### *Proof*

Note that  $\Pr(Z_T | \mathcal{D}_T)$  can be expanded as follows:

$$\begin{aligned}
\Pr(Z_T | \mathcal{D}_T) &= \Pr(z_T, Z_{T-1} | \mathcal{D}_T) = \Pr(z_T | \mathcal{D}_T) \Pr(Z_{T-1} | z_T, \mathcal{D}_T) \\
&= \Pr(z_T | \mathcal{D}_T) \Pr(z_{T-1}, Z_{T-2} | z_T, \mathcal{D}_T) \\
&= \Pr(z_T | \mathcal{D}_T) \Pr(z_{T-1} | z_T, \mathcal{D}_T) \Pr(Z_{T-2} | z_{T-1}, z_T, \mathcal{D}_T) \\
&= \Pr(Z_T | \mathcal{D}_T) \Pr(z_{T-1} | z_T, \mathcal{D}_T) \Pr(z_{T-2}, Z_{T-3} | z_{T-1}, z_T, \mathcal{D}_T) \\
&= \Pr(z_T | \mathcal{D}_T) \dots \underbrace{\Pr(z_{T-2} | z_{T-1}, z_T, \mathcal{D}_T)}_{\Pr(z_{T-2} | z_{t>(T-2)}, \mathcal{D}_T)} \Pr(Z_{T-3} | z_{T-2}, z_{T-1}, z_T, \mathcal{D}_T) \\
&\quad \vdots \\
&= \Pr(z_T | \mathcal{D}_T) \Pr(z_{T-1} | z_{t>(T-1)}, \mathcal{D}_T) \Pr(z_{T-2} | z_{t>(T-2)}, \mathcal{D}_T) \dots \Pr(z_1 | z_{t>1}, \mathcal{D}_T).
\end{aligned}$$

Let now  $y_{t>\bar{r}}^{ob} \equiv [y_{\bar{r}+1}^{ob}, y_T^{ob}]$  and  $z_{t>\bar{r}} \equiv [z_{\bar{r}+1}, z_T]$  denote the sets of all observable and state variables for periods between  $\bar{r} + 1$  up to  $T$ . Recall that the observation equation implies that, conditional on  $z_t$ ,  $y_t^{ob}$  does not depend on past values of the state variable. As shown below, the Markovian property copped with this fact implies let us show that

$\Pr(z_{T-j} | z_{l>T-j}, \mathcal{D}_T) = \Pr(z_{T-j} | z_{T-j+1}, \mathcal{D}_{T-j})$ . Indeed:

$$\begin{aligned} \Pr(z_{T-j} | z_{l>T-j}, \mathcal{D}_T) &= \Pr(z_{T-j} | z_{l>T-j}, y_{l>T-j}^{ob}, \mathcal{D}_{T-j}) \\ &= \frac{\Pr(z_{T-j}, y_{l>T-j}^{ob} | z_{l>T-j}, \mathcal{D}_{T-j})}{\Pr(y_{l>T-j}^{ob} | z_{l>T-j}, \mathcal{D}_{T-j})}, \\ &= \frac{\Pr(y_{l>T-j}^{ob} | z_{T-j}, z_{l>T-j}, \mathcal{D}_{T-j}) \Pr(z_{T-j} | z_{l>T-j}, \mathcal{D}_{T-j})}{\Pr(y_{l>T-j}^{ob} | z_{l>T-j}, \mathcal{D}_{T-j})} \end{aligned}$$

*Due to observed equation*

$$\begin{aligned} &= \frac{\Pr(y_{l>T-j}^{ob} | z_{l>T-j}, \mathcal{D}_{T-j}) \Pr(z_{T-j} | z_{l>T-j}, \mathcal{D}_{T-j})}{\Pr(y_{l>T-j}^{ob} | z_{l>T-j}, \mathcal{D}_{T-j})} \\ &= \Pr(z_{T-j} | z_{l>T-j}, \mathcal{D}_{T-j}) \\ &= \Pr(z_{T-j} | z_{T-j+1}, z_{l>T-j+1}, \mathcal{D}_{T-j}) \\ &= \frac{\Pr(z_{T-j}, z_{l>T-j+1} | z_{T-j+1}, \mathcal{D}_{T-j})}{\Pr(z_{l>T-j+1} | z_{T-j+1}, \mathcal{D}_{T-j})} \\ &= \frac{\Pr(z_{l>T-j+1} | z_{T-j}, z_{T-j+1}, \mathcal{D}_{T-j}) \Pr(z_{T-j} | z_{T-j+1}, \mathcal{D}_{T-j})}{\Pr(z_{l>T-j+1} | z_{T-j+1}, \mathcal{D}_{T-j})} \end{aligned}$$

*Markovian property*

$$\begin{aligned} &= \frac{\Pr(z_{l>T-j+1} | z_{T-j+1}, \mathcal{D}_{T-j}) \Pr(z_{T-j} | z_{T-j+1}, \mathcal{D}_{T-j})}{\Pr(z_{l>T-j+1} | z_{T-j+1}, \mathcal{D}_{T-j})} \\ &= \Pr(z_{T-j} | z_{T-j+1}, \mathcal{D}_{T-j}). \end{aligned}$$

Using our previous result, we conclude that

$$\Pr(Z_T | \mathcal{D}_T) = \Pr(z_T | \mathcal{D}_T) \Pr(z_{T-1} | z_T, \mathcal{D}_{T-1}) \dots \Pr(z_1 | z_2, \mathcal{D}_1).$$

■

**Proposition 2**

The conditional density of  $(y_{t-q+1} | z_{t+1}, \mathcal{D}_t)$  is proportional to

$$\Pr(y_{t-q+1} | z_{t+1}, \mathcal{D}_t) \propto \Pr(\tilde{e}_{t+1} | \tilde{z}_{t+1}, y_{t-q+1}, \mathcal{D}_t) \Pr(y_{t-q+1} | \tilde{z}_{t+1}, \mathcal{D}_t)$$

**Proof**

$$\begin{aligned} \Pr(y_{t-q+1} | z_{t+1}, \mathcal{D}_t) &= \Pr(y_{t-q+1} | y_{t+1}, \tilde{z}_t, \mathcal{D}_t) \\ &= \frac{\Pr(y_{t+1}, y_{t-q+1} | \tilde{z}_t, \mathcal{D}_t)}{\Pr(y_{t+1} | \tilde{z}_t, \mathcal{D}_t)} = \frac{\Pr(y_{t+1} | \tilde{z}_t, y_{t-q+1}, \mathcal{D}_t) \Pr(y_{t-q+1} | \tilde{z}_t, \mathcal{D}_t)}{\Pr(y_{t+1} | \tilde{z}_t, \mathcal{D}_t)} \end{aligned}$$

$$\begin{aligned} \text{Conditioning} & \\ \text{first term} & \\ \text{also on } \tilde{e}_{t+1} & \\ &= \frac{\Pr(y_{t+1}, \tilde{e}_{t+1} | \tilde{z}_t, y_{t-q+1}, \mathcal{D}_t) \Pr(y_{t-q+1} | \tilde{z}_t, \mathcal{D}_t)}{\Pr(\tilde{e}_{t+1} | y_{t+1}, \tilde{z}_t, y_{t-q+1}, \mathcal{D}_t) \Pr(y_{t+1} | \tilde{z}_t, \mathcal{D}_t)} \\ &= \frac{\Pr(y_{t+1} | \tilde{z}_t, y_{t-q+1}, \tilde{e}_{t+1}, \mathcal{D}_t) \Pr(\tilde{e}_{t+1} | \tilde{z}_t, y_{t-q+1}, \mathcal{D}_t) \Pr(y_{t-q+1} | \tilde{z}_t, \mathcal{D}_t)}{\Pr(\tilde{e}_{t+1} | y_{t+1}, \tilde{z}_t, y_{t-q+1}, \mathcal{D}_t) \Pr(y_{t+1} | \tilde{z}_t, \mathcal{D}_t)} \end{aligned}$$

$$\begin{aligned} \text{Conditioned} & \\ \text{on } \tilde{e}_{t+1}, \text{ the} & \\ \text{first term} & \\ \text{does not} & \\ \text{depend on} & \\ y_{t-q+1} & \\ &= \frac{\Pr(y_{t+1} | \tilde{z}_t, \tilde{e}_{t+1}, \mathcal{D}_t) \Pr(\tilde{e}_{t+1} | \tilde{z}_t, y_{t-q+1}, \mathcal{D}_t) \Pr(y_{t-q+1} | \tilde{z}_t, \mathcal{D}_t)}{\Pr(\tilde{e}_{t+1} | y_{t+1}, \tilde{z}_t, \mathcal{D}_t) \Pr(y_{t+1} | \tilde{z}_t, \mathcal{D}_t)}. \end{aligned}$$

Since terms  $\Pr(y_{t+1} | \tilde{z}_t, \tilde{e}_{t+1}, \mathcal{D}_t)$ ,  $\Pr(\tilde{e}_{t+1} | y_{t+1}, \tilde{z}_t, \mathcal{D}_t)$  and  $\Pr(y_{t+1} | \tilde{z}_t, \mathcal{D}_t)$  are not functions of  $y_{t-q+1}$ , they are constants with respect to  $\Pr(y_{t-q+1} | z_{t+1}, \mathcal{D}_t)$ .

Therefore, we obtain:

$$\Pr(y_{t-q+1} | z_{t+1}, \mathcal{D}_t) \propto \Pr(\tilde{e}_{t+1} | \tilde{z}_{t+1}, y_{t-q+1}, \mathcal{D}_t) \Pr(y_{t-q+1} | \tilde{z}_{t+1}, \mathcal{D}_t).$$

■







## References



- Basdevant, Olivier (2003), *On Applications of State-space Modeling in Macroeconomics*, Reserve Bank of New Zealand Discussion Paper Series, No. DP2003/02, Reserve Bank of New Zealand.
- Canova, Fabio (2007), *Methods for Applied Macroeconomic Research*, Princeton University Press, Princeton.
- Carter, Christopher K., and Robert Kohn (1994), "On Gibbs Sampling for State Space Models," *Biometrika*, Vol. 81, No. 3, August, pp. 541-553.
- Doan, Thomas, Robert B. Litterman, and Christopher A. Sims (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, Vol. 3, No. 1, January, pp. 1-100.
- Durbin, James, and Siem Jan Koopman (2012), *Time Series Analysis by State Space Methods*, 2nd ed., No. 38, Oxford Statistical Science, Oxford University Press, August.
- Eraker, Bjørn, Ching Wai (Jeremy) Chiu, Andrew T. Foerster, Tae Bong Kim, and Hernán D. Seoane (2015), "Bayesian Mixed Frequency VARs," *Journal of Financial Econometrics*, Vol. 13, No. 3, pp. 698-721.
- Frühwirth-Schnatter, Sylvia (1994), "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis*, Vol. 15, No. 2, March, pp. 183-202.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (2003), *Bayesian Data Analysis, Second Edition*, Chapman and Hall/CRC, Boca Raton, Fla., July.
- Hamilton, James D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, N. J.

- Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, Vol. 82, No. 1, pp. 35-45.
- Litterman, Robert B. (1986), "Forecasting with Bayesian Vector Autoregressions-five Years of Experience," *Journal of Business & Economic Statistics*, Vol. 4, No. 1, pp. 25-38.
- Prado, Raquel, and Mike West (2010), *Time Series*, Chapman & Hall, London.
- Schorfheide, Frank, and Dongho Song (2015), "Real-time Forecasting with a Mixed-frequency VAR," *Journal of Business & Economic Statistics*, Vol. 33, No. 3, pp. 366-380.
- West, Mike (1996), "Bayesian Time Series: Models and Computations for the Analysis of Time Series in the Physical Sciences," in K. M. Hanson and R. N. Silver (eds.), *Maximum Entropy and Bayesian Methods*, No. 79, Fundamental Theories of Physics, Springer Netherlands, pp. 23-34.
- West, Mike (1997), "Time Series Decomposition," *Biometrika*, Vol. 84, No. 2, January, pp. 489-494.
- West, Mike, and Jeff Harrison (1997), *Bayesian Forecasting and Dynamic Models*, Springer, New York.





# Table of Contents





Preface	91
About the authors	92
Acknowledgements	92
<b>1. Introduction .....</b>	<b>93</b>
<b>2. The Model .....</b>	<b>101</b>
<b>3. Gibbs Sampling .....</b>	<b>107</b>
3.1 The Dynamic Linear Model.....	107
3.1.1 Adapted FFBS Algorithm .....	111
3.2 Bayesian VAR .....	117
3.2.1 Computing the Likelihood Function .....	118
3.2.2 Joint Prior and Posterior .....	119
3.2.3 Specifying the Joint Prior Distribution.....	121
<b>4. Properties of mixed-frequency VAR: simulations .....</b>	<b>122</b>
<b>5. Empirical Analysis .....</b>	<b>128</b>

5.1 Estimating Real GDP at a Monthly Frequency.....	132
5.2 Estimating the Unemployment Rate with Irregularly Spaced Data.....	135
<b>Conclusion .....</b>	<b>138</b>
<b>Annex .....</b>	<b>142</b>
Annex A. Some Results.....	145
Matrix Algebra .....	145
Maximum Likelihood Estimator–Multiv Gaussian Distribution.....	146
Annex B. Proofs .....	149
Proposition 1 .....	149
Proposition 2 .....	151
Proof .....	151
<b>References .....</b>	<b>153</b>

CENTRO DE ESTUDIOS MONETARIOS LATINOAMERICANOS  
Asociación Regional de Bancos Centrales

[www.cemla.org](http://www.cemla.org)

