



UNIVERSIDAD NACIONAL DE COLOMBIA

TRABAJO FINAL DE GRADO - MAESTRÍA EN ESTADÍSTICA

Una metodología de identificación para el modelo factorial dinámico de umbrales

Autor:

Julian Alonso Cardenas Cardenas
C.C.: 1032471488
Email: juacardenasca@unal.edu.co

Director:

PhD. Sergio Calderon

Facultad de Ciencias

22 de julio de 2024

Una metodología de identificación para el modelo factorial dinámico de umbrales

Julian Alonso Cardenas Cardenas

Resumen

El objetivo general de este documento es explorar una metodología para la etapa de identificación del número de regímenes y el número de factores en un modelo factorial dinámico de umbrales. Para lograr este objetivo se hizo uso de las correlaciones canónicas muestrales del proceso de series de tiempo multivariado, junto con las ideas derivadas de una prueba de no linealidad. De esta manera se muestran indicios para identificar tanto el número de factores, como el número de umbrales. Finalmente, la metodología propuesta se aplica sobre conjuntos de datos simulados y reales.

Palabras clave: Modelo factorial, Modelo factorial dinámico de umbrales, Series de tiempo, Modelos no lineales.

An identification methodology for threshold dynamic factor model

Julian Alonso Cardenas Cardenas

Abstract

The main objective of this paper is to explore a methodology for the identification stage of the number of regimes and the number of factors in a dynamic threshold factor model. To achieve this objective, we used the canonical sample correlations of the multivariate time series process, together with insights derived from a nonlinearity test. In this way, we found some evidence to identify the number of factors and the number of thresholds. Finally, the proposed methodology is applied on simulated and real data sets.

Factorial model, Threshold dynamic factor model, Dynamic factors, Non linear models, Time series.

Índice

1. Introducción	6
2. Marco teórico.	7
2.1. Modelo factorial dinámico de umbrales	7
2.2. Elementos teóricos	8
2.2.1. Prueba de Tsay	8
2.2.2. Correlaciones canónicas	10
3. Objetivos.	11
3.1. Objetivo General	11
3.2. Objetivos Específicos	11
4. Metodología	11
5. Aplicación de la metodología	14
5.1. Caso 2 regímenes	14
5.2. Caso 3 regímenes	18
5.3. Simulación de control	23
5.4. Replicaciones del caso 2 regímenes	24
5.5. Casos particulares	26
5.5.1. Gran dimensión	26
6. Aplicación con datos reales	31
7. Conclusiones	37
8. Anexos	39

Índice de figuras

1.	Correlaciones canónicas cuadradas - Primer ejemplo	15
2.	Algoritmo Bry and Boschan (1971) - Valor propio más grande	15
3.	Algoritmo Bry and Boschan (1971) - Segundo valor propio más grande	15
4.	Corr. canónicas - 1er régimen - Primer ejemplo	16
5.	Corr. canónicas - 2do régimen - Primer ejemplo	16
6.	Estimaciones de los procesos factores - Primer ejemplo	17
7.	Correlaciones canónicas - Segundo ejemplo	18
8.	Correlaciones canónicas cuadradas - Primera iteración	20
9.	Correlaciones canónicas cuadradas - Segunda iteración	20
10.	Correlaciones canónicas cuadradas - Tercer umbral	21
11.	Correlaciones canónicas cuadradas - Simulación de control	23
12.	Correlaciones canónicas cuadradas - $r=1000$	25
13.	Corr. canónicas - 1er umbral - $r = 1000$	25
14.	Corr. canónicas - 2do umbral - $r = 1000$	25
15.	Correlaciones canónicas cuadradas - $k = 16$	28
16.	Corr. canónicas - 1er umbral - $k = 16$	29
17.	Corr. canónicas - 2do umbral - $k = 16$	30
18.	Acciones	31
19.	Acciones - Transformaciones	32
20.	Índice S&P 500 y su transformación	32
21.	Diagrama de correlaciones canónicas - $d = 0$	33
22.	Diagrama de correlaciones canónicas - $d = 1$	34
23.	Proceso de factores - Primer régimen	35
24.	Proceso de factores - Segundo régimen	36
25.	Correlaciones canónicas cuadradas - 1er umbral - Segundo ejemplo	39
26.	Correlaciones canónicas cuadradas - 2do umbral - Segundo ejemplo	40
27.	Algoritmo Bry and Boschan (1971) - Valor propio más grande	40
28.	Algoritmo Bry and Boschan (1971) - Segundo valor propio más grande	40
29.	Estimaciones del proceso de factores - Segundo ejemplo	41
30.	Algoritmo Bry and Boschan (1971) - Valor propio más grande - Primera iteración	41
31.	Algoritmo Bry and Boschan (1971) - Segundo valor propio más grande - Primera iteración	41
32.	Algoritmo Bry and Boschan (1971) - Valor propio más grande - Segunda iteración	42
33.	Algoritmo Bry and Boschan (1971) - Segundo valor propio más grande - Segunda iteración	42
34.	Correlaciones canónicas cuadradas - 1er umbral - $j = 3$	42
35.	Correlaciones canónicas cuadradas - 2do umbral - $j = 3$	43
36.	Estimaciones del proceso de factores - $j = 3$	43
37.	Algoritmo de puntos de giro - Primer valor propio	45
38.	Algoritmo de puntos de giro - Segundo valor propio	46
39.	Algoritmo de puntos de giro - Tercer valor propio	46

Índice de cuadros

1.	P-valores de la prueba de Peña and Poncela (2006) - Primer ejemplo	16
2.	P-valores de la prueba de Peña and Poncela (2006) - Segundo ejemplo	18
3.	P-valores de la prueba de Peña and Poncela (2006) - Caso 3 umbrales	22
4.	P-valores de la prueba de Peña and Poncela (2006) en el caso control	24
5.	Número de veces que se rechaza H_0 en la prueba de Peña and Poncela (2006) ($r = 1000$)	24
6.	Resultados de la prueba de Tsay (1998) sobre las replicas	26
7.	P-valores de la prueba de Peña and Poncela (2006) - Aplicación práctica $d = 0$	33
8.	P-valores de la prueba de Peña and Poncela (2006) - Aplicación práctica $d = 1$	34
9.	P-valores de la prueba de Peña and Poncela (2006) - Aplicación práctica - Caso estacionario	36
10.	Acciones del índice S&P 500 usadas	45

1. Introducción

Dentro del análisis de series de tiempo uno de los principales problemas que se presenta es el tratamiento de las series multivariadas. La primera respuesta a este problema consistió en extender las metodologías univariadas al caso multivariado, de esta manera se crearon los análisis a través de vectores ARMA. Este tipo de modelos, conocido como modelos VARMA¹, aunque muy usados en la práctica, carecen de parsimonia y desde este punto de vista, parece más recomendable aplicar un enfoque de reducción de dimensionalidad. Fue así como surgió el análisis de series de tiempo a través del modelo factorial dinámico. El objetivo de la metodología de factores dinámicos es encontrar los movimientos comunes que conservan la dependencia de las series originales aplicando métodos de reducción de dimensionalidad. El uso de esta clase de metodologías se remonta a [Anderson \(1963\)](#), pero quizá, los referentes metodológicos más importantes del modelo factorial dinámico dentro de la literatura son [Stock and Watson \(2002\)](#), quienes reúnen una gran cantidad de resultados sobre estas técnicas dentro de sus diversos trabajos. Referentes importantes del uso del modelo factorial para el desarrollo de este documento son [Stock and Watson \(2011\)](#), y [Pan and Yao \(2008\)](#), por solo nombrar algunos autores.

Por otro lado, también dentro del análisis de series de tiempo se da un problema muy particular, y es que muchos casos prácticos presentan comportamientos determinados por eventos externos a la dinámica de las series de tiempo como anuncios, leyes o situaciones específicas que generan cambios estructurales en el comportamiento de las series de tiempo, de este manera se induce un comportamiento *no lineal* en las series. Existen diversos acercamientos a este tipo de inconvenientes, uno de los más difundidos es el tratamiento de las series por medio de los modelos autoregresivos de tipo umbral (TAR). Estos son modelos en los que se usa un valor de umbral para distinguir rangos de valores o comportamientos de diversa índole en los datos. En este tipo de modelos se presenta no linealidad ya que los cambios de régimen dictaminados por la variable umbral, implican parámetros no constantes en el modelo. El origen de este tipo de modelos se remonta a los desarrollos presentados por [Tong and Lim \(1980\)](#) desde la perspectiva univariada, pero desde el punto de vista multivariado quienes construyeron un esquema general de análisis fueron [Tiao and Tsay \(1989\)](#). Así, surgen los modelos autoregresivos multivariados de tipo umbral (MTAR). Otros desarrollos referentes en las metodologías MTAR pueden encontrarse en [Calderón-Villanueva and Nieto \(2021\)](#), en donde se desarrolla el uso de la metodología en términos de pronóstico.

Teniendo en cuenta todo lo anterior, surge el marco teórico de este documento, el cual se apoya sobre la combinación de las dos metodologías anteriormente expuestas, el modelo factorial dinámico y los modelos autoregresivos de tipo umbral. Estas dos metodologías convergen dentro de la clase de modelos conocida como *modelos factoriales dinámicos de umbrales*, los cuales permiten analizar sistemas de series temporales que presenten comportamientos no lineales del tipo umbral. El objetivo de esta clase de modelos consiste en construir un esquema que permita analizar un modelo multivariado de dimensión menor a la original con la característica de ser no lineal. Estos modelos son útiles para analizar sistemas de series temporales con estructuras de tipo umbral y pueden interpretarse como una reparametrización del modelo TAR vectorial reduciendo significativamente el número de parámetros. El vector de factores comunes se representa mediante diferentes procesos autorregresivos que se activan cuando una variable determinada pasa un umbral dado. Adicionalmente, los regímenes pueden relacionarse con diferentes estados, que pueden ser útiles para estudiar aplicaciones en diversos campos como la economía o la hidrología, lo importante es que se analice una serie de tiempo dada la existencia de fenómenos determinados.

La referencia central para este tipo de modelos y la cual se usa dentro de este documento se presenta en el trabajo de [Correal and Peña \(2008\)](#) donde se introduce el modelo factorial dinámico de umbrales y se propone un método de estimación por medio del algoritmo EM con un procedimiento de búsqueda directa utilizando los algoritmos del filtro y de suavización de Kalman. Sin embargo, la metodología de [Correal and Peña \(2008\)](#) presenta dos dificultades. Por un lado, no se presenta una evidencia de la necesidad del modelo por medio de la identificación de la no linealidad y adicionalmente, no se presenta una herramienta que ayude en la etapa de identificación de los parámetros estructurales (número de factores, número de regímenes y valores umbrales), ya que se hace una búsqueda mediante *grillas* de valores sin conocer de antemano unos posibles candidatos de tales parámetros. Por lo tanto, la motivación principal de este documento es encontrar una herramienta metodológica que asista en esta labor. Para efectos de cumplir con este objetivo, se usa la forma del modelo factorial dinámico presentada en [Peña and Poncela \(2006\)](#) y complementada por [Bolívar et al. \(2021\)](#), donde se desarrollan además

¹Más información puede encontrarse en [Lütkepohl \(2006\)](#)

de una metodología para la identificación, un procedimiento completo para estimar y pronosticar este tipo de modelos usando las *correlaciones canónicas*. La idea central de este documento es integrar estos resultados junto con los planteamientos del modelo MTAR presentados en [Tiao and Tsay \(1989\)](#), para construir una metodología que intente dar luces para la identificación del número de regímenes y el número de factores, con el fin de que la estimación tenga un soporte previo, sin dejar de lado la evaluación de la necesidad del modelo por medio de la identificación de la no linealidad.

Estos problemas han sido abordados recientemente por [Massacci \(2017\)](#), [Liu and Chen \(2020\)](#) y [Wu \(2021\)](#), donde se esbozan pruebas y metodologías para lograr la identificación y la estimación de este tipo de modelos con diversos planteamientos sobre los supuestos en cada caso, la diferencia aquí es que estos métodos exponen sus metodologías siempre y cuando el número de factores comunes es conocido. Adicionalmente, en este trabajo se usan las bondades de las correlaciones canónicas, las cuales aún no han sido exploradas en la labor de la identificación de modelos con las características descritas.

Este documento se estructura de la siguiente manera. La presentación teórica del modelo, el planteamiento de la problemática y las herramientas teóricas usadas dentro de este estudio se muestran en la sección 2. En la sección 3 se presentan los objetivos del desarrollo de este trabajo. En la sección 4 se desarrollan las ideas generales de la metodología. La puesta en práctica de la metodología con ayuda de diversos casos de simulación se presenta en la sección 5. La sección 6 muestra el uso de la metodología desarrollada en un conjunto de datos reales, en particular, se usan datos del índice bursátil estadounidense S&P500 para dar una representación de la metodología. Se sigue con las conclusiones y las referencias respectivas y se finaliza con algunos anexos sobre las aplicaciones presentadas.

2. Marco teórico.

2.1. Modelo factorial dinámico de umbrales

Sea $\{Y_t\}$ una serie temporal de dimensión k . $Y_t = (y_{1t}, \dots, y_{kt})'$. Se asumirá el proceso Y_t de media cero sin pérdida de generalidad. Se dice que $\{Y_t\}$ sigue un modelo factorial dinámico de umbrales con c regímenes de órdenes p_1, p_2, \dots, p_c y variable umbral w_t si:

$$Y_t = \Delta f_t + u_t \tag{1}$$

$$f_t = \sum_{i=1}^{p_j} \phi_i^{(j)} f_{t-i} + \Gamma^{(j)} a_t, \tag{2}$$

teniendo que $w_{t-d} \in (\gamma_{j-1}, \gamma_j]$ con $j = 1, \dots, c$. Así, cada valor de j genera una ecuación para f_t en (2) y tales ecuaciones, las cuales generan la dinámica de cada uno de los c regímenes comentados anteriormente, pueden representarse como las siguientes c ecuaciones:

$$f_t = \begin{cases} \sum_{i=1}^{p_1} \phi_i^{(1)} f_{t-i} + \Gamma^{(1)} a_t & \text{si } w_{t-d} \in (\gamma_0, \gamma_1] \\ \sum_{i=1}^{p_2} \phi_i^{(2)} f_{t-i} + \Gamma^{(2)} a_t & \text{si } w_{t-d} \in (\gamma_1, \gamma_2] \\ \vdots & \\ \sum_{i=1}^{p_c} \phi_i^{(c)} f_{t-i} + \Gamma^{(c)} a_t & \text{si } w_{t-d} \in (\gamma_{c-1}, \gamma_c] \end{cases}$$

donde:

- $\{w_t\}$ es un proceso estocástico unidimensional observable y estacionario.
- $\{f_t\}$ es un proceso no observable de media cero de dimensión $r < k$
- $\{u_t\}$ es ruido blanco de dimensión k con matriz de covarianzas Σ_u diagonal y definida positiva. $\{u_t\}$ es independiente de $\{f_{t-h}\}$ para $h \geq 0$.

- $\{a_t\}$ es ruido blanco de dimensión r con matriz de covarianzas la identidad I_r . $\{a_t\}$ es independiente de $\{f_{t-h}\}$ para $h \geq 1$.
- $\{w_t\}$, $\{u_t\}$ y $\{a_t\}$ deben ser independientes entre sí.

Los parámetros del modelo son: $-\infty = \gamma_0 < \gamma_1 < \dots < \gamma_{c-1} < \gamma_c = \infty$ conocidos como parámetros umbral, d que es un número entero positivo que indica el retardo de la variable umbral, la matriz de carga de los factores Δ de dimensión $k \times r$, $\phi_i^{(j)}$ y $\Gamma^{(j)}$ para $j = 1, \dots, c$; determinan la dinámica del factor según el umbral, con $\Gamma^{(j)}$ matrices diagonales y definidas positivas de dimensión $(r \times r)$.

El modelo indica que se obtienen componentes comunes no lineales debido a que la dinámica de los factores viene determinada por medio del nivel que se presente en la variable umbral. Por lo tanto, se dice que los factores $\{f_t\}$ siguen un modelo MTAR si cumplen la forma de la ecuación (2) donde $\{w_t\}$ puede ser una combinación lineal estacionaria de $\{Y_t\}$ o una variable exógena.

En [Correal and Peña \(2008\)](#) se establecen condiciones para que el modelo definido en (1) y (2) sea identificable. Para empezar, Δ debe cumplir con que $\text{rango}(\Delta) = r$ y $\Delta' \Delta = I_r$, esto es un resultado proveniente de la definición del modelo de factores dinámicos desarrollado en [Peña and Poncela \(2006\)](#). Adicionalmente, las matrices $\Gamma^{(j)}$ deben ser matrices diagonales y definidas positivas para $j = 1, \dots, c$.

Dadas todas la definiciones y las condiciones anteriores, el objetivo consiste en considerar el modelo planteado en (1) y (2), poniendo énfasis en varios puntos en específico. En primera instancia, la identificación de la no linealidad en el proceso de factores $\{f_t\}$, dado que este proceso es no observable. Segundo, la identificación de la dimensión del proceso $\{f_t\}$ y por último, la evaluación del comportamiento de $\{w_t\}$ para postular unos posibles valores de los parámetros umbral γ_j . Para este fin, se van a usar diversas herramientas, que como se podrá observar en el desarrollo del documento, pueden esbozar una idea general de los objetivos planteados. Las posibles metodologías a seguir para este fin se muestran a continuación.

2.2. Elementos teóricos

2.2.1. Prueba de Tsay

La prueba presentada a continuación fue desarrollada por [Tsay \(1998\)](#) para evaluar la hipótesis de no linealidad de tipo umbral sobre una serie de tiempo multivariada. Lo siguiente es un resumen de lo presentado por tales autores. Sea $\{U_t\}$ un proceso estocástico de dimensión k , se dice que $\{U_t\}$ sigue un MTAR con z_t variable de umbrales si:

$$U_t = c_j + \sum_{i=1}^p \phi_i^{(j)} U_{t-i} + \sum_{i=1}^q \beta_i^{(j)} X_{t-i} + \epsilon_t^{(j)}, \tag{3}$$

con $r_{j-1} < z_{t-d} \leq r_j$, con d constante de retardo, c_j vectores constantes, $\phi_i^{(j)}$ matrices de transición que determinan la dinámica de $\{U_t\}$, X_t matriz de variables regresoras junto con $\beta_i^{(j)}$ los parámetros asociados a tales regresores y $\epsilon_t^{(j)}$ el componente idiosincrático de cada serie. Todo lo anterior para $j = 1, \dots, s$. Lo anterior, es una generalización de la ecuación (2), con lo cual, se puede observar que en este caso se generan s ecuaciones. En general, el test consiste en observar el siguiente sistema de hipótesis: $H_0 : \{U_t\}$ es lineal vs la alternativa $\{U_t\}$ sigue un MTAR. Sea $S = \{z_{(1)}, \dots, z_{(n)}\}$ la variable de rezago ordenada de menor a mayor. Asumiendo p, q y d conocidos se define:

$$U_{t(i)+d} = \mathcal{X}_{t(i)+d} \Phi + \epsilon_{t(i)+d}, \tag{4}$$

con $t = h+1, \dots, n$, $h = \max(p, q, d)$, $\mathcal{X}'_t = (\mathbf{1}, U'_{t-1}, \dots, U'_{t-p}, X'_{t-1}, \dots, X'_{t-q})$ y $t(i)$ la t -ésima observación ordenada respecto a la variable umbral z_{t-d} ordenada de menor a mayor. Lo anterior para $i = 1, \dots, n-h$. Se hace una regresión donde solo se cambia el orden en que cada observación de $\{U_t\}$ entra en la regresión pero la dinámica se mantiene igual. Este concepto es de gran relevancia en este documento, debido a que es la idea

sobre la que se fundamenta la metodología planteada más adelante.

Sea $\Phi_{(m)}$ la estimación de mínimos cuadrados de Φ con los primeros m valores ordenados de z_{t-d} . Se definen los residuos de predicción ϵ_t y los residuos de predicción estandarizados η_t como:

$$\epsilon_{t(m+1)+d} = Y_{t(m+1)+d} - \Phi_{(m)} \mathcal{X}_{t(m+1)+d}, \quad (5)$$

$$\eta_{t(m+1)+d} = \epsilon_{t(m+1)+d} / (1 + \mathcal{X}'_{t(m+1)+d} V_m \mathcal{X}_{t(m+1)+d}), \quad (6)$$

donde:

$$V_m = \left[\sum_{i=1}^m \mathcal{X}_{t(i)+d} \mathcal{X}'_{t(i)+d} \right]^{-1},$$

Lo anterior dentro de la regresión dada por:

$$\eta_{t(l)+d} = \mathcal{X}_{t(l)+d} \varphi + w_{t(l)+d}. \quad (7)$$

Con $l = m_0, \dots, n-h$. La hipótesis $H_0 : \varphi = 0$ se prueba con la estadística $C(d) \sim \mathcal{X}_{k(pk+qv+1)}^2$ con $C(d)$:

$$C(d) = (n - h - m_0 - (kp + vq + 1)) (\log(|S_0|) - \log(|S_1|)), \quad (8)$$

$$S_0 = \frac{1}{n - h - m_0} \sum_{m_0+1}^{n-h} \eta'_{t(l)+d} \eta_{t(l)+d}, \quad (9)$$

$$S_1 = \frac{1}{n - h - m_0} \sum_{m_0+1}^{n-h} w'_{t(l)+d} w_{t(l)+d}. \quad (10)$$

Todo el procedimiento anterior estructura una prueba para evaluar no linealidad de tipo umbral sobre procesos que sigan lo planteado en la ecuación (3). Se puede observar, fácilmente, que la ecuación (3) es una generalización de la ecuación (2) y si retomamos el planteamiento general del modelo en (1) y (2), se puede observar que $\{f_t\}$, es decir, el proceso de factores, es el que sigue la dinámica del modelo MTAR planteada en (3). De tal manera que la aplicación de la prueba desarrollada por [Tsay \(1998\)](#) debería evaluarse sobre $\{f_t\}$. El inconveniente que surge aquí es que $\{f_t\}$ es no observable, por lo tanto se presenta la necesidad de encontrar una estimación de $\{f_t\}$ para comprobar que el proceso sí sea no lineal de tipo umbral por medio de tal prueba. Debido a que el foco del documento es hacer énfasis en la etapa de identificación, el objetivo es buscar una estimación previa para $\{f_t\}$ para proceder con la aplicación de la prueba.

Para cumplir con este objetivo, se usa el supuesto planteado anteriormente que dictamina que: $\Delta' \Delta = I$. Adicionalmente, se define nuevamente que dado $w_{t-d} \in (\gamma_{j-1}, \gamma_j]$, la variable de umbrales al igual que en (2), con $j = 1, \dots, c$. Con estos hechos se tiene que, para cualquier régimen j :

$$\Delta' Y_t = f_t + \Delta' U_t \Leftrightarrow f_t = \Delta' Y_t - \Delta' U_t, \quad \forall t.$$

Además:

$$\Delta' Y_t = \sum_{i=1}^{p_j} \phi_i^{(j)} f_{t-i} + \Gamma^{(j)} a_t + \Delta' U_t,$$

$$\Delta' Y_t = \sum_{i=1}^{p_j} \phi_i^{(j)} [\Delta' Y_{t-i} - \Delta' U_{t-i}] + \Gamma^{(j)} a_t + \Delta' U_t,$$

$$\Delta' Y_t = \sum_{i=1}^{p_j} \phi_i^{(j)} \Delta' Y_{t-i} - \sum_{i=1}^{p_j} \phi_i^{(j)} \Delta' U_{t-i} + \Gamma^{(j)} a_t + \Delta' U_t,$$

Si se toma como $Q_t = \Delta'Y_t$ y $V_t = \Delta'U_t$ se tiene que, para el j -ésimo régimen:

$$Q_t = \sum_{i=1}^{P_i} \phi_i^{(j)} Q_{t-i} + \underbrace{\Gamma^{(j)} a_t + V_t - \sum_{i=1}^{P_i} \phi_i^{(j)} V_{t-i}}_{\sigma_t^{(j)}}, \quad (11)$$

$$Q_t = \sum_{i=1}^{P_i} \phi_i^{(j)} Q_{t-i} + \sigma_t^{(j)}. \quad (12)$$

En los supuestos del modelo planteado en (3) $\{a_t\}$ y $\{u_t\}$ son ruidos blancos y son independientes. Bajo estos supuestos la expresión $\sigma_t^{(j)}$ debería ser ruido blanco y depende de $j = 1, \dots, c$. Con estos hechos, se cumplen las condiciones de la ecuación (3). Por tanto, $\{Q_t\} = \{\Delta'Y_t\} \sim \text{MTAR}$. A su vez, se afirma en [Peña and Poncela \(2006\)](#) que $\{Q_t\}$ es una estimación *previa* del proceso de factores subyacente. Por lo tanto, este hecho provee las bases para la aplicación de esta prueba dentro del contexto de este documento.

2.2.2. Correlaciones canónicas

La referencia central de este documento para el tratamiento del modelo dinámico factorial son las ideas planteadas en [Peña and Poncela \(2006\)](#). En esta subsección se resumen las herramientas que usan estos autores para la etapa de identificación del modelo factorial dinámico, en particular, la definición y el uso de las correlaciones canónicas para identificar el número de factores comunes. De esta manera, se comienza definiendo el modelo de factores dinámicos como lo definen en [Peña and Poncela \(2006\)](#). Sea $\{Y_t\}$ una serie de tiempo multivariada de dimensión m , se dice que sigue un modelo factorial dinámico si:

$$Y_t = P f_t + e_t \quad (13)$$

Donde $\{f_t\}$ es el vector r -dimensional de factores comunes, P es la matriz de carga de los factores de dimensión $m \times r$ y $\{e_t\}$ un proceso de ruido blanco con distribución normal de media vector de 0 y matriz de varianzas diagonal Σ_e . Se observa que este proceso es similar al definido en la ecuación (1), sin embargo, para este caso, el modelo que se asume para el proceso $\{f_t\}$ es un modelo $VARMA(p, q)$ ².

Para la etapa de identificación del modelo planteado, [Peña and Poncela \(2006\)](#) plantean el uso de las *correlaciones canónicas*. Esta metodología se remonta a los trabajos de [Hotelling \(1935\)](#) y en resumen, la correlación canónica busca que para dos vectores aleatorios X y Y de dimensiones p y q se encuentren combinaciones lineales de tipo:

$$a_1 x_1 + \dots + a_p x_p = U, b_1 y_1 + \dots + a_q y_q = V,$$

Tal que la correlación entre U y V sea máxima para cualesquiera constantes a_i y b_j . Por construcción, pueden existir más de un par de variables canónicas como U y V que cumplan con la definición. En total, existen tantas parejas de variables canónicas como el mínimo entre p y q . Más detalles sobre la metodología de correlaciones canónicas pueden encontrarse en [Monroy \(2007\)](#).

Esta metodología es importante, ya que [Peña and Poncela \(2006\)](#) demuestran que para el modelo dinámico factorial definido en (13) y bajo los supuestos anteriormente planteados, el número de factores comunes, r , es equivalente al número de correlaciones canónicas diferentes de 0 entre Y_t y Y_{t-k} . Para demostrar esta hipótesis definen la matriz de dimensión $m \times m$:

$$M(k) = [E(Y_t Y_t')]^{-1} E(Y_t Y_{t-k}') [E(Y_{t-k} Y_{t-k}')]^{-1} E(Y_{t-k} Y_t') \quad (14)$$

²Más detalles sobre el modelo $VARMA$ puede encontrarse en [Lütkepohl \(1993\)](#)

Se demuestra que el número de correlaciones canónicas entre Y_{t-k} y Y_t iguales a 0, es igual al número de valores propios iguales a 0 de la matriz $M(k)$, y dado que el $\text{rango}(M(k)) = r$, este número es $m - r$. Así las cosas, el número de factores comunes es r y este es equivalente a las correlaciones canónicas diferentes de 0 entre Y_{t-k} y Y_t . Para el caso de muestra finita, sea T la longitud de la serie, las correlaciones canónicas cuadradas muestrales entre Y_t y Y_{t-k} son los valores propios de la matriz $\hat{M}_1(k)$:

$$\hat{M}_1(k) = \left[\sum_{t=k+1}^T (Y_t Y_t') \right]^{-1} \sum_{t=k+1}^T (Y_t Y_{t-k}') \left[\sum_{t=k+1}^T (Y_{t-k} Y_{t-k}') \right]^{-1} \sum_{t=k+1}^T (Y_{t-k} Y_t') \quad (15)$$

Se demuestra adicionalmente que sin importar si Y_t es estacionario o no, cuando T tiende a infinito, $\hat{M}_1(k)$ tiene $m - r$ valores propios que convergen en probabilidad a 0. En consecuencia, r es la dimensión del proceso de factores. De igual manera, al considerar los valores propios de $\hat{M}_1(k)$ como, $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \dots \leq \hat{\lambda}_m$, se muestra que el estadístico:

$$S_{m-r} = -(T - k) \sum_{j=1}^{m-r} \log(1 - \hat{\lambda}_j), \quad (16)$$

se distribuye asintóticamente como una distribución $\chi_{(m-r)2}^2$.

Adicionalmente, los autores establecen que los vectores propios asociados a los r valores propios más grandes de la matriz $\hat{M}_1(k)$, pueden considerarse como una estimación *previa* de la matriz de pesos de los factores P . Con todas las herramientas teóricas ya definidas, la idea general es diseñar una metodología que pueda llegar a traer información que permita cumplir con los objetivos que se describen a continuación.

3. Objetivos.

3.1. Objetivo General

Explorar una metodología para la etapa de identificación del número de regímenes y el número de factores en un modelo factorial dinámico de umbrales.

3.2. Objetivos Específicos

- Proponer una metodología para la identificación del número de regímenes para el modelo factorial dinámico de umbrales.
- Usar las autocorrelaciones muestrales del proceso junto con las ideas de la prueba de no linealidad de Tsay para identificar el número de factores.
- Evaluar el funcionamiento de una modificación de la prueba de no linealidad de Tsay en el modelo factorial dinámico de umbrales.
- Ilustrar la metodología propuesta a un conjunto de datos reales o simulados.

4. Metodología

Los objetivos plantean que se debe explorar una metodología para identificar el número de factores y de regímenes dentro del modelo planteado en las ecuaciones (1) y (2). Como se ha mencionado anteriormente, la metodología se basa en la combinación de dos ideas. La primera de ellas es la propuesta por Tsay (1998), en la cual se construye una regresión donde solo se cambia el orden en que cada observación entra en una regresión secuencial, pero la dinámica se mantiene igual; de esta manera, formulan una prueba para evaluar la hipótesis de no linealidad sobre un modelo MTAR. Dado que en el modelo presentado en (1) y (2), es $\{f_t\}$ el proceso que teóricamente sigue un MTAR, la aplicación de la prueba desarrollada en Tsay (1998) no es posible, ya que

este es no observable. Por otro lado, en [Peña and Poncela \(2006\)](#), se establece que el cálculo de las correlaciones canónicas por medio de la matriz $\hat{M}_1(k)$, da indicios del comportamiento del proceso de factores no observable $\{f_t\}$ en un modelo de factores dinámicos, en particular, del número de factores comunes del proceso observado $\{Y_t\}$. De acuerdo a lo anterior, y retomando el modelo planteado en (1) y (2), la idea que se plantea aquí es evaluar las correlaciones canónicas del proceso $\{Y_t\}$ en un modelo factorial dinámico de umbrales, y estudiar si al igual que en el caso del modelo de factores dinámicos, esta métrica contiene información sobre el proceso $\{f_t\}$.

Sin embargo, dentro del modelo factorial dinámico de umbrales $\{f_t\}$ es no lineal de tipo umbral, por lo tanto, con el fin de tener en cuenta esta característica dentro de la metodología, se integra la idea de [Tsay \(1998\)](#), la cual consiste en evidenciar posibles cambios en la estructura del modelo, por medio una métrica que tenga en cuenta a los umbrales. En el caso de [Tsay \(1998\)](#), se evalúan los residuos de una regresión ordenada, mientras que en este trabajo, se evalúan los cambios de estructura del modelo dentro del cálculo de las correlaciones canónicas, es decir, ordenar las observaciones respecto a la variable de umbrales y secuencialmente, computar la matriz $\hat{M}_1(k)$ y obtener sus valores propios. Se espera que la dinámica de tales valores propios muestren un cambio de estructura del modelo, y adicionalmente, den idea del número de factores del modelo, tal y como se plantea en los objetivos de este documento.

Dado todo lo anterior, se propone la siguiente metodología, donde el proceso en general se puede resumir en los siguientes pasos:

- **PASO 1:** Se ordena de menor a mayor la dupla de observaciones Y_t y Y_{t-k} , respecto a la variable umbral w_{t-d} . Lo anterior para algún retardo d de la variables umbral y para algún k . Este orden es importante ya que se busca mantener la dinámica temporal, pero evidenciar respecto al orden, los posibles cambios en la estructura del modelo dados los regímenes. Esta es la idea que proviene de la estructura de la prueba de [Tsay \(1998\)](#). Este paso puede ser iterativo para hallar valores más viables d y k , pero como en [Peña and Poncela \(2006\)](#), se sugiere encontrar k por criterios de información a través de un ajuste de un modelo VAR. Mientras que w_{t-d} se va a asumir como conocido, el cual es un supuesto usual dentro de este tipo de modelos.
- **PASO 2:** Se construye la matriz $\hat{M}_1(k)_{n_l}$, como lo muestra la Ecuación (15), para la dupla de observaciones ordenadas Y_t y Y_{t-k} , donde n_l para algún l^3 es el número de observaciones ordenadas por w_{t-d} con las que se construye $\hat{M}_1(k)_{n_l}$. Con este punto n_l definido, se empiezan a agregar datos uno a uno hasta llegar a T , construyendo en cada paso la matriz $\hat{M}_1(k)_{n_l}$, y extrayendo en cada uno de estos pasos los valores propios de tal matriz. De esta manera, se obtienen las correlaciones canónicas cuadradas muestrales (λ_i) para la dupla de observaciones ordenadas Y_t y Y_{t-k} .
- **PASO 3:** Para cada valor propio λ_i con $i = 1, \dots, m$, de cada una de las matrices $\hat{M}_1(k)_{n_l}$ generadas, se debe construir un gráfico donde se observe en el eje x la cantidad de observaciones ordenadas respecto a la variable umbral (n_l) y en el eje y la evolución de cada uno de los m valores propios de $\hat{M}_1(k)_{n_l}$. Si las observaciones siguen la dinámica de las ecuaciones (1) y (2), es decir, provienen de un proceso que se puede representar como un *modelo factorial dinámico de umbrales*, la gráfica propuesta para los λ_i debe hacer evidente algún cambio en la estructura del modelo, debido a que, si el nivel de w_{t-d} define una no linealidad de tipo umbral sobre las observaciones, los niveles de λ_i deben mostrar algún tipo de quiebre o cambio de dinámica debido a la existencia de los regímenes.
- **PASO 4:** Para asistir la identificación de un punto de quiebre específico dentro de la gráfica comentada, se utiliza el algoritmo de [Bry and Boschan \(1971\)](#)⁴, el cual es una herramienta econométrica para encontrar *puntos de quiebre* en una serie de tiempo. Para usar este algoritmo, se asume que el desarrollo de los λ_i componen una serie de tiempo indexada por el orden n_l hasta llegar a T , de esta manera, al aplicar el algoritmo, se encuentran *puntos candidatos* donde posiblemente exista el cambio de dinámica ya comentado y donde, según el objetivo de la metodología, se establece la aparición de un nuevo régimen. El punto candidato a ser escogido como el cambio de régimen será el punto en el tiempo más cercano al punto de cambio en la evolución identificada en cada diagrama de valores propios.

³Gracias a diversas pruebas se establece este punto inicial como un 20 % de la muestra. Este porcentaje asegura tener *suficientes* observaciones para mostrar algún comportamiento que se pueda ir generalizando, aunque este número depende evidentemente del tamaño de muestra T y de la cantidad de observaciones en cada umbral.

⁴Una explicación más explícita de este algoritmo se puede encontrar en los anexos.

- **PASO 5:** Las observaciones que estén hasta antes del punto de corte identificado, forman un primer grupo de observaciones y a su vez, se asumen como pertenecientes a un primer régimen. Estas observaciones son removidas de la muestra y se aplica la metodología del **PASO 1** al **PASO 5**, hasta que el diagrama de los valores propios λ_i para las observaciones restantes sea estable en su desarrollo, es decir, el comportamiento de estos valores no haga evidente algún cambio de estructura que se asuma pueda provenir de una no linealidad de tipo umbral. El número de regímenes identificado corresponde al número de grupos de observaciones que se construyan bajo estas iteraciones.
- **PASO 6:** Con el número de regímenes ya identificado, se usan las observaciones pertenecientes a cada uno de estos regímenes y se calcula la matriz $\hat{M}_1(k)$ como en el **PASO 2** y el **PASO 3** dentro de cada régimen definido. Con los valores propios de tal matriz, se identifica el número de factores r a través de la aplicación de la prueba de [Peña and Poncela \(2006\)](#) para cada grupo de observaciones. Esta prueba es viable en este paso, ya que dentro de cada régimen, la dinámica se puede asumir como lineal.
- **PASO 7:** Al extraer los vectores propios que están asociados a los r valores propios ya identificados como diferentes de 0, se puede construir para cada grupo, una estimación preliminar de la matriz de carga de los factores Δ_j con $j = 1, \dots, c$. Con cada estimación de Δ , se construye una estimación de f_t por medio de la relación $\hat{f}_t^{(j)} = \Delta_j' Y_t$ que se establece desde los supuestos del modelo.
- **PASO 8:** Con cada una de estas estimaciones, se puede observar la hipótesis de no linealidad en $\hat{f}_t^{(j)}$ haciendo uso de la prueba de [Tsay \(1998\)](#). Este paso complementa la metodología y dota al proceso de validez ya que se justifica la no linealidad del proceso de factores mediante su estimación, ya que en su estado de la naturaleza este es no observable.

Bajo la estructura anterior, la siguiente sección de este documento ilustra diversos trabajos de simulación que evidencian la información que tienen las correlaciones canónicas para la identificación del modelo, haciendo evidente que la dinámica mostrada en la evolución de estos valores puede traer ideas sobre el número de factores pero además, sobre los puntos posibles en donde la variable de umbrales cambia el comportamiento del proceso de factores lo que lleva a identificar el número de regímenes.

5. Aplicación de la metodología

5.1. Caso 2 regímenes

Para introducir el proceso metodológico de la sección anterior, se muestra un ejemplo particular bajo un modelo que sigue la estructura de la sección 2.1, en particular de las ecuaciones (1) y (2). En primer lugar se genera f_t de la siguiente manera:

- Se generan **2** factores comunes que provienen de un modelo MTAR con variable de rezago $w_t = \frac{1}{2}w_{t-1} + \eta_t$, es decir, un modelo AR con parámetro $\frac{1}{2}$, el proceso de ruido blanco $\{\eta_t\}$ se simula como una distribución normal estándar.

- La estructura del modelo MTAR se establece como. Si $w_{t-1} < 0$

$$\phi_1^{(1)} = \begin{pmatrix} 0.8 & 0 \\ 0 & -0.9 \end{pmatrix},$$

- Si $w_{t-1} \geq 0$

$$\phi_1^{(2)} = \begin{pmatrix} -0.8 & 0 \\ 0 & 0.9 \end{pmatrix},$$

- Las matrices de varianzas se establecen como matrices identidad para ruido en los dos regímenes $\Gamma_{(j)}$ para $j = 1, 2$.

Con f_t ya generado, se puede obtener Y_t tal que la matriz de carga de los factores genera un proceso de dimensión 5:

$$\Delta = \begin{pmatrix} 0.29 & 0.94 \\ 0.54 & -0.05 \\ 0.34 & -0.06 \\ 0.47 & -0.23 \\ 0.52 & -0.22 \end{pmatrix},$$

Para finalizar la simulación, el proceso de ruido $\{u_t\}$ es generado por una distribución normal multivariada estándar. La simulación se construye con un periodo de inicialización de $n_0 = 100$ en la construcción de los factores y se simulan un total de $T = 2000$ observaciones que corresponden a 1000 para cada uno de los regímenes. La metodología se comienza a aplicar al ordenar la pareja de observaciones Y_t y Y_{t-1} respecto al proceso w_{t-1} como se indica en el **PASO 1** de la metodología. Siguiendo con el **PASO 2**, se puede comenzar a construir $\hat{M}_1(k)_{n_l}$ con $l = 400$ que es un 20% del total de observaciones. Punto a punto se agrega una nueva observación ordenada por w_{t-1} y se hace el cálculo de $\hat{M}_1(k)_{n_l}$ extrayendo los valores propios en cada uno de estos puntos. En particular, la evolución de los valores propios de las matrices $\hat{M}_1(k)_{n_l}$ se muestra en la Figura 1. En esta grafica, el eje x contiene una a una la cantidad de observaciones ordenadas con las que se calcula $\hat{M}_1(k)_{n_l}$ y el eje y , los valores propios de estas matrices $\hat{M}_1(k)_{n_l}$, cada recuadro corresponde a un valor propio de $\hat{M}_1(k)$ ordenados de manera descendente. Esta gráfica es la que se establece en el **PASO 3** de la metodología.

Tal como se propone en el **PASO 3** y **PASO 4** de la metodología, se procede a reconocer una no linealidad de tipo umbral, debido a un cambio de quiebre en el desarrollo de los valores propios de la Figura 1. Al observar esta gráfica, es evidente que en los dos valores propios más grandes existe un marcado cambio de dinámica alrededor de la observación ordenada n_{1000} , lo que da indicios de la no linealidad bajo la hipótesis de un cambio de estructura en el modelo y que va acorde con la construcción de la simulación planteada. Este punto se refuerza con el algoritmo de [Bry and Boschan \(1971\)](#), el cual es aplicado a la serie de valores propios como se indica en el **PASO 4**. Los puntos candidatos como cambio de régimen para las gráficas de los dos primeros valores propios son mostradas en las Figuras 2 y 3, las líneas verticales rojas en el desarrollo de la serie de tiempo cortan el punto en el cual se tiene un candidato a punto de quiebre, el valor más cercano al cambio de dinámica será el valor escogido como punto de corte para el cambio de régimen. Para este caso, el valor más cercano al cambio de tendencia se da en n_{1052} en el primer diagrama y n_{1082} en el segundo diagrama de la Figura 1, mientras que,

la línea vertical en cada recuadro de la Figura 1 es el punto candidato escogido, el cual es el n_{1052} que es el más cercano al valor real simulado, solo por propósitos de este ejemplo.

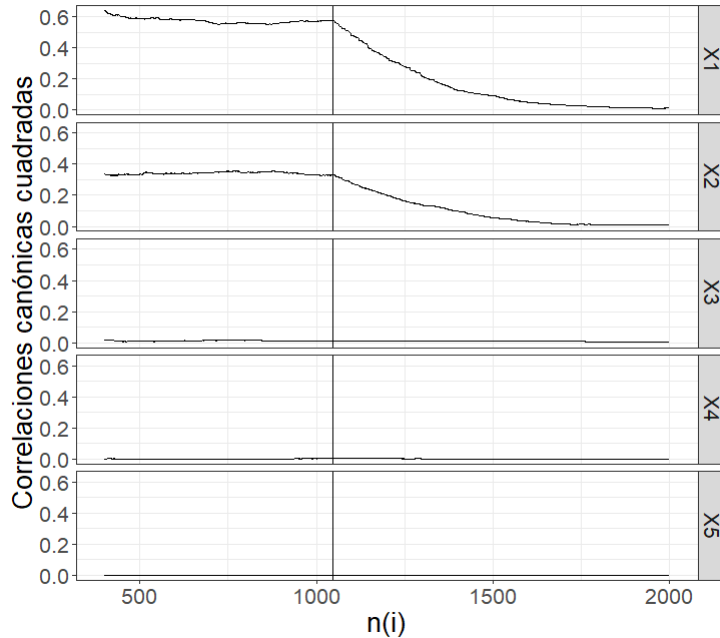


Figura 1: Correlaciones canónicas cuadradas - Primer ejemplo

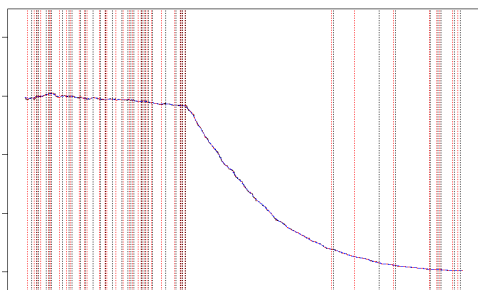


Figura 2: Algoritmo Bry and Boschan (1971) - Valor propio más grande

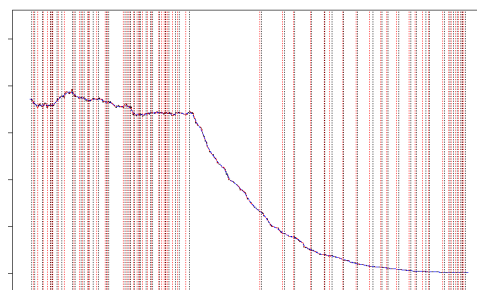


Figura 3: Algoritmo Bry and Boschan (1971) - Segundo valor propio más grande

Con este índice de observaciones ordenadas (n_{1052}), se puede obtener una *estimación preliminar* del corte de la variable umbral como $\hat{\gamma}_1 = -0.09$. De esta manera, siguiendo el **PASO 5** de la metodología, todas las observaciones que estén antes de este punto de corte identificado se asumen que pertenecen a un primer régimen, mientras que, las observaciones restantes pertenecen a un posible segundo régimen. Por lo tanto, se repite el cálculo de la matriz $\hat{M}_1(k)$ pero esta vez para cada uno de los dos grupos de observaciones generados por el punto de corte, y que se asumen corresponden a los dos regímenes tal como se expresa en el **PASO 5** de la metodología. Los valores propios de este nuevo cálculo se muestran en las Figuras 4 y 5. Se puede observar en estos gráficos que, el comportamiento de los 5 valores propios es estable durante todo el periodo y no tiene cambios bruscos, por lo cual, se puede suponer que la dinámica es lineal dentro de cada régimen definido en las Figuras 4 y 5 y por tanto, se puede asumir que se tienen **2** regímenes. Con esto se completa el **PASO 5** de la metodología.

Reforzando lo evidenciado en la Figura 1, en las Figuras 4 y 5 se muestra que existen solo dos valores propios sistemáticamente diferentes de 0 en la evolución de los gráficos. Esto se comprueba al poner en práctica

el **PASO 6** de la metodología y aplicar la prueba de [Peña and Poncela \(2006\)](#). Los resultados se pueden ver en el Cuadro 1, en donde se evidencia que en los dos casos no se rechaza la hipótesis nula de 2 factores comunes, con lo que se puede afirmar que existen, desde cualquiera de los dos regímenes, al menos dos factores dinámicos.

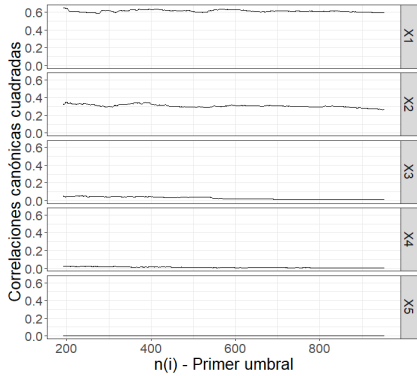


Figura 4: Corr. canónicas - 1er régimen - Primer ejemplo

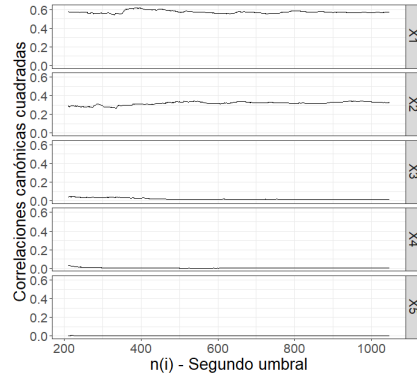


Figura 5: Corr. canónicas - 2do régimen - Primer ejemplo

H_0	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$
Primer umbral	0	0	0.19	0.78	0.77
Segundo umbral	0	0	0.31	0.65	0.80

Cuadro 1: P-valores de la prueba de [Peña and Poncela \(2006\)](#) - Primer ejemplo

Ahora, siguiendo con el **PASO 7** de la metodología, se construyen estimaciones de Δ con el fin de construir también una estimaciones de f_t por medio de la relación $\hat{f}_t = \hat{\Delta}' Y_t$. Ahora, tomando como columnas de $\hat{\Delta}$ los vectores propios asociados a los dos valores propios más grandes del cálculo de $\hat{M}_1(k)$ con todas las observaciones en cada régimen, se obtienen las siguientes matrices de carga estimadas:

$$\hat{\Delta}_1 = \begin{pmatrix} -0.92 & 0.33 \\ 0.13 & 0.58 \\ 0.12 & 0.25 \\ 0.18 & 0.50 \\ 0.26 & 0.47 \end{pmatrix} \quad \hat{\Delta}_2 = \begin{pmatrix} 0.98 & -0.02 \\ -0.12 & 0.53 \\ -0.04 & 0.47 \\ 0.11 & 0.40 \\ -0.01 & 0.57 \end{pmatrix},$$

Se puede evidenciar que, estas matrices son una rotación de la matriz Δ planteada para la construcción de este ejemplo. Por lo tanto, se construyen procesos de factores con estas matrices. Estos procesos se muestran en la Figura 6, en azul se presenta la estimación correspondiente al primer régimen y en rojo al segundo. Teniendo ya los procesos $\hat{f}_t^{(1)}$ y $\hat{f}_t^{(2)}$, se puede hacer uso de la prueba de [Tsay \(1998\)](#) como se indica en el **PASO 8**. Al aplicar esta prueba sobre los dos procesos generados, se obtienen p-valores muy cercanos a 0, con lo cual se rechaza la hipótesis de linealidad en el proceso de factores resultante. Esto culmina con el proceso metodológico planteado, obteniendo una identificación del número de regímenes (2) y el número de factores (2), además del punto de corte de la variable de umbrales ($\hat{\gamma}_1 = -0.09$).

Como se ha dicho, el paso de estimación no hace parte de los objetivos centrales de este documento, aún así, se evidencia que los vectores propios de $\hat{M}_1(k)$ pueden dar una estimación *cercana* de la matriz de pesos, teniendo en cuenta la identificación del número de umbrales y el número de factores que se mostró para este caso específico.

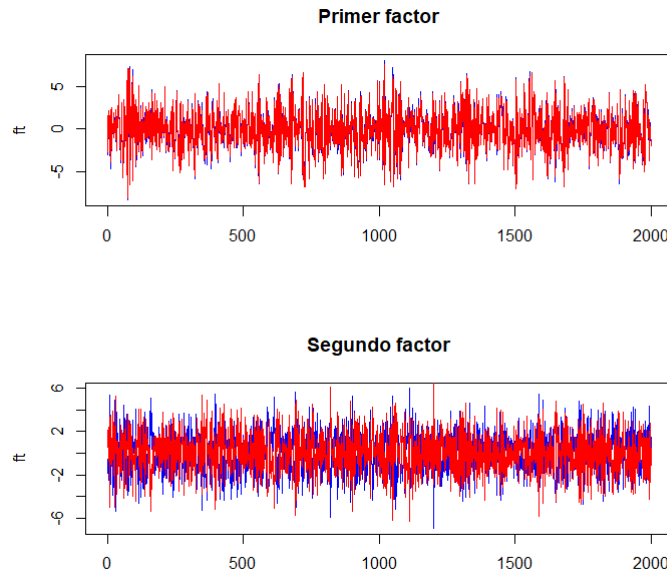


Figura 6: Estimaciones de los procesos factores - Primer ejemplo

Para extender los resultados mostrados se presenta una variación de las *matrices de transición* de los factores, tal que los regímenes no definan un cambio de signo, si no un cambio de magnitud en estos valores. Aunque en la práctica, este no sea un caso tan frecuente, debido a que es muy posible que no se logre identificar plenamente, se pone en práctica para observar la capacidad de la metodología aún en casos tan sensibles y que podrían tener algún problema en su identificación. Por lo tanto, para este ejemplo se establece que, si $w_{t-1} < 0$

$$\phi_1^{(1)} = \begin{pmatrix} 0.9 & 0 \\ 0 & -0.4 \end{pmatrix},$$

Si $w_{t-1} \geq 0$

$$\phi_1^{(2)} = \begin{pmatrix} 0.4 & 0 \\ 0 & -0.9 \end{pmatrix},$$

De nuevo, las matrices de varianzas se establecen como matrices identidad para el ruido en los dos regímenes. De igual manera, Γ_j son matrices identidad para $j = 1, 2$ y las demás condiciones se mantienen iguales al primer caso, esto incluye a la matriz de pesos del factor que se genera también para un proceso de dimensión 5 y el número total de observaciones en la muestra y por régimen.

Los 3 primeros pasos de la metodología se resumen en la Figura 7, en donde se observa la construcción del diagrama de valores propios de la matriz $\hat{M}_1(k)$ para este caso. En esta gráfica, se puede ver como se va dando poco a poco un cambio de tendencia, pero a pesar de que no sea tan evidente como en el caso anterior, sí se puede observar que a medida que entran nuevas observaciones parece definirse un cambio de pendiente al comportamiento inicial, esto es para los dos valores propios más grandes. Hay que remarcar, que en el ejemplo #1 hay una diferencia marcada de signo y magnitud de los parámetros de las matrices de transición del modelo de factores, y por tanto, se puede observar un corte muy claro en la evolución de estos valores propios respecto al número de observaciones. Pero para este caso, fue difícil identificar tal punto de corte, sin embargo, este valor fue escogido como uno de los candidatos al aplicar el algoritmo de [Bry and Boschan \(1971\)](#), en particular este valor fue n_{1008} con un valor de la variable umbral $\gamma_1 = -0.003$, y se muestra en la línea vertical en cada diagrama para cada valor propio de la Figura 7. Las gráficas de la aplicación del algoritmo pueden ser encontradas en los anexos⁵ y se puede observar una gran cantidad de candidatos dada la naturaleza del ejemplo y la descripción que se hizo

⁵Figura 27 en los anexos.

de las limitaciones en esta simulación particular. Aún a pesar de todo lo comentado, sí se muestra en la Figura 7, como en los dos primeros casos existe este cambio de comportamiento buscado en la metodología propuesta.

Adicionalmente, al hacer la clasificación de las observaciones dado el punto de corte anteriormente mencionado, se construyen las matrices $\hat{M}_1(k)$ para cada régimen⁶, al igual que el caso anterior, se observa que ya no existe ningún cambio de dinámica para ninguno de los valores propios en los dos regímenes, por lo tanto, se usa la prueba de Peña and Poncela (2006) y se aplica dentro de cada régimen (Tabla 2). La prueba valida que existen al menos dos factores comunes en cada caso. Las conclusiones finales son similares al primer ejemplo. Primero, las pruebas de Tsay (1998) muestran que las estimaciones $\{f_t^{(j)}\}$ para $j = 1, 2$ son no lineales de tipo umbral. Segundo, las matrices de carga $\hat{\Delta}_1$ y $\hat{\Delta}_2$ son de nuevo rotaciones de la matriz Δ original y por último, se obtiene una identificación plena del número de regímenes, del número de factores y una estimación preliminar del punto de corte de la variable de umbrales. Las estimaciones de los procesos $\{f_t^{(j)}\}$ ⁷ para $j = 1, 2$ y las matrices de carga generadas a través de los vectores propios de $\hat{M}_1(k)$ también se presentan en los anexos.

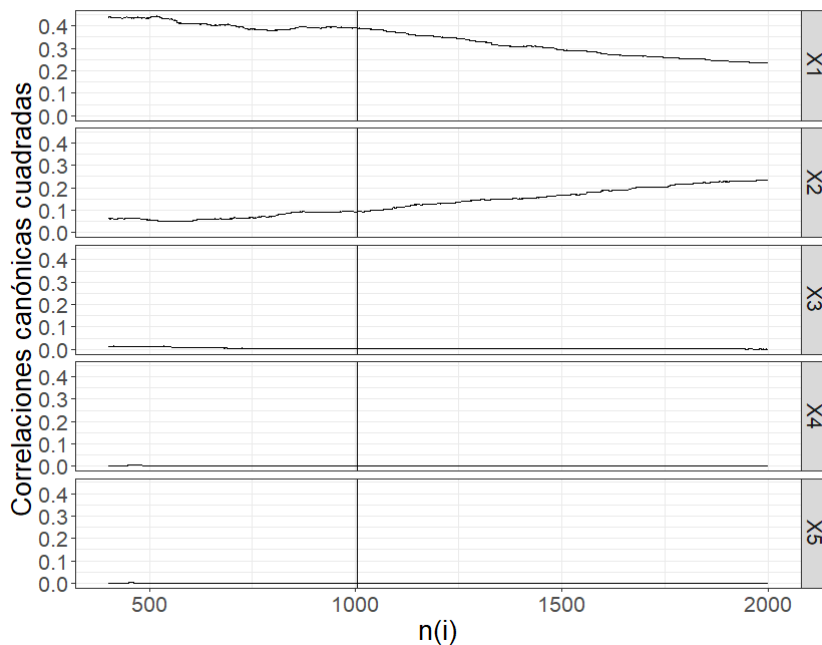


Figura 7: Correlaciones canónicas - Segundo ejemplo

H_0	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$
Primer umbral	0	0	0.56	0.97	0.72
Segundo umbral	0	0	0.59	0.24	0.35

Cuadro 2: P-valores de la prueba de Peña and Poncela (2006) - Segundo ejemplo

El siguiente caso trata de generalizar la cantidad de umbrales para mostrar como funcionaria la metodología cuando se tienen más de 2 umbrales y da luces del uso de la metodología de una manera más amplia.

5.2. Caso 3 regímenes

Esta sección presenta la metodología para un caso en el cual se muestran más de dos regímenes. Con esta particularidad, la metodología varía un poco respecto a los casos presentados anteriormente, específicamente

⁶Figuras 25 y 26 de los anexos.

⁷Figura 29 de los anexos.

en el paso **5** de la metodología propuesta. En tal paso se establece que cuando se identifica un punto de corte en el diagrama de valores propios, se clasifican las observaciones antes de este punto de corte y se asumen que pertenecen a un primer régimen. Para las observaciones restantes, se vuelven a repetir los pasos **1** a **5** hasta que el último subconjunto de observaciones restantes muestre un comportamiento estable en los valores propios de la matriz $\hat{M}_1(k)$. Este proceso es fácilmente generalizable, por lo tanto, se muestra únicamente para el caso de 3 regímenes. Así las cosas, la simulación de este caso tiene las siguientes características:

- Se generan **2** factores comunes que provienen de un modelo MTAR con variable de rezago $w_t = \frac{1}{2}w_{t-1} + \eta_t$, es decir, un modelo AR con parámetro $\frac{1}{2}$, el ruido $\{\eta_t\}$ se simula como una distribución normal estándar. Esto es análogo al primer caso.

- La estructura de modelo MTAR que genera los factores se establece así:

- Si $w_{t-1} < -0.53$

$$\phi_1^{(1)} = \begin{pmatrix} 0.8 & 0 \\ 0 & -0.9 \end{pmatrix},$$

- Si $w_{t-1} \geq -0.53$ y $w_{t-1} \leq 0.38$

$$\phi_1^{(2)} = \begin{pmatrix} -0.8 & 0 \\ 0 & 0.9 \end{pmatrix},$$

- Si $w_{t-1} > 0.38$

$$\phi_1^{(3)} = \begin{pmatrix} 0.6 & 0 \\ 0 & -0.6 \end{pmatrix},$$

- Las matrices de varianzas se establecen como matrices identidad para el ruido en los tres regímenes. Γ_j son matrices identidad para $j = 1, 2, 3$. Adicionalmente la matriz de carga Δ se define como:

$$\Delta = \begin{pmatrix} 0.29 & 0.94 \\ 0.54 & -0.05 \\ 0.34 & -0.06 \\ 0.47 & -0.23 \\ 0.52 & -0.22 \end{pmatrix},$$

La cual es la misma matriz del primer caso y las demás condiciones se mantienen análogas sin pérdida de generalidad.

La metodología se aplica de manera similar al comienzo, tal que se ordenan la dupla valores Y_t y Y_{t-1} respecto a w_{t-1} y bajo esta configuración se empieza a construir $\hat{M}_1(k)$ en cada punto del tiempo ordenado, esto siguiendo los pasos 1 y 2 de la metodología. De estas matrices se obtienen los valores propios y se construyen los diagramas correspondientes a los 5 valores propios como se indica en el paso 3. Estos resultados se muestran en la Figura 8, donde se puede observar dos marcados cambios en la evolución de los 2 valores propios más grandes de $\hat{M}_1(k)$. Conjuntamente, siguiendo el paso 4 de la metodología, se construye el algoritmo de [Bry and Boschan \(1971\)](#) para los diagramas correspondientes a los dos valores propios más grandes⁸, donde los puntos de quiebre escogidos (n_{940} y n_{962}), generan unas estimaciones previas del valor de corte del primer régimen de -0.59 y -0.61 correspondientes a los dos valores propios más grandes, los cuales son relativamente cercanos al valor real (-0.53). De estos dos valores, se escoge el más cercano al valor real, por lo tanto, el primer punto de corte es $\hat{\lambda}_1 = -0.59$. Todo lo mencionado anteriormente corresponde a una primera iteración de la metodología.

⁸En los anexos Figura 32

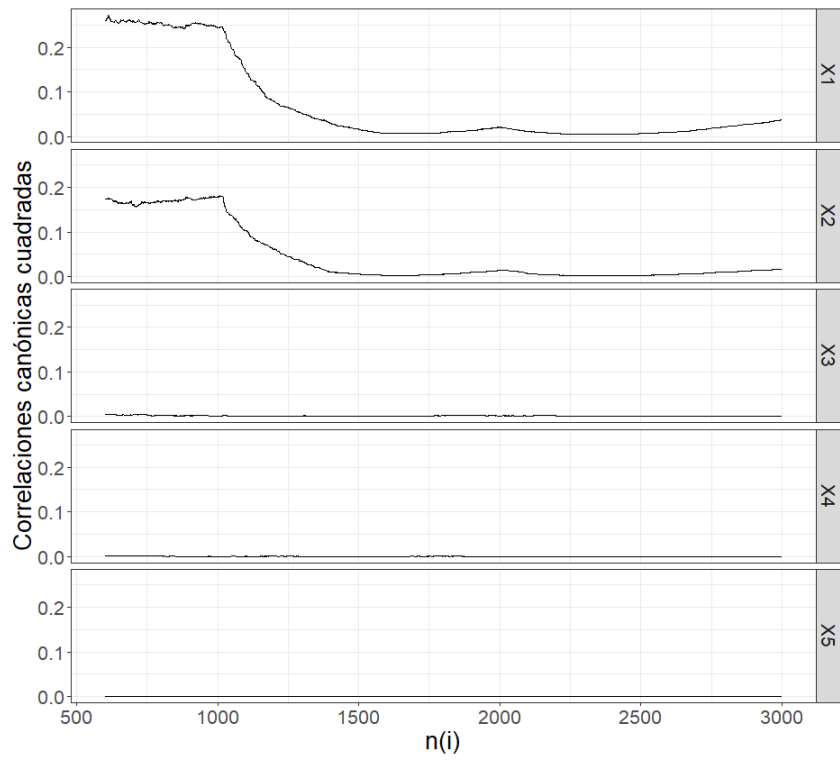


Figura 8: Correlaciones canónicas cuadradas - Primera iteración

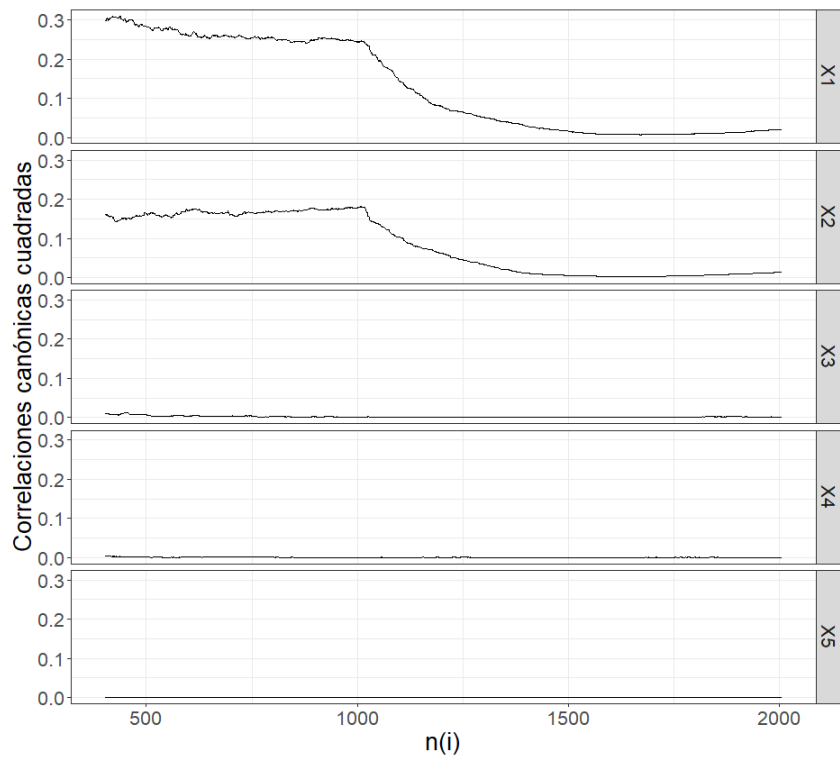


Figura 9: Correlaciones canónicas cuadradas - Segunda iteración

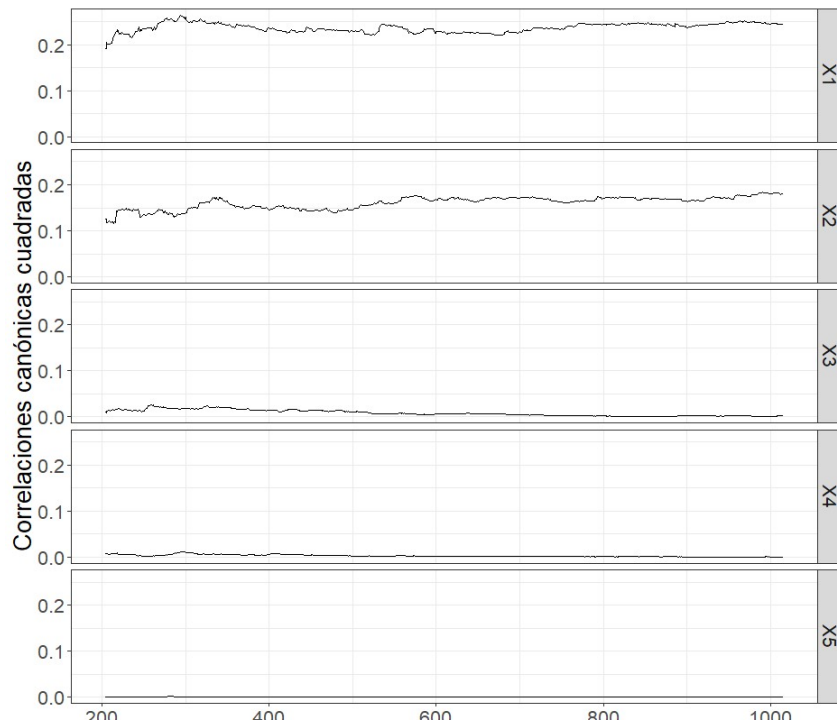


Figura 10: Correlaciones canónicas cuadradas - Tercer umbral

Ahora, con el primer punto de corte identificado, las observaciones que están antes de este punto de corte son excluidas ya que se asume que pertenecen al primer régimen. Con las observaciones restantes se vuelve a iniciar la metodología desde el paso 1, ordenando la dupla Y_t y Y_{t-k} respecto a la variable umbral y sus valores restantes. Esto se convierte en una segunda *iteración* del procedimiento, donde se construyen nuevas $\hat{M}_1(k)^{(2)}$ y se genera una nueva evolución de los valores propios de tal matriz, los resultados de esta *iteración* se pueden observar en la Figura 9, donde al tener ya *aisladas* las observaciones pertenecientes al primer régimen, se puede ver solo un cambio marcado de tendencia en los dos mayores valores propios de esta nueva matriz. En los anexos, se muestran los puntos de quiebre candidatos de la segunda iteración del procedimiento gracias al algoritmo de [Bry and Boschan \(1971\)](#). De los puntos candidatos, los más cercanos al punto de quiebre para cada valor propio, generan estimaciones del punto de corte de la variable de umbrales en 0.32 Y 0.34, que de nuevo, son valores muy cercanos al punto de corte original. De nuevo, por razones de simulación, se escoge $\hat{\lambda}_2 = 0.34$ como el nuevo punto de corte.

Para seguir con la aplicación de la metodología propuesta, se aíslan las observaciones pertenecientes tanto al primer como segundo régimen ya definidos por los valores $\hat{\lambda}_1$ y $\hat{\lambda}_2$ en las dos *iteraciones* anteriores. Así, se construye el último diagrama de evolución de valores propios de $\hat{M}_1(k)^{(2)}$. La evolución de los valores propios se muestra en la Figura 10, donde se puede ver que ya no existe ningún cambio *fuerte* en la dinámica de la evolución de los valores propios, y por tanto, se puede dar por terminada la identificación del número de regímenes. Los diagramas de correlaciones canónicas cuadradas de los dos primeros umbrales son mostrados en los anexos, aunque su dinámica se puede intuir desde las Figuras 8 y 9, donde se puede observar que antes del punto de corte definido en cada iteración, su comportamiento se puede asumir como estable.

La metodología sigue al construir las matrices $\hat{M}_1(k)$ para cada uno de los 3 grupos de observaciones establecidos. La identificación del número de factores para cada régimen se complementa con los resultados del Cuadro 3, donde se presentan las pruebas de [Peña and Poncela \(2006\)](#) para cada uno de los tres regímenes encontrados. Se evidencia claramente la existencia de por lo menos dos factores comunes en cualquiera de los 3 casos. Para finalizar, las estimaciones de las matrices de carga Δ se pueden observar en los anexos, donde de nuevo, se obtienen rotaciones de la matriz de carga original que se simuló, al igual que los casos de la sección

anterior. Finalmente, en la Figura 36 de los anexos, se presentan las 3 estimaciones del proceso de factores $\{f_t\}$, cada una correspondiente a cada régimen identificado. Para los tres casos, se aplicó la prueba de Tsay (1998) encontrando evidencias de no linealidad de tipo umbral.

Todos los casos hasta ahora considerados, muestran la capacidad de las correlaciones canónicas cuadradas para la identificación del modelo, sin importar el número de regímenes que se presenten, ya que sin pérdida de generalidad, este procedimiento se puede replicar para $j = 1, \dots, c$, teniendo en cuenta que se tengan *suficientes* observaciones por cada régimen, para poder observar una dinámica estable dentro de la evolución de las correlaciones canónicas cuadradas en cada umbral.

H_0	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$
Primer umbral	0	0	0.53	0.32	0.09
Segundo umbral	0	0	0.11	0.68	0.74
Tercer umbral	0	0	0.72	0.76	0.86

Cuadro 3: P-valores de la prueba de Peña and Poncela (2006) - Caso 3 umbrales

5.3. Simulación de control

En todos los casos de simulación anteriormente presentados, la metodología fue aplicada bajo el hecho de que el proceso de factores $\{f_t\}$ sigue un modelo MTAR y por tanto, las observaciones $\{Y_t\}$ siguen un modelo factorial dinámico de umbrales, tal como se describió en las ecuaciones (1) y (2). Es un hecho que la metodología desarrollada en este documento fue diseñada para el caso descrito, sin embargo, surge la siguiente pregunta: ¿Que sucede con la metodología cuando se aplica a un conjunto de datos que no tiene estas características? En particular, es interesante observar el caso cuando el proceso de factores $\{f_t\}$ es lineal. Cuando esto sucede, el modelo de las ecuaciones (1) y (2) se simplifica a un caso particular del modelo que se presenta en la ecuación (13), es decir, el modelo factorial dinámico de Peña and Poncela (2006). Por lo tanto, esta sección tiene como objetivo hacer evidentes las diferencias en los comportamientos de las correlaciones canónicas cuadradas de $\{Y_t\}$ cuando se está bajo el contexto del modelo de factores dinámicos, respecto a los comportamientos que se han visto cuando se supone que se sigue un modelo factorial dinámico de umbrales.

Para cumplir con este objetivo, se construye una simulación donde se toman condiciones similares a la simulación en 5.1. La principal diferencia radica en que para hacer la simulación del proceso de factores $\{f_t\}$, se restringe al caso $\phi_1^{(1)} = \phi_1^{(2)} = \phi_1$, con:

$$\phi_1 = \begin{pmatrix} 0.8 & 0 \\ 0 & -0.9 \end{pmatrix},$$

Esto implica que el proceso de factores generado es de dimensión 2 y que, además, no existiran regímenes y por lo tanto, no es necesario hacer la simulación de una variable de umbrales. Adicionalmente, para generar $\{Y_t\}$ se usa la misma matriz de transición Δ que se usó en el ejemplo 5.1, esto lleva a obtener un proceso de observaciones de dimensión 5.

Finalmente, las matrices de varianza y covarianza de los procesos de ruido $\{u_t\}$ y $\{a_t\}$ se asumen como matrices identidad y se simulan a través de distribuciones normales multivariadas. Todo la anterior sin pérdida de generalidad. Al igual que los casos anteriores, se simulan $T = 2000$ observaciones con un periodo de inicialización de 100.

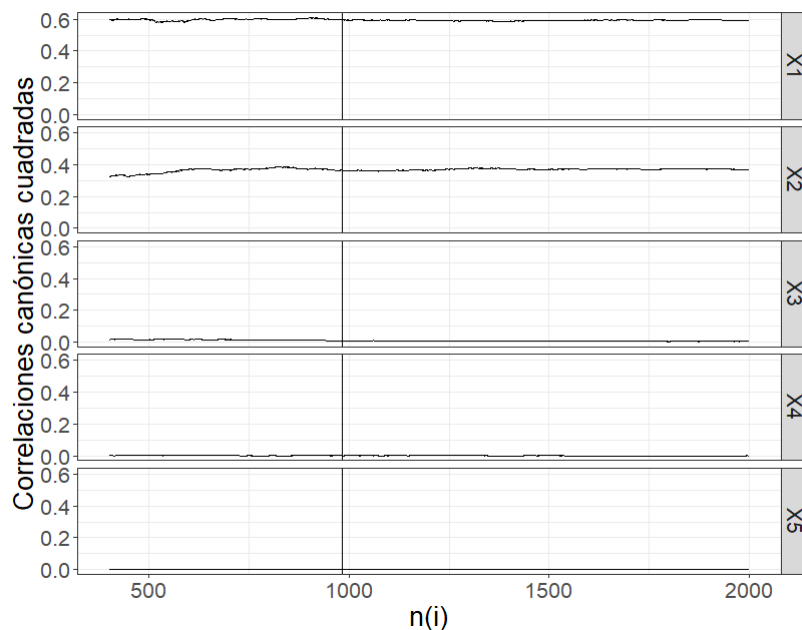


Figura 11: Correlaciones canónicas cuadradas - Simulación de control

Con las observaciones simuladas a la mano se siguen los mismos pasos metodológicos, específicamente los pasos 1, 2 y 3 de la metodología planteada. Así, se obtiene el diagrama de correlaciones canónicas cuadradas de la Figura 11, en donde se puede ver que no existe un cambio de dinámica en el desarrollo de los dos valores propios más grandes de la matriz $\hat{M}_1(k)$ calculada para este caso particular. Aunque se intente encontrar un cambio de dinámica en estos valores, esto puede deberse únicamente a la aleatoriedad resultante del procedimiento y no es claro en ningún punto. Esta es la principal diferencia que se evidencia respecto a las gráficas de todos los casos anteriores, en donde, debido al cambio de régimen, se hace evidente un cambio de dinámica en el comportamiento de los valores propios de $\hat{M}_1(k)$. Esta evidencia confirma aún más la capacidad de la metodología para detectar un comportamiento no lineal de tipo umbral.

Siguiendo con los resultados de esta simulación, dado que no existe tal no existe cambio de dinámica comentado, no es aplicado el algoritmo de Bry and Boschan (1971) debido a que no hay evidencia de un claro punto de quiebre, por lo tanto, se aplica directamente la prueba de Peña and Poncela (2006) sobre toda la muestra ordenada. Los resultados de la prueba son mostrados en el Cuadro 4, donde se puede observar que no se rechaza la hipótesis de $r = 2$ factores. Para finalizar, se construye con todas las observaciones, una estimación de Δ con el fin de evaluar la hipótesis de no linealidad sobre una estimación de $\{f_t\}$. Siguiendo las indicaciones de la metodología, se usa la prueba de Tsay (1998). Al calcular el estadístico de prueba se obtiene un p-valor de 0.18, con lo cual no se rechaza la hipótesis nula de linealidad y por tanto, se concluye que tal estimación del proceso de factores sigue la estructura del modelo planteado en la ecuación (13), es decir, el caso del modelo de factores de Peña and Poncela (2006).

H_0	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$
P-value	0	0	0.71	0.94	0.97

Cuadro 4: P-valores de la prueba de Peña and Poncela (2006) en el caso control

5.4. Replicaciones del caso 2 regímenes

Para apoyar los resultados de las simulaciones ejemplo mostradas anteriormente, se construyen réplicas del ejercicio de simulación de la Sección 5.1 con el fin de evaluar el comportamiento general del procedimiento. Para esto se construyen $r = 1000$ réplicas de cada caso donde se busca evaluar el comportamiento de las pruebas de Peña and Poncela (2006) y Tsay (1998), de tal manera que se puede observar algo de la variabilidad del procedimiento y parte de la capacidad de la metodología propuesta.

Para iniciar, se construye el diagrama de correlaciones canónicas para las 1000 réplicas siguiendo la metodología propuesta. Ahora, con el fin de resumir toda la información generada por cada iteración y para cada replica, se construye el diagrama con el promedio de cada valor propio en cada i -ésima posición temporal sobre las 1000 replicas, además, para observar el comportamiento de la variabilidad se construye una medida basada en dos veces la desviación estándar de cada valor propio en cada i -ésima posición hacia arriba y hacia abajo del promedio calculado. Esto se muestra en la Figura 12, donde se puede evidenciar un comportamiento similar a la Figura 1 de la Sección 5.1, pero con la medida de variabilidad incorporada. Se puede ver que es consistente el desarrollo de los valores propios de la matriz $M_1(k)$ hasta el punto marcado con la línea vertical, que es donde se simula el cambio de régimen generado por la variable umbral. También es de destacar que las correlaciones canónicas asociadas a los tres menores valores son sistemáticamente 0 tal como se observó en el ejemplo en 5.1., esto muestra que para este caso el procedimiento parece consistente.

H_0	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$
Primer umbral	1000	1000	260	62	11
Segundo umbral	1000	1000	286	52	5

Cuadro 5: Número de veces que se rechaza H_0 en la prueba de Peña and Poncela (2006) ($r = 1000$)

Siguiendo la propuesta metodológica se procede a separar las observaciones pertenecientes al primer y al segundo régimen, en este caso asumiendo que se conoce el punto de corte, solo para observar el comportamiento

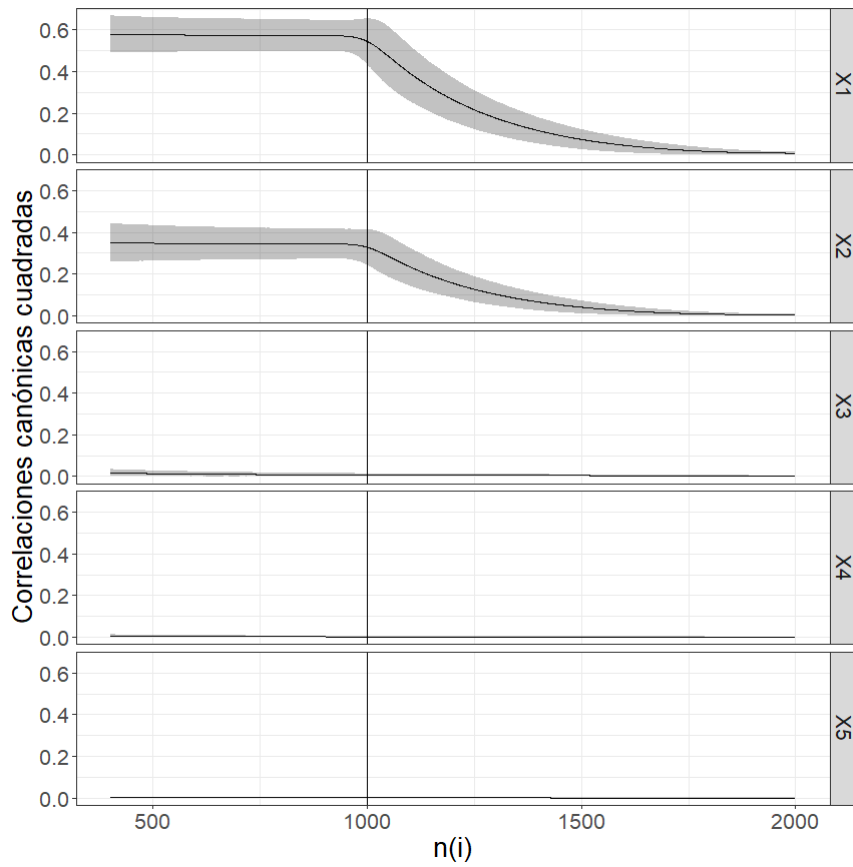


Figura 12: Correlaciones canónicas cuadradas - $r=1000$

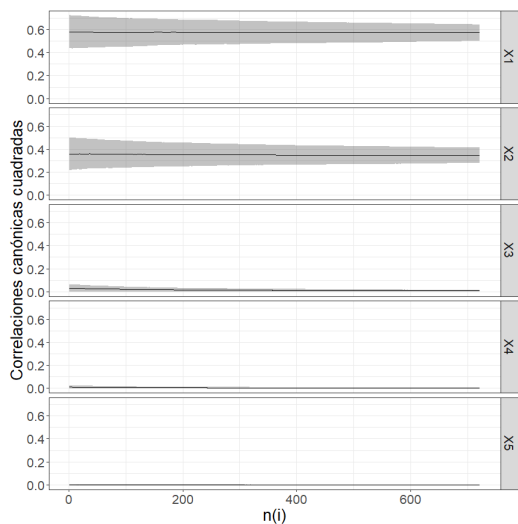


Figura 13: Corr. canónicas - 1er umbral - $r = 1000$

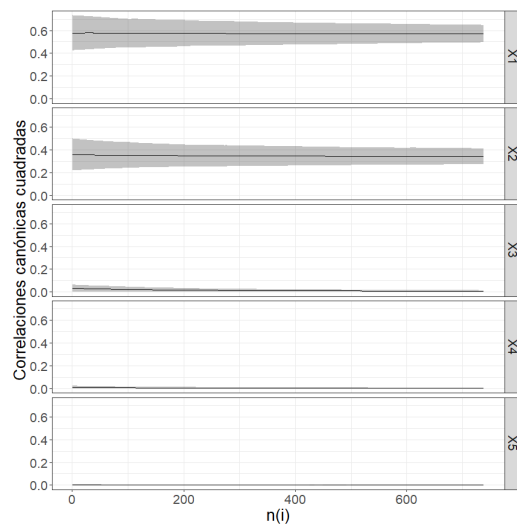


Figura 14: Corr. canónicas - 2do umbral - $r = 1000$

de las réplicas y se procede a construir los diagramas de valores propios de manera independiente para cada uno de los regímenes. De nuevo, estos diagramas se generan con el promedio de cada valor propio en cada

i -ésima posición temporal sobre las 1000 replicas, y la desviación estándar de cada valor propio en cada i -ésima posición hacia arriba y hacia abajo de tal promedio. Los resultados de estos diagramas se pueden encontrar en las Figuras 13 y 14 donde se puede ver que el comportamiento de los valores propios dentro de cada régimen es estable, simulando el comportamiento que ya se había observado en Peña and Poncela (2006). Además, se puede observar estabilidad en los valores y como en los primeros registros de cada umbral parece haber algo más de *variabilidad* resultante del procedimiento. Para asegurar este resultado y siguiendo los pasos de la metodología, se hacen las pruebas para el número de factores de Peña and Poncela (2006), los resultados sobre las 1000 replicas pueden encontrarse en el Cuadro 5, donde en todos los casos se rechaza la hipótesis de $r = 0$ y $r = 1$ factores, con lo cual se establece que sistemáticamente, en los dos umbrales considerados existen por lo menos $r = 2$ factores comunes, lo cual es consistente con el procedimiento de simulación.

Dentro de este ejemplo de replicación del procedimiento hay que resaltar un par de particularidades. El primero, es que no se aplica el algoritmo de Bry and Boschan (1971), ya que este procedimiento al ser una apoyo para el reconocimiento de los puntos de quiebre en la evolución de los diagramas de valores propios, se vuelve un procedimiento particular en cada caso, por lo tanto, su replicación no es necesaria. Por otro lado, los ejercicios de construcción del proceso de factores bajo la estimación de Δ propuesta no son resumidos, pero en ejercicios internos se establecen las mismas conclusiones, se obtienen estimaciones de Δ que son rotaciones de la matriz original y adicionalmente, los procesos de factores resultantes se establecen como no lineales comprobados a través de la prueba de Tsay (1998) en la mayoría de los casos. Esto se puede observar en el Cuadro 6⁹. Múltiples ejercicios de replications fueron construidos, con el fin de observar la viabilidad del proceso, en su gran mayoría y bajo diferentes adecuaciones a las ya aquí mostradas los resultados fueron alentadores. Aún así, en la siguiente sub-sección se presentan casos en los cuales la metodología pierde capacidad.

Δ	Número de veces que se rechaza H_0 en la prueba de Tsay
Primer umbral	985
Segundo umbral	979

Cuadro 6: Resultados de la prueba de Tsay (1998) sobre las replicas

5.5. Casos particulares

5.5.1. Gran dimensión

Durante el desarrollo del documento se ha tratado un caso muy particular donde la dimensión del proceso de serie de tiempo observado es de 5, aún así, el procedimiento propuesto se puede extender fácilmente a casos donde la dimensionalidad es mucho más alta. En esta sección se hace una revisión de algunos casos y resultados asociados a extender la dimensión de la serie de nuevo tomando como base el ejemplo de la Sección 5.1. Así que, f_t se obtiene de la siguiente manera:

- Se generan 2 factores comunes que provienen de un modelo MTAR con variable de rezago $w_t = 0.5w_{t-1} + \eta_t$, es decir, un modelo AR con parámetro 0.5, el ruido $\{\eta_t\}$ se simula como una distribución normal estándar.
- Según la estructura MTAR se establece: si $w_{t-1} < 0$

$$\phi_1^{(1)} = \begin{pmatrix} 0.8 & 0 \\ 0 & -0.9 \end{pmatrix},$$

- Si $w_{t-1} \geq 0$

$$\phi_1^{(2)} = \begin{pmatrix} -0.8 & 0 \\ 0 & 0.9 \end{pmatrix},$$

- Las matrices de varianzas se establecen como identidades para el ruido en los dos regimenes Γ_j son matrices identidad para $j = 1, 2$.

⁹Para cada régimen se asume un nivel de significancia del 5%.

Con f_t ya generado, se puede obtener Y_t tal que la matriz de carga de los factores genera un proceso de dimensión 16. De nuevo, se generan 2000 observaciones como longitud total de la serie de tiempo multivariada.

$$\Delta = \begin{pmatrix} -0.05 & 0.65 \\ 0.17 & -0.27 \\ 0.26 & 0.31 \\ 0.07 & -0.15 \\ -0.11 & -0.33 \\ 0.08 & -0.26 \\ -0.25 & -0.11 \\ 0.13 & 0.16 \\ -0.17 & 0.18 \\ 0.20 & -0.39 \\ 0.15 & 0.24 \\ -0.42 & -0.08 \\ 0.41 & -0.31 \\ 0.56 & 0.19 \\ -0.10 & 0.03 \\ -0.06 & 0.42 \end{pmatrix}$$

Para finalizar la construcción, el proceso de ruido $\{u_t\}$ es generado por una distribución normal multivariada estándar. La simulación se construye con un periodo de inicialización de $n_0 = 100$ en la construcción de los factores.

La Figura 15 presenta la construcción del diagrama de valores propios de la matriz $M_1(k)$ para la serie simulada de dimensión 16 siguiendo la metodología que se ha trabajado durante el desarrollo del documento. Es evidente que la gráfica muestra un comportamiento similar al que ya se ha venido mostrando, con dos valores propios marcadamente más grandes y con un desarrollo diferente a los demás, por lo menos en las observaciones pertenecientes al primer régimen, y son estos valores, los que muestran un cambio de dinámica en el punto donde empiezan a ingresar observaciones asociadas al segundo régimen.

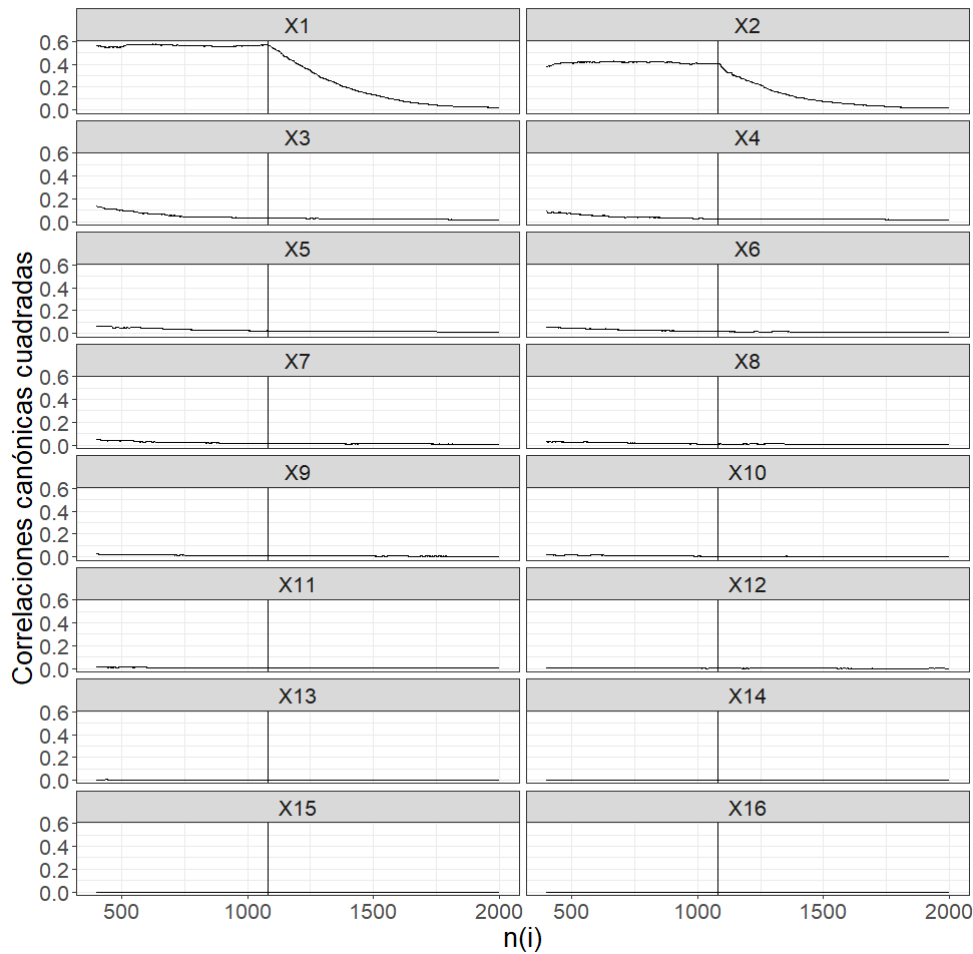


Figura 15: Correlaciones canónicas cuadradas - $k = 16$

La conclusión en este caso es la misma a los análisis anteriores, se tienen dos valores propios en donde se hace explícito un cambio de dinámica y es fácil reconocer un punto de corte claro que llevaría a pensar la existencia de no linealidad de tipo umbral. Siguiendo con la metodología, se construyen los diagramas de valores propios de las matrices $M_1(k)$ para las observaciones clasificadas en cada régimen (Figuras 16 y 17), estas muestran un patrón similar a las gráficas anteriormente mostradas, solo que se empieza a diferenciar que los valores propios en algunos recuadros, al principio de su evolución no son 0. Esto es más claro sobre todo en el tercer y cuarto recuadro, aunque, los valores son cercanos a 0, no son del todo nulos. Ya al final de la evolución de los diagramas, todos estos valores propios parecen ya ser 0, esto llevaría a pensar que un análisis desde el gráfico y con toda la muestra traería la conclusión adecuada. Sin embargo, en el estado de la naturaleza deberían ser nulos, por lo tanto, el procedimiento demora en reconocer esta característica para este caso, si se compara respecto a las simulaciones anteriores donde el desarrollo siempre parecía ser nulo.

Para reforzar esta conclusión, se construye la prueba de número de factores de Peña and Poncela (2006) para cada uno de los regímenes encontrados. Para este ejemplo particular, se encuentra que se rechazan las hipótesis nulas de $r = 0$, $r = 1$, $r = 2$, $r = 3$ y $r = 4$, afirmando entonces que el proceso tiene al menos 5 factores comunes. Se construyeron replicas para diversos escenarios no solo de la dimensión del proceso de factores $\{f_t\}$ si no del número de regímenes y el número de muestra utilizado, evidenciando que la prueba de Peña and Poncela (2006) aplicada a este tipo de escenarios no da buenos resultados a medida que el número de dimensiones va creciendo. En conclusión, lo que se tiene es que el uso de la prueba de Peña and Poncela (2006) para hallar el número de factores comunes en modelos factoriales de umbrales no muestra los resultados deseados, cuando la dimensión

del proceso observado es grande. Sin embargo, gráficamente puede tenerse un criterio del cual apoyarse y poder concluir de buena manera, debido a que los diagramas muestran un comportamiento esperado.

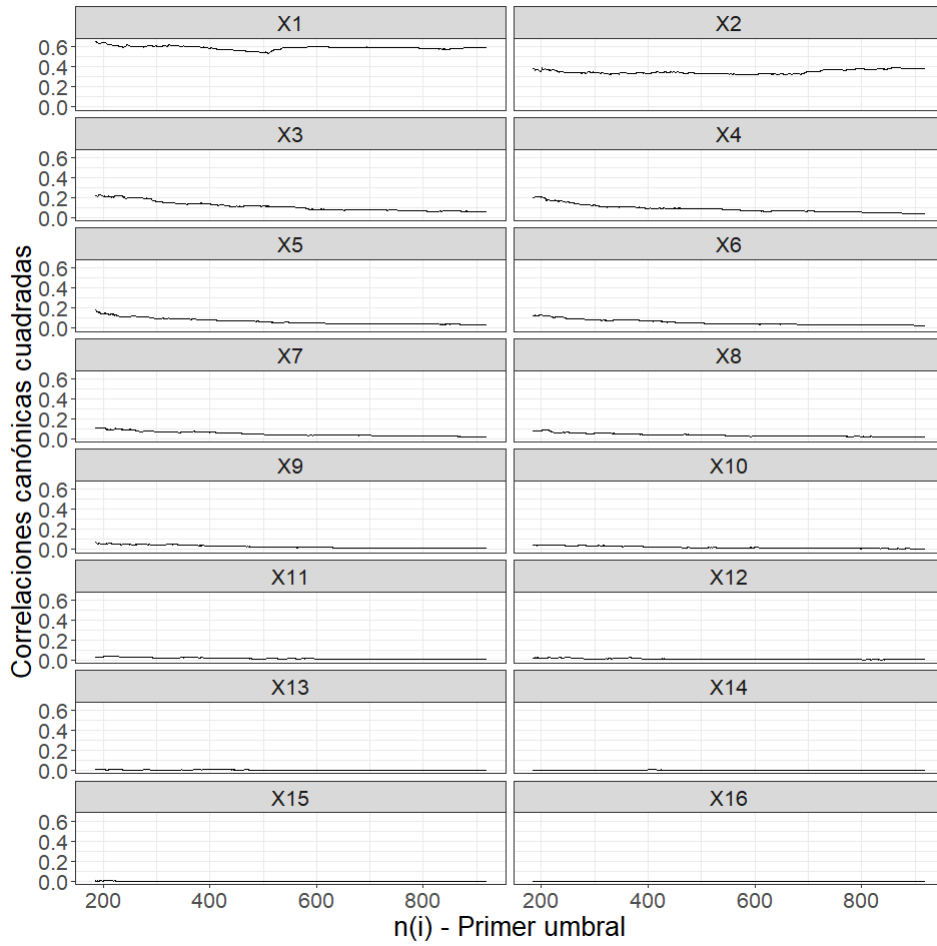


Figura 16: Corr. canónicas - 1er umbral - $k = 16$

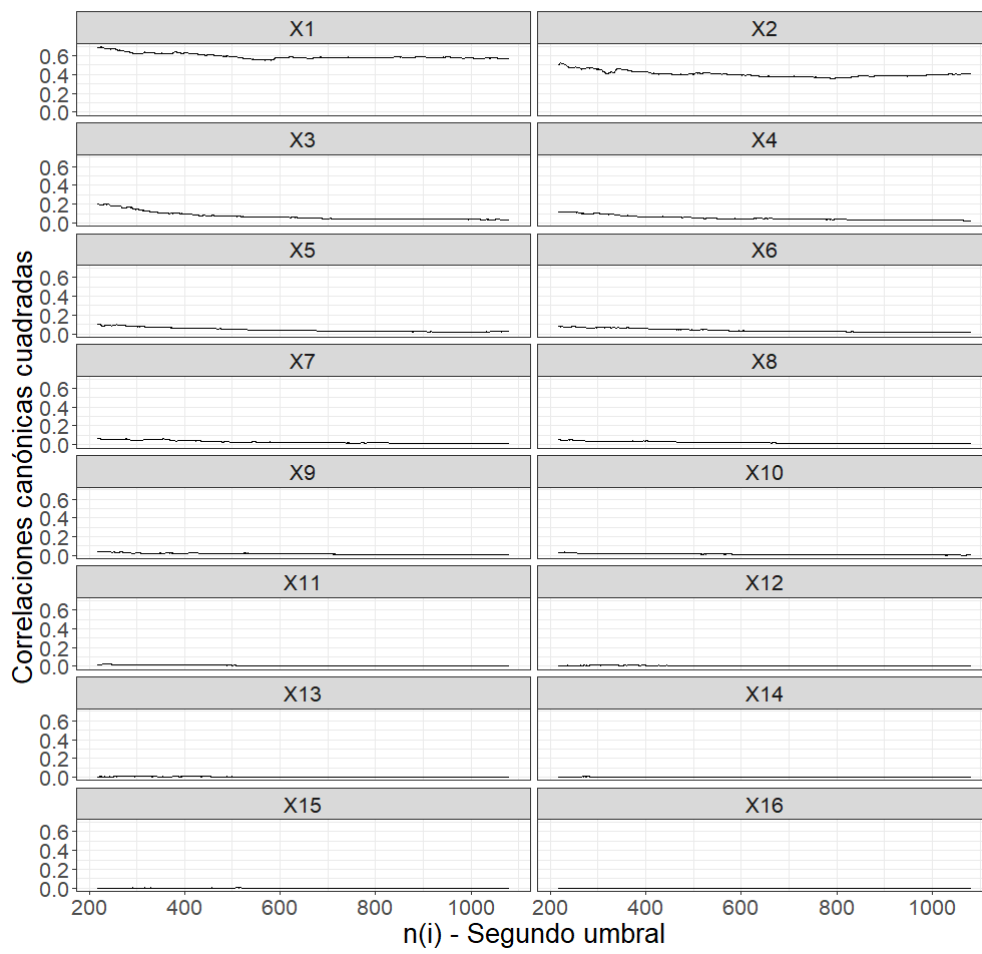


Figura 17: Corr. canónicas - 2do umbral - $k = 16$

6. Aplicación con datos reales

Para dar un ejemplo del uso del esquema metodológico propuesto, se usa un conjunto de datos compuesto por los precios de cierre de 9 de las principales acciones del índice bursátil S&P500 el cual es considerado el índice más representativo de la situación real del mercado Norteamericano¹⁰. Esta información está disponible en frecuencia diaria para los años 2014 y 2015, para un total de 502 observaciones. Los nombres de las acciones utilizadas son mostradas en el Cuadro 10 en los anexos. Adicionalmente, el comportamiento particular de cada una de estas acciones es mostrado en la Figura 18. Ahora, para seguir las especificaciones de la metodología, se transforman mediante el logaritmo y se obtienen los cambios diarios, estos ajustes son mostrados en la Figura 19.

Por otro lado, dentro del contexto metodológico en el que se desarrolla este trabajo es necesaria una *variable de umbrales*, la cual define la dinámica conjunta de las series de tiempo comentadas anteriormente. De manera natural, se puede pensar que la serie candidata como variable de umbrales es el propio índice S&P500, esta serie también es transformada por medio de la primera diferencia y el logaritmo para obtener la estacionariedad requerida (Ver Figura 20).

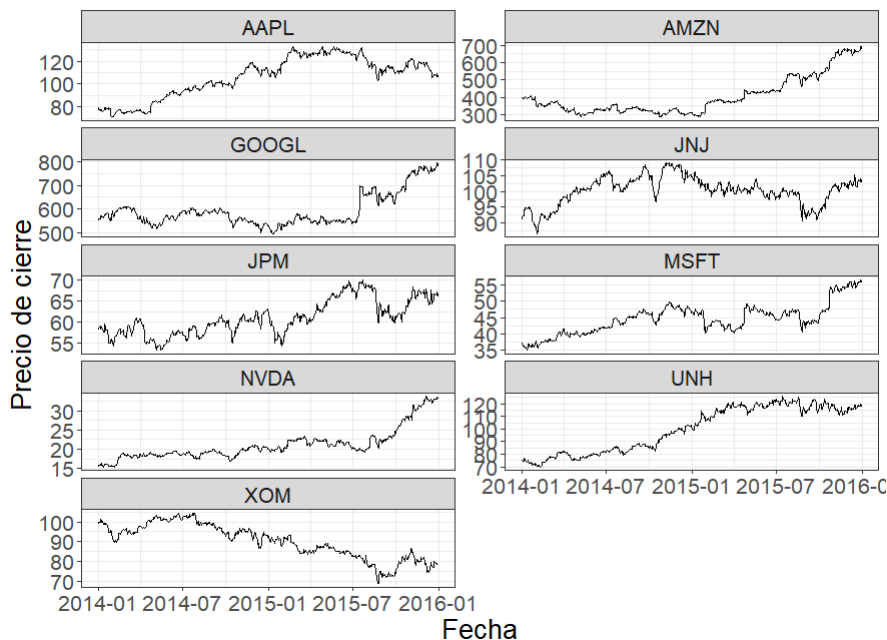


Figura 18: Acciones

El objetivo de este ejemplo será evaluar la dinámica de las acciones respecto a la evolución del índice S&P500 para algún rezago d . La metodología se aplica en varias etapas bajo los pasos metodológicos ya definidos. En principio, se busca un valor candidato k para evaluar la matriz $\hat{M}_1(k)$, para esto, se ajustan diversos modelos VAR para las observaciones de la Figura 19 y el mejor modelo respecto al criterio de información de Akaike (AIC)¹¹ es el escogido como valor candidato. El valor encontrado fue $k = 1$. Para la segunda etapa del procedimiento, se busca construir el esquema de valores propios de la matriz $\hat{M}_1(k)$, para $k = 1$ ordenando la pareja de observaciones $\{Y_t\}$ y $\{Y_{t-1}\}$ respecto a la transformación del índice S&P500 pero esta vez, para los rezagos $d = 0, 1$. Eso se hace con el fin de evaluar más posibilidades del impacto de la variable de rezagos, lo que significa que se mostrarán dos resultados diferentes ya que se dan dos ordenamientos diferentes. Además, evaluar los rezagos $d = 0, 1$ de la transformación del índice S&P500, significa desde la práctica que los cambios en el S&P500 en t o $t - 1$ definen la dinámica de los precios de las acciones en t , lo cual tiene sentido desde

¹⁰Información adicional sobre el indicador puede ser encontrada en <https://www.spglobal.com/spdji/en/documents/methodologies/methodology-sp-us-indices.pdf>

¹¹Información adicional sobre este criterio puede encontrarse en Akaike (2011).

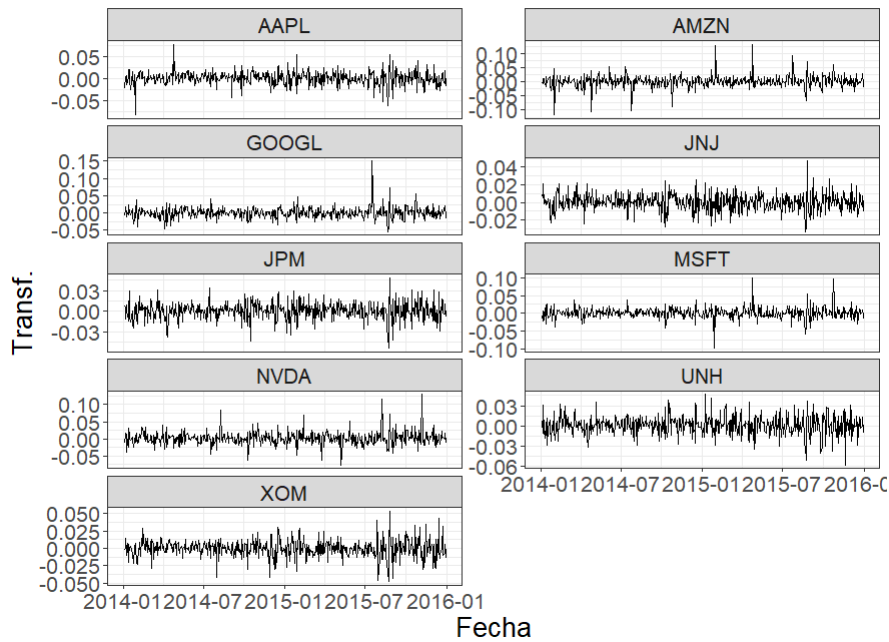


Figura 19: Acciones - Transformaciones

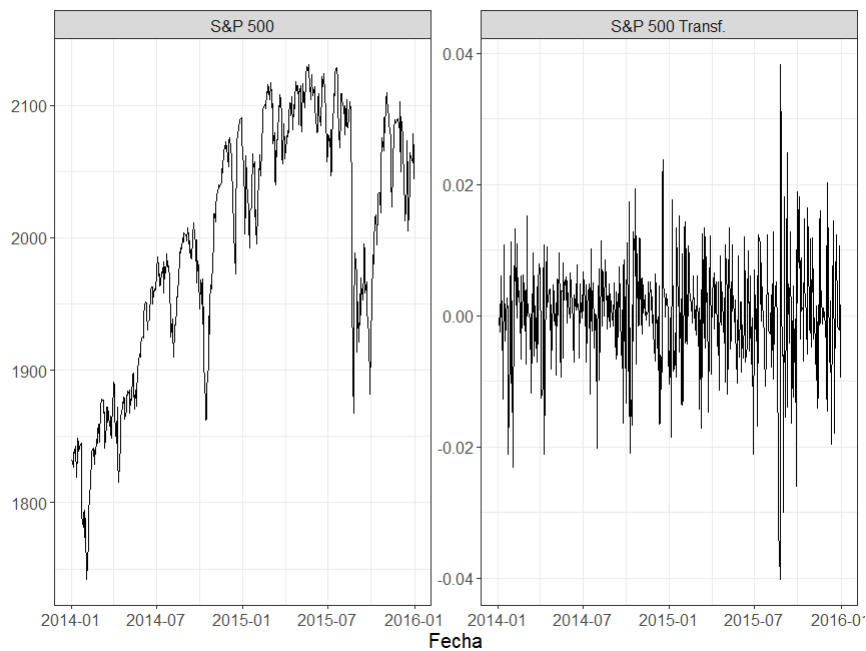


Figura 20: Índice S&P 500 y su transformación

el ejemplo planteado. El diagrama de las correlaciones valores propios de $\hat{M}_1(k)$ cuando el orden es respecto a $S\&P500_{d=0}$ para $k = 1$ es mostrado en la Figura 21, donde se evidencian dos cosas. En primera instancia, los primeros 3 valores propios muestran una dinámica que destaca respecto a los demás, y segundo, dentro de estos 3 gráficos existe un cambio de dinámica alrededor de la observación 250. Para confirmar este punto de corte se aplica el algoritmo de [Bry and Boschan \(1971\)](#) para las tres primeras series, dando como resultado puntos de

quiebre cercanos alrededor de este punto¹², lo cual confirma un tipo de no linealidad similar a lo identificado por medio de los ejercicios de simulación. Más específicamente, este punto de corte genera un valor umbral de $\gamma_1 = 0.003$ y es el que define la dinámica de las series desde el nivel del índice S&P 500. Al ser cercano a 0 se puede suponer que choques positivos en el índice generan una dinámica, mientras que choques negativos generan otra dinámica diferente.

Ahora, para seguir con esta implementación se aplica la prueba de [Peña and Poncela \(2006\)](#) para los 2 grupos generados por el umbral, encontrando que, en los dos casos, se tienen a lo más 3 factores comunes dinámicos. Los p-valores asociados a estas pruebas son encontrados en la [Tabla 7](#).

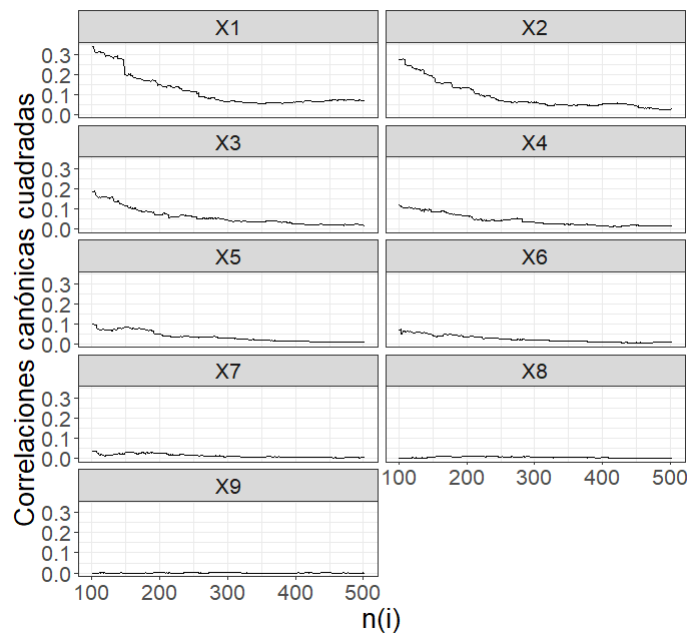


Figura 21: Diagrama de correlaciones canónicas - $d = 0$

H_0	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$
Primer umbral	0	0	0	0.01	0.97	0.98	0.99	0.88
Segundo umbral	0	0	0	0	0.13	0.41	0.07	0.20

Cuadro 7: P-valores de la prueba de [Peña and Poncela \(2006\)](#) - Aplicación práctica $d = 0$

Además, si suponemos $d = 1$ y se realiza el cálculo de la matriz $\hat{M}_1(k)$ con este ordenamiento, se puede observar en la [Figura 22](#) y la [Tabla 8](#) resultados similares a los encontrados en el caso $d = 0$, solo que en esta ocasión el punto de corte parece estar alrededor de 200, tomando este valor como corte, dada la sugerencia también evidenciada en los postulados del algoritmo [Bry and Boschan \(1971\)](#), se obtiene un valor umbral $\gamma = 0.001$ muy similar al encontrado en el primer caso. Adicionalmente parece que existen dos valores propios marcadamente mayores a los demás, pero al analizar la [Tabla 8](#) se puede encontrar que a lo más, existen 3 factores comunes en los 2 regímenes encontrados.

Siguiendo con la aplicación de la metodología, se obtienen las matrices de carga Δ estimadas por medio de los vectores propios de la matriz $\hat{M}_1(k)$ en los dos subconjuntos definidos por los regímenes encontrados, estas matrices son:

¹²Ver [Figuras 37, 38 y 39](#) en los anexos

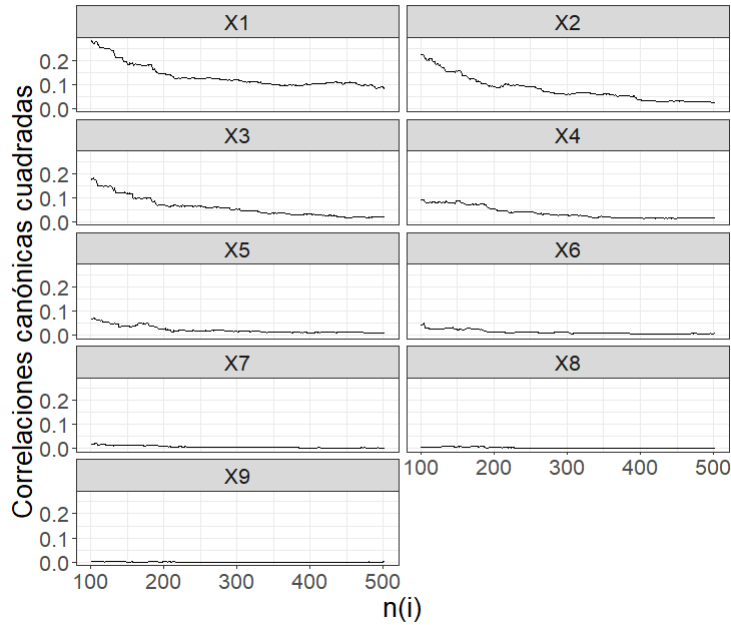


Figura 22: Diagrama de correlaciones canónicas - $d = 1$

H_0	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$
Primer umbral	0	0	0	0	0.99	0.99	0.94	0.82
Segundo umbral	0	0	0	0	0.98	0.98	0.88	0.62

Cuadro 8: P-valores de la prueba de Peña and Poncela (2006) - Aplicación práctica $d = 1$

$$\hat{\Delta}_1 = \begin{pmatrix} 0.14 & -0.41 & 0.35 \\ 0.06 & 0.45 & 0.31 \\ -0.54 & -0.17 & 0.07 \\ 0.61 & -0.42 & -0.66 \\ -0.44 & -0.16 & -0.45 \\ -0.04 & -0.20 & -0.01 \\ 0.20 & -0.03 & 0.27 \\ 0.24 & 0.37 & -0.23 \\ 0.16 & 0.46 & -0.10 \end{pmatrix} \quad \hat{\Delta}_2 = \begin{pmatrix} 0.04 & -0.25 & -0.59 \\ -0.12 & -0.07 & 0.25 \\ 0.47 & -0.26 & -0.13 \\ -0.55 & -0.51 & 0.61 \\ 0.29 & 0.52 & 0.31 \\ 0.48 & 0.23 & 0.17 \\ -0.33 & 0.30 & -0.06 \\ -0.13 & -0.04 & 0.03 \\ -0.13 & -0.44 & -0.26 \end{pmatrix}$$

Con estas matrices a la mano, se construyen los procesos $\{f_t\}$ estimados para cada uno de los regímenes encontrados. Para finalizar la aplicación de la metodología, se aplica la prueba de Tsay (1998) a estos procesos generados $\{\hat{f}_t^{(j)}\}$ para $j = 1, 2$, dando como resultado, que en los dos casos, se rechaza la hipótesis nula de linealidad, por lo tanto, se puede asumir que los procesos resultantes son no lineales de tipo umbral. En las Figuras 23 y 24 puede encontrarse las gráficas de estos procesos.

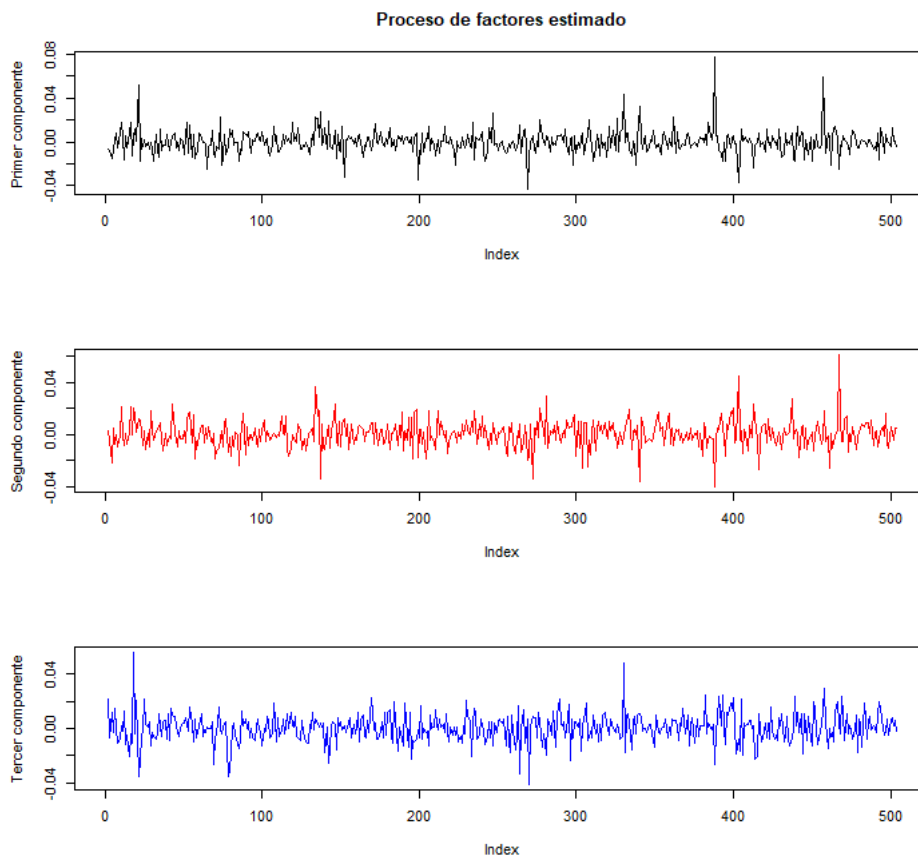


Figura 23: Proceso de factores - Primer régimen

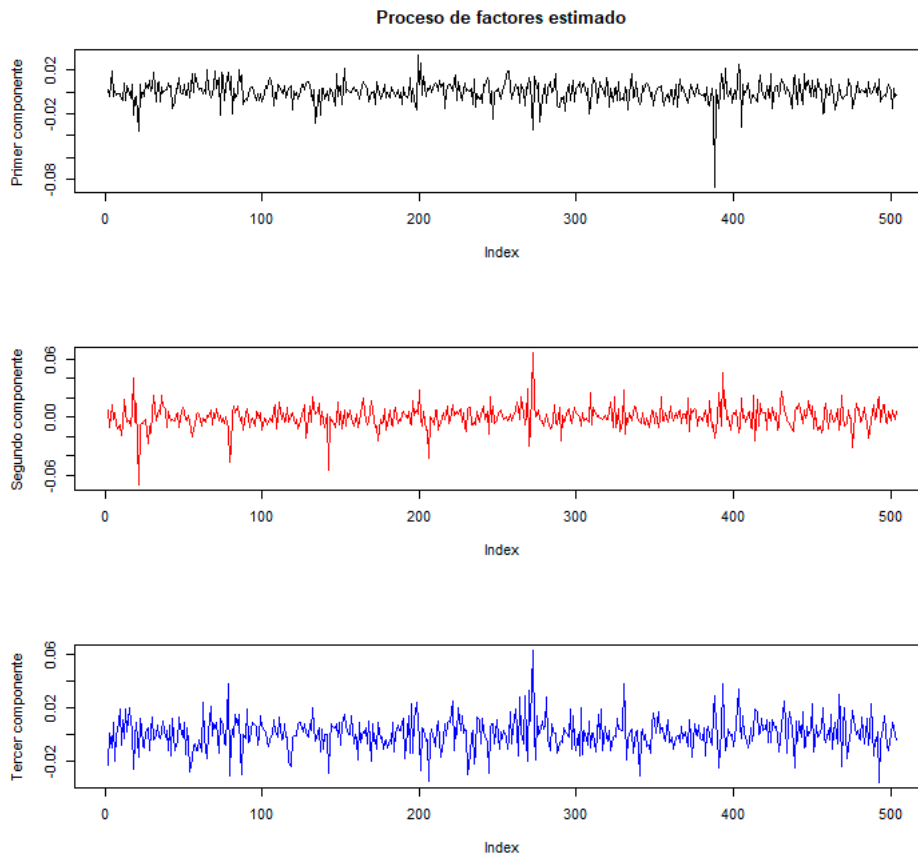


Figura 24: Proceso de factores - Segundo régimen

Hay que destacar que estos pasos que se siguieron dentro de esta aplicación, sugieren que pueden haber hasta 3 factores comunes, que el valor umbral puede ser cercano a 0 y que pueden existir 2 regímenes, lo que cumple con las expectativas de identificación planteadas en los objetivos de este trabajo. Finalmente, para hacer una comparación respecto al caso *lineal* y observar los posibles cambios que se presentan al hacer uso del modelo propuesto, se presenta en la Tabla 9 los resultados de la prueba de Peña and Poncela (2006) suponiendo que no existen los regímenes y aplicando sobre todas las observaciones, esta prueba muestra que solo existe 1 factor. Debido a esto, puede que se este obviando alguna dinámica dentro del proceso considerado.

H_0	$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$
Primer umbral	0.003	0.19	0.03	0.002	0.99	0.99	0.99	0.95

Cuadro 9: P-valores de la prueba de Peña and Poncela (2006) - Aplicación práctica - Caso estacionario

7. Conclusiones

Se muestra en el documento una metodología que toma algunos conceptos de [Peña and Poncela \(2006\)](#) y [Tsay \(1998\)](#) para la identificación del modelo factorial dinámico de umbrales a través de las correlaciones canónicas. En el desarrollo del documento se pudo ver que el procedimiento propuesto hace una identificación tanto del número de regímenes como el número de factores de una manera correcta bajo los escenarios planteados y las consideraciones comentadas. Se puede ver cómo bajo los diversos casos de simulación, se puede establecer que las correlaciones canónicas contienen la información para hacer una identificación plena de tales valores, esto dota a la propuesta de [Correal and Peña \(2008\)](#) de herramientas para poder procesar de mejor manera el paso de estimación, al conocer previamente valores de lo que se puede considerar como *hiperparámetros* dentro de su procedimiento e incluso una estimación previa de la matriz de pesos de los factores.

Este documento abre un camino en el cual se pueden evaluar las correlaciones canónicas y sus características dentro de los análisis de la identificación de modelos de este tipo, sobre todo teniendo en cuenta la ordenación dictaminada desde la variable umbral, la cual, conserva la dinámica de asociación temporal pero ayuda a hallar cambios en los patrones de la información. Como se comentó, esta idea se recogió desde los resultados de [Tsay \(1998\)](#), y finalmente se logró una integración adecuada con las ideas de [Peña and Poncela \(2006\)](#) al utilizar el cálculo de las correlaciones canónicas cuadradas y la prueba de número de factores para lograr los objetivos planteados en este trabajo. Un punto adicional a recalcar en este documento es que también se hacen explícitas ciertas limitaciones que pueda llegar a tener la metodología, poniendo énfasis en el caso de gran dimensión, no por los resultados mostrados a través de los gráficos, sino de un limitante de la prueba desarrollada por [Peña and Poncela \(2006\)](#) en este caso particular. Por otro lado, las replicaciones de algunos de los casos de simulación dotan de mayor validez a la metodología y motivan a seguir evaluando este camino buscando aplicaciones más diversas de las correlaciones canónicas en este contexto.

Para finalizar, hay que destacar que el ejercicio realizado con datos reales pone en evidencia la capacidad de la metodología para postular tanto el número de regímenes como el número de factores, cuando se plantea un modelo factorial dinámico de umbrales como posible modelo de los datos. Varias rutinas de estimación buscan los parámetros dentro de una grilla de posibilidades y al reducir el conjunto de posibilidades gracias a estos parámetros postulados se puede evidenciar la potencial utilidad de la metodología aquí propuesta. Aun así, la principal conclusión de este documento gira en torno a la gran utilidad que pueden tener las correlaciones canónicas en este tipo de modelos y cómo se puede incluso llegar a estimar todos los componentes a través de este camino.

Referencias

- Akaike, H. (2011). Akaike's information criterion. *International encyclopedia of statistical science*, pages 25–25.
- Anderson, T. W. (1963). The use of factor analysis in the statistical analysis of multiple time series. *Psychometrika*, 28:1–25.
- Bolívar, S., Nieto, F. H., and Peña, D. (2021). On a new procedure for identifying a dynamic common factor model. *Revista Colombiana de Estadística*, 44(1):1–21.
- Bry, G. and Boschan, C. (1971). Front matter to cyclical analysis of time series: Selected procedures and computer programs". In *Cyclical analysis of time series: Selected procedures and computer programs*, pages 13–2. NBER.
- Calderón-Villanueva, S. A. and Nieto, F. H. (2021). Forecasting with multivariate threshold autoregressive models. *Revista Colombiana de Estadística*, 44(2):369–383.
- Correal, M. E. and Peña, D. (2008). Threshold dynamic factor model. *Revista Colombiana de Estadística*, 31(2):183–192.
- Hotelling, H. (1935). The most predictable criterion. *Journal of educational Psychology*, 26(2):139.
- King, R. G. and Plosser, C. I. (1994). Real business cycles and the test of the adelmans. *Journal of Monetary Economics*, 33(2):405–438.
- Liu, X. and Chen, R. (2020). Threshold factor models for high-dimensional time series. *Journal of Econometrics*, 216(1):53–70.
- Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*. Springer, 2 edition.
- Lütkepohl, H. (2006). Forecasting with VARMA models. *Handbook of economic forecasting*, 1:287–325.
- Massacci, D. (2017). Least squares estimation of large dimensional threshold factor models. *Journal of Econometrics*, 197(1):101–129.
- Monroy, L. G. D. (2007). *Estadística multivariada: inferencia y métodos*. Universidad Nacional de Colombia.
- Pan, J. and Yao, Q. (2008). Modelling multiple time series via common factors. *Biometrika*, 95(2):365–379.
- Peña, D. and Poncela, P. (2006). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference*, 136(4):1237–1257.
- Stock, J. H. and Watson, M. (2011). Dynamic factor models. *Oxford Handbooks Online*.
- Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179.
- Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(2):157–195.
- Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. In *Exploration Of A Nonlinear World: An Appreciation of Howell Tong's Contributions to Statistics*, pages 9–56. World Scientific.
- Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *journal of the american statistical association*, 93(443):1188–1202.
- Wu, J. (2021). Estimation of high dimensional factor model with multiple threshold-type regime shifts. *Computational Statistics & Data Analysis*, 157:107153.

8. Anexos

Anexo 1 - Algoritmo Bry-Boschan

El algoritmo de Bry-Boschan ([Bry and Boschan \(1971\)](#)) tiene como intención establecer los puntos de quiebre de una serie de tiempo por medio de sus máximos y mínimos locales, entendiendo como máximo o mínimo local, un valor que es máximo o mínimo en un intervalo fijo de la serie de tiempo. El algoritmo fue diseñado para series mensuales, sin embargo, puede ser aplicado a una gran cantidad de casos haciendo pequeñas variaciones al desarrollo.

Los detalles del algoritmo pueden encontrarse en la referencia original, pero, en [King and Plosser \(1994\)](#) se hace un paso a paso que se resume en las siguientes líneas. El algoritmo inicia reemplazando valores que se consideren extremos dentro de la serie de tiempo, esto, para lograr hallar los ciclos de manera más efectiva. Como segundo paso, el algoritmo determina ciclos dentro de la serie calculando un promedio móvil de orden 12 y encuentra los puntos altos o bajos de la serie con ayuda de las 5 observaciones anteriores y posteriores de cada paso que se este analizando. Además, se restringe la duración de un ciclo a mínimo 15 pasos, eliminando picos y valles de los ciclos de menor tamaño y con esto se logra una alternancia de los puntos de quiebre con los picos mas altos y los valles mas bajos. Finalmente, los puntos de quiebre se vuelven a hallar en la serie original mediante promedios móviles de corto plazo (de 3 a 6 pasos). Los puntos de quiebre que presenta la metodología, son los picos o los valles de los ciclos de mayor tamaño y que no se consideran como valores extremos dentro de la serie.

Anexo 2 - Anexos de la sección 5.1

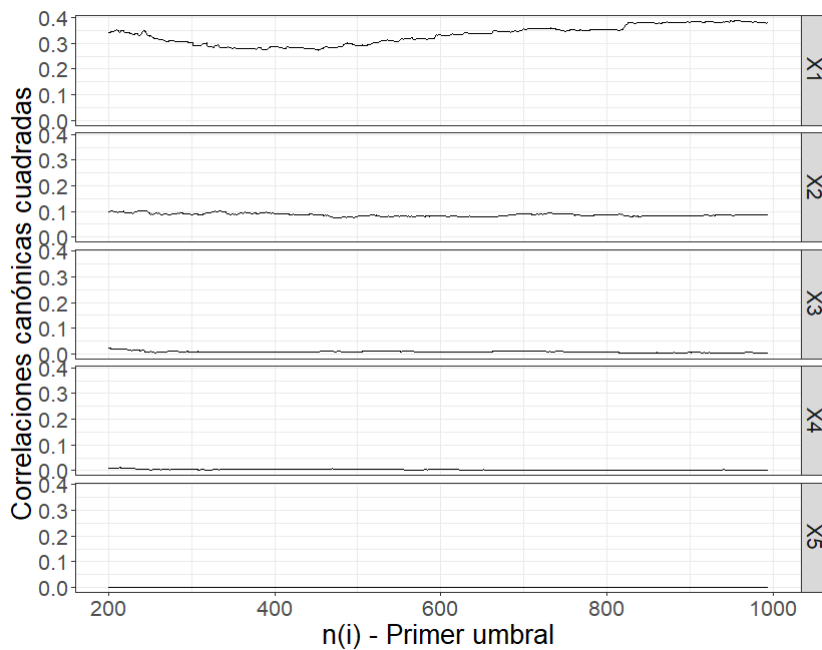


Figura 25: Correlaciones canónicas cuadradas - 1er umbral - Segundo ejemplo

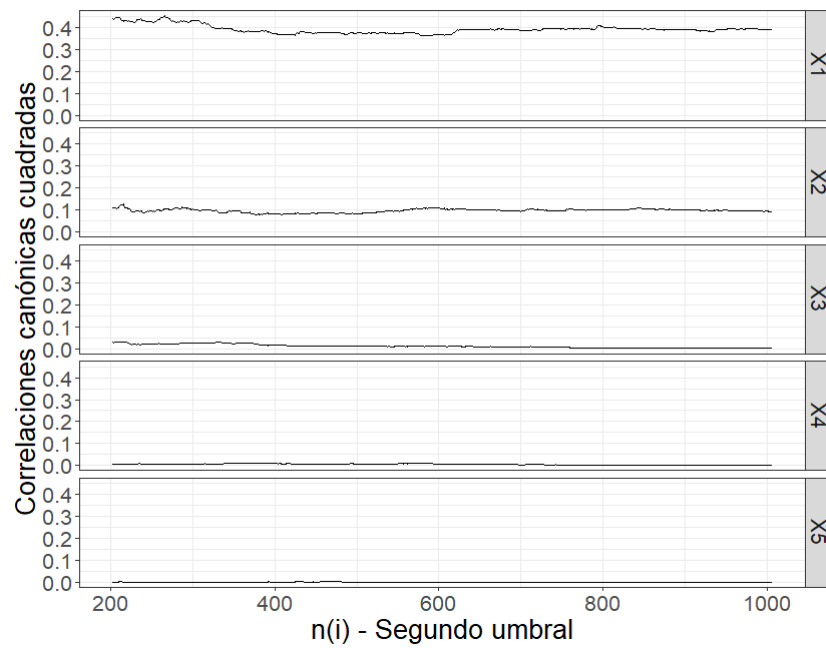


Figura 26: Correlaciones canónicas cuadradas - 2do umbral - Segundo ejemplo

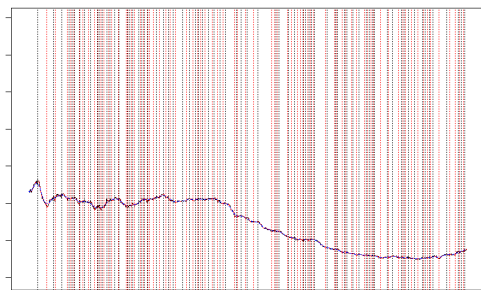


Figura 27: Algoritmo [Bry and Boschan \(1971\)](#) - Valor propio más grande

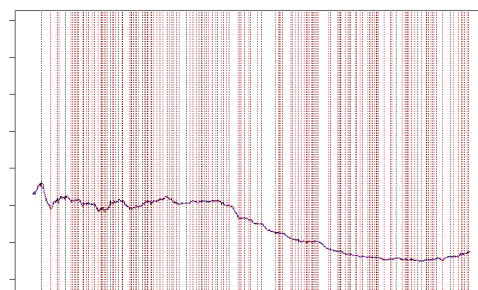


Figura 28: Algoritmo [Bry and Boschan \(1971\)](#) - Segundo valor propio más grande

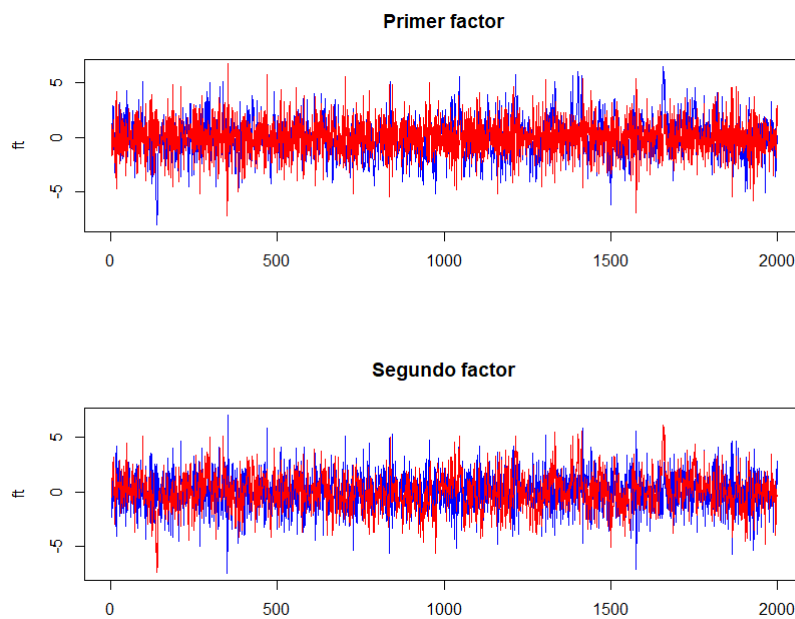


Figura 29: Estimaciones del proceso de factores - Segundo ejemplo

Estimaciones preliminares de Δ

$$\hat{\Delta}_1 = \begin{pmatrix} 0.92 & -0.20 \\ 0.08 & -0.54 \\ -0.11 & -0.44 \\ -0.21 & -0.18 \\ -0.26 & -0.65 \end{pmatrix} \quad \hat{\Delta}_2 = \begin{pmatrix} -0.21 & 0.92 \\ -0.48 & -0.07 \\ -0.39 & 0.09 \\ -0.53 & -0.28 \\ -0.52 & -0.22 \end{pmatrix}$$

Anexo 3 - Anexos de la sección 5.2

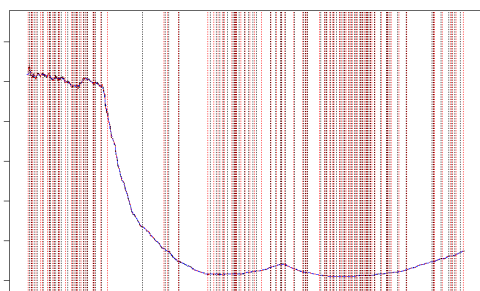


Figura 30: Algoritmo Bry and Boschan (1971) - Valor propio más grande - Primera iteración

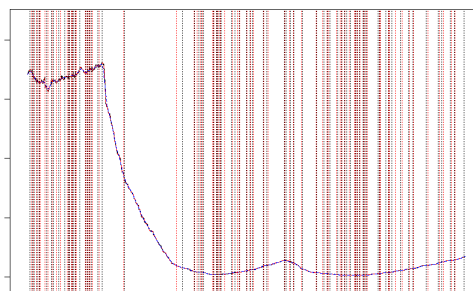


Figura 31: Algoritmo Bry and Boschan (1971) - Segundo valor propio más grande - Primera iteración

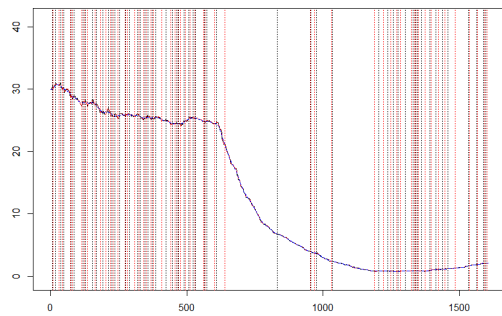


Figura 32: Algoritmo Bry and Boschan (1971) - Valor propio más grande - Segunda iteración

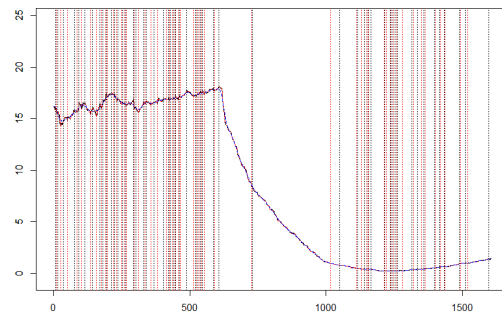


Figura 33: Algoritmo Bry and Boschan (1971) - Segundo valor propio más grande - Segunda iteración

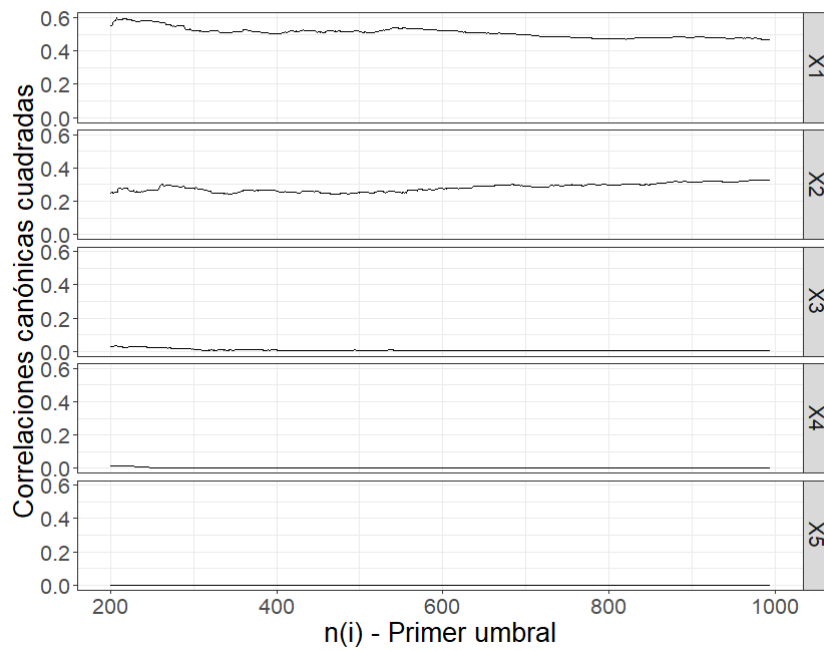


Figura 34: Correlaciones canónicas cuadradas - 1er umbral - $j = 3$

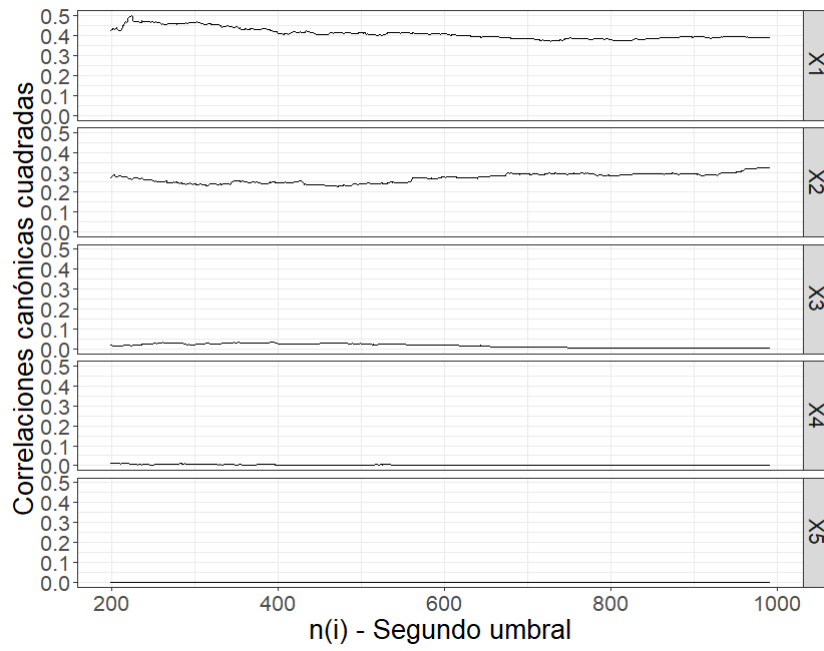


Figura 35: Correlaciones canónicas cuadradas - 2do umbral - $j = 3$

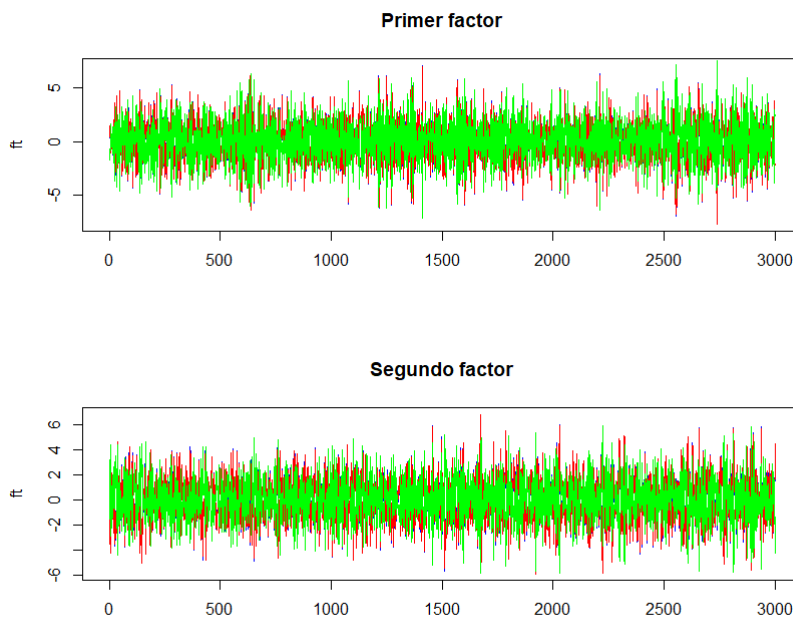


Figura 36: Estimaciones del proceso de factores - $j = 3$

Estimaciones preliminares de Δ para $j = 3$

$$\hat{\Delta}_1 = \begin{pmatrix} -0.87 & -0.45 \\ 0.21 & -0.52 \\ 0.01 & -0.41 \\ 0.27 & -0.41 \\ 0.33 & 0.41 \end{pmatrix} \quad \hat{\Delta}_2 = \begin{pmatrix} -0.85 & -0.40 \\ 0.14 & -0.52 \\ -0.02 & -0.39 \\ 0.29 & -0.40 \\ 0.39 & -0.49 \end{pmatrix} \quad \hat{\Delta}_3 = \begin{pmatrix} 0.86 & 0.41 \\ -0.19 & 0.51 \\ -0.19 & 0.12 \\ -0.24 & 0.33 \\ -0.33 & 0.66 \end{pmatrix}$$

Anexo 4 - Anexos de la sección 6

Código	Nombre
AAPL	Apple Inc.
AMZN	Amazon
GOOGL	Alphabet Class A
JNJ	Johnson & Johnson
JPM	JP Morgan Chase
MSFT	Microsoft
NVDA	NVIDIA
UNH	United Health Group
XOM	Exxon Mobil

Cuadro 10: Acciones del índice S&P 500 usadas

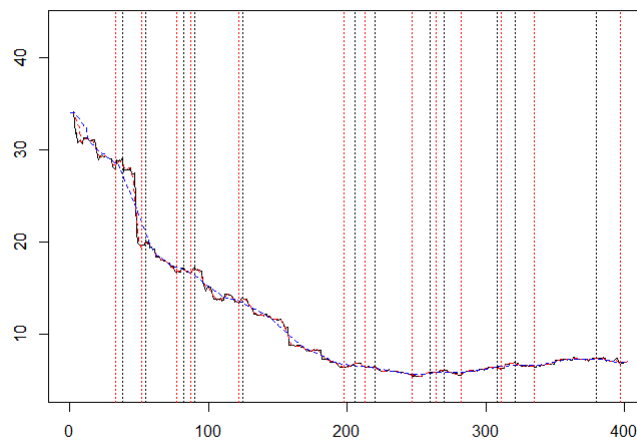


Figura 37: Algoritmo de puntos de giro - Primer valor propio

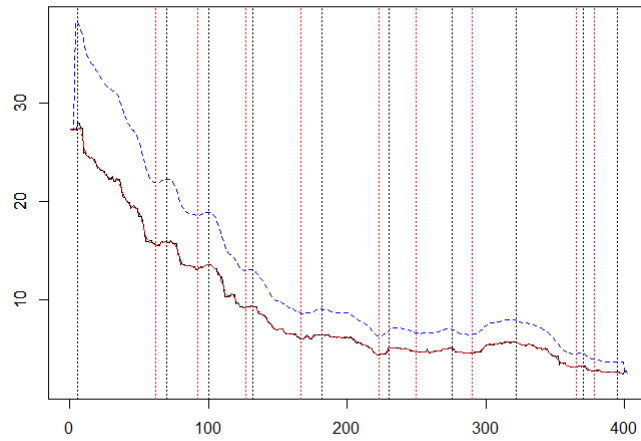


Figura 38: Algoritmo de puntos de giro - Segundo valor propio

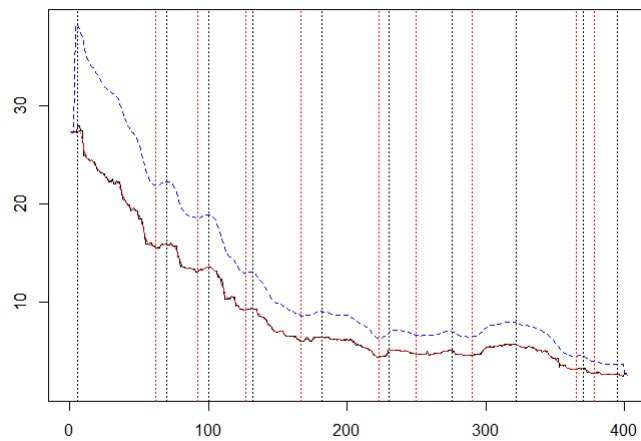


Figura 39: Algoritmo de puntos de giro - Tercer valor propio