

PONTIFICIA UNIVERSIDAD  
CATOLICA DEL PERÚ

Escuela de Posgrado



**Estimación del ciclo financiero utilizando métodos de análisis  
exploratorios: Aplicación a Perú en el periodo 2000-2020.**

Tesis para obtener el grado académico de Magíster en  
Estadística

que presenta:

*Lilian Katherine Atoche Murrieta*

Asesor:  
*Sergio Camiz*

Lima, 2022

## Resumen

La presente tesis tiene por objetivo identificar el ciclo financiero del Perú utilizando series temporales financieras como por ejemplo tasas de interés activas y pasivas, monto de ahorro total en el sistema financiero, la liquidez de las entidades financieras, entre otros. La información tiene frecuencia mensual y abarca los periodos 2000-2020. Para determinar el ciclo financiero se utilizan técnicas de análisis exploratorias. Primero empezamos con el análisis de componentes principales el cual nos permitió identificar si existen componentes comunes entre las variables. Luego utilizamos una clasificación jerárquica (HFC por sus siglas en inglés) para agrupar las variables en grupos homogéneos. Este análisis nos permitió asociar a cada grupo una variable representativa el cual permitió identificar los componentes del ciclo financiero. Finalmente utilizamos una tercera técnica exploratoria conocida como análisis evolutivo el cual tiene por objetivo identificar si la correlación de las series varía a lo largo del tiempo y cómo varía ésta. Los resultados del análisis exploratorio confirman la existencia de tres componentes del ciclo financiero. El primer componente identifica el componente de larga duración del ciclo financiero. Este componente mide la evolución del sistema financiero a lo largo del tiempo. El segundo componente mide la evolución de las tasas activas y pasivas en el sistema financiero el cual es un componente de mediano plazo. El tercer componente podemos asociarlo con la volatilidad que tiene el mercado financiero peruano.

Palabras claves: ciclo financiero, componentes principales, análisis jerárquico, análisis evolutivo, variable representativa.

## Abstract

The aim of this thesis is to identify the financial cycle of Peru using financial time series such as active and passive interest rates, total savings amount in the financial system, liquidity of financial entities, among others. The information is monthly and covers the periods 2000-2020. Exploratory analysis techniques are used to determine the financial cycle. We first started with the principal components analysis which allowed us to identify if there are common components among the variables. We then use a hierarchical classification (HFC) to group the variables into homogeneous groups. This analysis allowed us to associate each group with a representative variable which allowed us to identify the components of the financial cycle. Finally, we use a third exploratory technique known as evolutionary analysis, which aims to identify whether the correlation of the series varies over time. The results of the exploratory analysis confirm the existence of three components of the financial cycle. The first component identifies the long-term component of the financial cycle. This component measures the evolution of the financial system over time. The second component measures the evolution of lending and deposit rates in the financial system, which is a medium-term component. The third component can be associated with the volatility of the Peruvian financial market.

Keywords: financial cycle, principal components analysis, hierarchical analysis, evolutionary analysis, representative variable.

# Índice general

<b>1</b>	<b>Planteamiento y justificación del tema seleccionado</b>	<b>4</b>
1.1.	Determinación del problema . . . . .	4
1.2.	Identificación del problema . . . . .	6
1.3.	Justificación . . . . .	7
<b>2</b>	<b>Objetivos</b>	<b>10</b>
<b>3</b>	<b>Descripción de la metodología de investigación</b>	<b>11</b>
3.1.	Tipo de investigación . . . . .	11
3.2.	Datos . . . . .	11
3.3.	Plan de análisis de datos . . . . .	13
<b>4</b>	<b>Marco Teórico</b>	<b>14</b>
4.1.	Análisis de Componentes principales (PCA) . . . . .	14
4.1.1.	Inercia . . . . .	14
4.1.2.	Racionalidad del PCA . . . . .	15
4.1.3.	Reglas de detención . . . . .	18
4.2.	Clasificación Factorial Jerárquica (HFC) . . . . .	19
4.3.	Análisis Evolutivo (EPCA) . . . . .	21
<b>5</b>	<b>Resultados</b>	<b>23</b>
5.1.	Análisis de componentes principales . . . . .	23
5.1.1.	Análisis del primer componente . . . . .	24
5.1.2.	Análisis del segundo componente . . . . .	25
5.1.3.	Análisis del tercer componente . . . . .	27
5.2.	Clasificación Factorial Jerárquica . . . . .	30
5.2.1.	Estudio de los nodos superiores . . . . .	32
5.2.2.	Estudio de las clases . . . . .	42
5.3.	Síntesis de los resultados . . . . .	48
5.4.	Análisis Evolutivo . . . . .	51
5.4.1.	Análisis evolutivo sobre las 5 variables representativas . . . . .	51
5.4.2.	Análisis evolutivo del primer grupo . . . . .	53
5.4.3.	Análisis evolutivo del segundo grupo . . . . .	56
5.4.4.	Análisis evolutivo del tercer grupo . . . . .	58
<b>6</b>	<b>Conclusiones</b>	<b>62</b>
	<b>Referencias</b>	<b>64</b>

## Capítulo 1

# Planteamiento y justificación del tema seleccionado

### 1.1. Determinación del problema

En esta tesis se estudiará el ciclo financiero de Perú en los años 2000-2020 utilizando métodos de análisis exploratorio. Una definición teórica bastante aceptada en la literatura del ciclo financiero es la que propone Borio (2012): “*el ciclo financiero es el resultado de la interacción entre las percepciones de valor y riesgo, las actitudes de los agentes hacia el riesgo y las restricciones financieras, que se refuerzan entre sí, y se traducen en auges y caídas del sector financiero*”. Borio (2012) señala que las restricciones financieras se reflejan en el volumen de créditos, mientras que las percepciones de valor y riesgo se observan en los precios de las viviendas. El estudio del ciclo financiero ha cobrado mayor interés luego de la crisis financiera 2007-2008 en la cual se observó que los choques que afectaron al mercado financiero se amplificaron y propagaron a la economía agregada (Pérez y Vilchez, 2018). Plašil et al.(2016) consideran que empíricamente la forma más parsimoniosa de medir el ciclo financiero es utilizando el ratio créditos a PBI. Dichos autores estimaron el ciclo financiero para República Checa y concluyeron que el ciclo financiero ayuda a predecir los riesgos crediticio a nivel agregado. Sobre el ratio créditos a PBI se pueden aplicar diferentes técnicas econométricas para extraer el ciclo financiero como por ejemplo el filtro Hodrick-Prescott<sup>1</sup> o Christiano-Fitzgerald<sup>2</sup> (Pérez y Vilchez, 2018). El-Baz (2018) estimó el ciclo financiero en Arabia Saudita considerando sólo la variable de volumen de crédito para la cual se identificaron los periodos de expansión y recesión utilizando el algoritmo de puntos de inflexión de Bry-Boschan<sup>3</sup>, demostrando que los periodos de expansión tienen una duración entre 4-49 trimestres y las contracciones entre 4-9 trimestres. Sin embargo, Drehmann, Borio y Tsatsaronis (2012) sostienen que el ciclo financiero puede medirse sintéticamente utilizando las variables índice de precio de las viviendas, el ratio crédito a PBI y el volumen de créditos. Analizan-

---

<sup>1</sup>El filtro Hodrick-Prescott (HP) es una técnica para suavizar las series de tiempo utilizando un parámetro de suavizamiento ( $\lambda$ ) que depende de la frecuencia de los datos. Una vez estimado se calcula el ciclo de la serie por diferencia entre la serie original y la serie suavizada.

<sup>2</sup>El filtro Christiano-Fitzgerald (CF) es una técnica para extraer ciclos basado en el análisis espectral de la serie. La ventaja de esta técnica es que permite extraer ciclos con diferentes frecuencias. Por ejemplo, en la literatura se conoce que el ciclo económico de una serie tiene frecuencias entre 8 y 32 trimestres, mientras que el ciclo financiero tiene frecuencias entre 32 y 120 trimestres.

<sup>3</sup>El algoritmo busca puntos de inflexión (picos/valles) en una serie  $y_t$ . Se define que en la serie  $y_t$  ocurre un pico en  $t$  si  $y_{t-k}, \dots, y_{t-k+1} < y(t) < y_{t+1}, \dots, y_{t+k}$  donde  $k$  debe establecerse, por ejemplo,  $k = 2$  para datos trimestrales,  $k = 5$  para datos mensuales y  $k = 1$  para datos anuales.  $k$  se denomina parámetro de *ventana simétrica*. Así mismo, las *fases* (de pico a valle o de valle a pico) y *ciclos* (de pico a pico o de valle a valle) deben tener una duración mínima introducidas por el investigador.

do los datos de siete países (Australia, Alemania, Japón, Noruega, Suecia, Reino Unido y Estados Unidos) los autores extraen las frecuencias de mediano plazo (32-120 trimestres) de las series utilizando el filtro Christiano-Fitzgerald y las promedian para obtener un indicador agregado del ciclo financiero. Por su parte, Shen et al.(2018) crean un índice del ciclo financiero agregado para China empleando el ratio crédito a PBI, el precio de viviendas y el índice de acciones. En cada serie se identifican puntos de inflexión (puntos de las series que son máximos o mínimos locales) con la metodología de Harding y Pagan (2002)<sup>4</sup>. Luego de obtener los puntos de inflexión en cada serie se agregan según la metodología de Drehmann et al.(2012)<sup>5</sup>. Claessens et al.(2011) estiman el ciclo financiero para 21 economías de la OCDE utilizando tres series: volumen de crédito, precio de viviendas y precio de acciones. Utilizando el algoritmo de Harding y Pagan de puntos de inflexión demuestran que las recesiones de los ciclos financieros (representados por esas tres variables) tienen una duración entre 5-8 trimestres y las expansiones tienen mayor duración dependiendo de la serie que se analice: en el precio de las acciones la duración es en promedio 22 trimestres, en el precio de viviendas 14 trimestres y el volumen de créditos 8 trimestres. Plašil et al.(2016) estiman el ciclo financiero para la República Checa utilizando variables que capturen las *percepciones* de valor y riesgo como la cantidad de nuevos préstamos bancarios, índice de precio de viviendas, diferencial de tasas de interés, índice de acciones, ratio deuda a ingreso de los hogares y el déficit de cuenta corriente (% PBI). Luego de estandarizar las variables, se agregan en un solo indicador utilizando la metodología de Holló et al.(2012) el cual considera la correlación por pares de variables y un peso que mide la importancia de cada variable del indicador agregado<sup>6</sup>. Estos pesos son calculados en base a 30,000 simulaciones y se elige la combinación de pesos que realice la mejor predicción de las pérdidas crediticias seis trimestres hacia adelante. Existen otras metodologías de estimación para identificar el ciclo financiero como por ejemplo, modelos de factores dinámicos (Menden y Proaño, 2017; Adarov, 2018; Ramos, 2020), modelos

<sup>4</sup>El algoritmo de puntos de inflexión de Harding y Pagan (2002) es una modificación del algoritmo de Bry-Boschan para datos anuales. Se define que en la serie  $y_t$  ocurre un pico en  $t$  si  $[(y_t - y_{t-1}) > 0, (y_t - y_{t-2}) > 0]$  y  $[(y_t - y_{t+1}) > 0, (y_t - y_{t+2}) > 0]$ . De manera similar, un valle ocurre en  $t$  si  $[(y_t - y_{t-1}) < 0, (y_t - y_{t-2}) < 0]$  y  $[(y_t - y_{t+1}) < 0, (y_t - y_{t+2}) < 0]$ . Shen et al.(2018) asumieron una duración del ciclo de al menos 3 años.

<sup>5</sup>La metodología consiste en los siguientes pasos: **Paso 1** Calcular el número mínimo de años hasta el pico (valle) el cual denotaremos como  $DP_t^{Y_i}$  ( $DT_t^{Y_i}$ ) para cada serie  $Y_i$  y cada punto de tiempo  $t$ , **Paso 2** Calculamos la mediana  $MP_t(MT_t)$  a través de todas  $DP_t^{Y_i}$  ( $DT_t^{Y_i}$ ), y **Paso 3** Buscar puntos mínimos locales en  $MP_t(MT_t)$ . Estos son puntos de inflexión del ciclo común si los picos (valles) de todas las series individuales no están más de 3 años de distancia del pico (valle) común identificado en el paso 2.

<sup>6</sup>El método de agregación puede expresarse en la siguiente fórmula:

$$Indice = (w^o s_t)' C_t (w^o s_t) ,$$

donde  $w = (w_1, w_2, \dots, w_n)$  es un vector de pesos e indica la importancia relativa de las  $n$  variables individuales,  $s_t = (s_{1,t}, s_{2,t}, \dots, s_{n,t})$  es el vector de subindicadores en el tiempo  $t$ , y  $(w \circ s_t)$  representa la multiplicación elemento por elemento de estos vectores (también conocido como el producto de Hadamard). La matriz  $C_t$  contiene los valores de los coeficientes de correlación por pares  $\rho_{t,ij}$  entre las variables  $i$  y  $j$  en el tiempo  $t$ . Los coeficientes de correlación se estiman de forma recursiva utilizando el método *EWMA* con factor de suavizado  $\lambda = 0,94$ :

$$\begin{aligned} \sigma_{t,ij} &= \lambda \sigma_{t-1,ij} + (1 - \lambda) \tilde{s}_{t,i} \tilde{s}_{t,j} \\ \sigma_{t,i}^2 &= \lambda \sigma_{t-1,i}^2 + (1 - \lambda) \tilde{s}_{t,i} \tilde{s}_{t,i} \\ \rho_{t,ij} &= \sigma_{t,ij} / (\sigma_{t,i} \sigma_{t,j}) \\ \tilde{s}_{t,i} &= (s_{t,i} - 0,5) \end{aligned}$$

multivariados estructurales (Rüinstler y Vlekke, 2016) y componentes no observables (Galati et al., 2016).

Por otro lado, también se han empleado metodologías de exploración (principalmente análisis de componentes principales - PCA) para medir el ciclo financiero de un país. Krznar y Matheson (2017) separan la medición del ciclo financiero entre un enfoque de mediano plazo y otro de corto plazo. Con respecto al primero, se utiliza el volumen de crédito al cual se le aplica el filtro econométrico de Hodrick-Prescott para obtener el ciclo financiero, mientras que en el segundo enfoque se utilizan nueve variables financieras (diferencial de tasas de interés, precio de acciones, precios de vivienda, crédito total, Embig, tipo de cambio y tasas de interés) a las cuales se extrae el primer componente principal y se le aplica el filtro de Christiano y Fitzgerald para encontrar el ciclo financiero en frecuencias de 4.75-50 años. Karfakis y Karfaki (2018) estiman el ciclo financiero para Grecia utilizando componentes principales aplicado a las series del crédito al sector no financiero privado y el precio de viviendas. Al igual que Borio (2012), estas variables reflejan los cambios en las percepciones y actitudes sobre el riesgo financiero. Strohsal et al.(2017) miden el ciclo financiero en Estados Unidos y Reino Unido como el primer componente principal de las series precio de vivienda y volumen de créditos. Los autores consideran que desde 1985 la duración de los ciclos financieros han aumentando pasando de una duración de 6 años a 14.7 años. Ma y Zhang (2016) estiman el ciclo financiero para las cuatro principales economías del mundo: Estados Unidos, China, Japón y Reino Unido. Los autores utilizan PCA a partir de un conjunto de series financieras tales como tipo de cambio real, crecimiento del dinero (M2), precio de viviendas, precio de acciones, spread bancario, tasas de interés de largo plazo y prima por riesgo.

Como se ha expuesto anteriormente, no hay una metodología estándar en la literatura acerca de cómo medir el ciclo financiero de un país es por ello que se han adoptado diferentes mediciones dada la naturaleza abstracta del ciclo financiero (Menden y Proaño, 2017). En el caso de las series utilizadas, muchos autores suelen utilizar el volumen de créditos y el precio de acciones de un país. Sin embargo, como señala Schüler, Hiebert y Peltonen (2015), idealmente el ciclo financiero debería resumir el comovimiento a lo largo del tiempo de una gama de variables del sector financiero que cubran variables de precio y cantidades. De igual manera, Stremmel (2015) indica que para medir los ciclos financieros debe utilizarse la mayor cantidad de información financiera posible. Por tal motivo, como parte del trabajo exploratorio se utilizará todo el conjunto de información financiera disponible y relevante para medir el ciclo financiero del Perú.

## **1.2. Identificación del problema**

En el Perú, la literatura relacionada al ciclo financiero es escasa. Pérez y Vilchez (2018) analizan la evolución del ciclo financiero en el Perú durante el periodo 1994-2018 utilizando el ratio crédito a PBI. A partir de esta variable, los autores estiman el ciclo financiero utilizando filtros econométricos como el filtro de Hodrick-Prescott y el filtro de Christiano-Fitzgerald, y evidencian que hubo un periodo expansivo durante los periodos 1994-1998, luego de este periodo el ciclo financiero mostró una caída y posteriormente una recuperación a mediados del año 2000 manteniéndose esta tendencia hasta los años 2013-2014. Por otro lado, Lahura

et al.(2013) estiman el ciclo financiero utilizando el filtro de Kalman<sup>7</sup> considerando en la *ecuación de observación* como variable dependiente el ratio crédito a PIB y como variables explicativas el logaritmo del PBI, el logaritmo de los términos de intercambio y un componente no observable que captura las desviaciones del crédito respecto a sus fundamentos macroeconómicos. Esta última variable representa una medida del ciclo financiero y se asume que en la *ecuación de estado* sigue un proceso autorregresivo de orden uno. Los autores determinaron dos fases de auge crediticio comprendidos entre finales de 2008 e inicios de 2009 y entre abril y junio de 2019. Un estudio más reciente se encuentra en Ramos (2020) quien tiene por objetivo estimar el ciclo financiero y analizar la relación de causalidad con el ciclo económico en el Perú. El autor emplea el método de factores dinámicos para estimar el ciclo financiero considerando cuatro variables representativas de cada mercado financiero del Perú: Volumen de créditos al sector privado (mercado de créditos), Índice General de la Bolsa de Valores de Lima (mercado de acciones), Embig Perú (mercado de bonos) y el tipo de cambio (mercado de divisas). El autor encuentra que existen dos componentes del ciclo financiero, uno de mediano plazo y otro de corto plazo. El primero alcanza su valor mínimo en abril del 2007, periodo anterior a la crisis financiera del 2008; mientras que el segundo lo hace en febrero del 2009. En cuanto a la relación de la causalidad, el autor evidencia que hay una causalidad del primer componente del ciclo financiero al ciclo económico en frecuencias de mediano plazo (mayores a 10 trimestres) y que el ciclo económico no causa a ningún componente del ciclo financiero.

Ramos (2020) considera que los estudios del ciclo financiero en el Perú se han enfocado en estudiar solo las variables crediticias, dejando de lado el resto de variables financieras. Es por ello que este autor sugiere utilizar una variable que representa a cada uno de los cuatro mercados financieros. Sin embargo, a pesar de utilizar diferentes series financieras, no se ha encontrado en la literatura un análisis exploratorio que permita determinar cuáles son las variables más importantes que puedan representar adecuadamente el ciclo financiero y las características de las series analizadas.

La presente tesis hará uso de diferentes series financieras disponibles y de técnicas exploratorias que servirán para futuros trabajos de investigación para determinar patrones en las variables financieras y las características del ciclo financiero del Perú.

### 1.3. Justificación

Durante la crisis financiera del 2007-2008 quedó demostrado que las fluctuaciones de los mercados financieros tienen repercusiones sobre la actividad económica. Durante esos

---

<sup>7</sup>El filtro de Kalman es el principal algoritmo para estimar sistemas dinámicos especificados en la forma de estado-espacio (*state-space*). La primera ecuación es una ecuación de medida, también conocida como ecuación de *señal* o de *observación*:

$$y_t = H_t\beta_t + Z_tA_t + \epsilon_t$$

donde:  $y_t$  es un vector de variables observables de orden  $n \times 1$ ,  $H_t$  es una matriz de orden  $n \times m$ ,  $Z_t$  es una matriz de variables exógenas de orden  $n \times k$ ,  $\beta_t$  es un vector de variables de estado no observables de orden  $m \times 1$ ,  $A_t$  es una matriz de parámetros  $k \times 1$  y  $\epsilon_t$  es el error de observación. La segunda es la *ecuación de estado* (variables *no observables*):

$$\beta_t = \mu_t + F_t\beta_{t-1} + S_tv_t$$

donde:  $F_t$  es una matriz  $m \times m$ ,  $\mu_t$  es un vector  $m \times 1$ ,  $S_t$  es una matriz  $m \times g$  y  $v_t$  es un vector de perturbaciones no correlacionadas  $m \times g$ .

años, Estados Unidos mostró que una caída del precio de las viviendas generó el aumento de las cuotas de los créditos hipotecarios que a su vez ocasionó la disminución del ingreso disponible de las familias impactando en un menor dinamismo del consumo de los hogares y una reducción de la inversión residencial (De la Cuba, 2009). La recesión de Estados Unidos repercutió sobre el resto de economías desarrolladas y emergentes terminando en una crisis financiera mundial. En el Perú el impacto de la crisis financieras se dio lugar durante los primeros trimestres del 2009 pasando de un crecimiento de 2% en el primer trimestre del 2009 a una caída de 0.4% en el tercer trimestre del 2009. Luego de la crisis financiera, se ha prestado mayor atención a los mercados financieros y sus vínculos con la macroeconomía. Al respecto, se regresó al estudio del ciclo financiero: un concepto que fue introducido en la década de los 90's (Cagliarini y Price, 2017) y definido inicialmente por Borio et al.(2001) como la expansión y posterior contracción del volumen de crédito y el precio de activos, acompañados por cambios en los estándares crediticios. Las propiedades del ciclo financiero han sido estudiadas ampliamente entre las cuales destacan: mayor duración con respecto al ciclo económico, predictor de crisis financieras, alto grado de sincronización con el ciclo económico, causalidad al ciclo económico (Ramos, 2020) y como indicador adelantado para predecir el crecimiento económico (Krznar y Matheson, 2017). Debido a las propiedades que posee el ciclo financiero, resulta importante estudiar esta variable para la formulación de políticas monetarias y macroprudenciales de un país. Con respecto a las políticas monetarias, algunos autores están a favor de que el Banco Central debe ajustar su tasa de interés de referencia ante desequilibrios financieros (Billio y Petronevich, 2017). Sin embargo, Cagliarini y Price (2017) señalan que no es recomendable que la política monetaria soporte la carga del ciclo financiero porque podría implicar grandes cambios en las tasas de interés lo cual podría ser costoso e ineficiente para una economía. Así mismo, Smets (2013) señala que un Banco Central puede perder credibilidad al mantener dos objetivos: estabilidad financiera (control del ciclo financiero) y otro de estabilidad monetaria (control de la inflación). En Perú, Ramos (2020) evidenció que la reacción del Banco Central al ciclo financiero no genera aumento de bienestar social a la economía. Cagliarini y Price (2017) indica que la política monetaria por sí sola no puede gestionar el ciclo financiero, sino que debe apoyarse en las políticas macroprudenciales. Estas últimas tienen por objetivo fortalecer la resistencia del sistema financiero y moderar la prociclicidad del ciclo financiero debido a que un sistema financiero mal regulado puede generar pérdidas potenciales en términos empleo y producción (Rojas, 2010). Choy y Chang (2014) señalan que las políticas macroprudenciales de Perú han tenido por objetivo limitar el crecimiento excesivo del crédito y precio de las viviendas, evitar la falta de liquidez de los bancos, disminuir la prociclicidad entre las variables financieras y económicas, reducir la volatilidad del tipo de cambio y los movimientos abruptos de los flujos de capitales y limitar el volumen de operaciones de las AFP's en el exterior.

En conclusión, el estudio y análisis del ciclo financiero de un país es importante porque advierte la formación de desequilibrios financieros que pueden generar impactos negativos en la economía como por ejemplo pérdida de empleo, inflación y caída de producción. Con la identificación del ciclo financiero, pueden adoptarse diferentes políticas para reducir su volatilidad e impacto. Por tales motivos, es importante estudiar el ciclo financiero del Perú

con un análisis exploratorio para que sirva como punto de partida para futuras investigaciones en cuanto a la selección de variables que pueden representar el ciclo financiero y su posterior estimación, así como la identificación de la relación entre las variables a lo largo del tiempo. Además, la identificación del ciclo financiero peruano servirá como un indicador de alerta temprana para las autoridades pertinentes de modo que se adopten las herramientas monetarias o macroprudenciales que permitan alcanzar una estabilidad macrofinanciera.

## Capítulo 2

### Objetivos

Por lo expuesto en la sección anterior, el objetivo de la tesis es identificar el ciclo financiero del Perú para el periodo 2000-2020 utilizando métodos de análisis de datos exploratorios. Empezaremos recopilando las series temporales más relevantes y disponibles para representar el mercado financiero peruano. En primer lugar se analizarán dichas series a través del análisis de componentes principales para identificar si hay componentes comunes entre ellas y cuáles pueden identificar el ciclo financiero peruano. Así mismo, se agregarán las series en grupos homogéneos utilizando una clasificación factorial jerárquica (HFC, por sus siglas en inglés) para conseguir grupos homogéneos bajo el punto de vista de su correlación. A cada grupo se le asociará una variable representativa y también un componente principal el cual permitirá representar los ciclos financieros, económicos entre otros.

Los resultados de los dos análisis previos se compararán: el objetivo es averiguar las relaciones entre los factores obtenidos del PCA y las variables representativas y componentes principales obtenidos de los grupos formados. Posteriormente, se analizarán las series de manera evolutiva, es decir aplicando el análisis en componentes principales en una ventana móvil, para ver la estabilidad de las relaciones ya encontradas o cómo estas varían a lo largo del tiempo.

## Capítulo 3

# Descripción de la metodología de investigación

### 3.1. Tipo de investigación

El tipo de estudio es exploratorio debido a que el objetivo de la tesis es identificar las características del ciclo financiero del Perú utilizando series de tiempo comprendidas entre el 2000-2020 con periodicidad mensual. No obstante, este análisis está acompañado de una revisión teórica de la definición, propiedades e importancia del ciclo financiero. La selección de variables es producto de la crítica de diferentes autores que indican que el ciclo financiero no puede representarse sólo con variables de crédito sino con un conjunto de series financieras que representen el mercado financiero peruano.

### 3.2. Datos

Se utilizan series de tiempo disponibles en el Banco Central de Reservas del Perú. Desde el sitio web <https://estadisticas.bcrp.gob.pe/estadisticas/series/> se descargaron datos mensuales desde el periodo 2000 hasta 2020 para 17 series el cual constituye la tabla de datos **X**. Las series financieras elegidas se encuentran en la tabla 3.1.

Cuadro 3.1: Descripción de las variables

Nº	Descripción	id
1	Activos internos netos de empresas bancarias (millones S/)	actnt
2	Circulante (millones S/.)	clnt
3	Ahorro total en el Sistema Financiero (millones S/.)	aho
4	Monto de encaje (millones S/.)	encj
5	Crédito total en el Sistema Financiero (millones S/.)	cred
6	Índice General Bolsa Valores de Lima	igbvl
7	Valor de los fondos de las AFPs (millones S/.)	spp
8	Tipo de cambio promedio interbancario (S/. Por USD)	tc
9	Liquidez de las empresas bancarias (millones S/.)	liq
10	Coefficiente de dolarización (%)	dol
11	Reservas Internacionales Netas (millones S/.)	rin
12	Palanca global de las empresas bancarias	plca
13	Tasa interés activa en moneda nacional	tamn
14	Tasa interés pasiva en moneda nacional	tpmn
15	Tasa interés activa en moneda extranjera	tamex
16	Tasa interés pasiva en moneda extranjera	tpmex
17	Embig Perú (pbs)	embig

De acuerdo a la literatura, las principales series de tiempo utilizadas para estimar el ciclo financiero de un país son el volumen de créditos, precio de viviendas y el precio de acciones (Claessens et al., 2011; Borio, 2012; Shen et al., 2018). También existen investigaciones que suelen considerar el ratio de crédito a PBI (Drehmann et al., 2012; Pérez y Vilchez, 2018). Sin embargo, para el caso peruano debemos tener en cuenta que el mercado financiero en Perú está determinado principalmente por un sistema bancario concentrado en cuatro entidades bancarias y un mercado de bonos poco profundo. Además existen una limitación en las estadísticas de precio de viviendas debido a que se concentra solo para distritos de Lima y la profundidad de la información es corta. Por tal motivo, al igual que Ramos (2020) se eligieron las cuatro principales series de tiempo que representan el mercado financiero peruano: volumen de créditos (mercado crediticio), Índice General Bolsa Valores de Lima (mercado de capitales), tipo de cambio (mercado cambiario) y EMBIG Perú<sup>1</sup> (mercado de bonos). Adicionalmente, para representar las inversiones de las AFPs se eligió la serie del valor de fondo de las AFPs<sup>2</sup>. Por otro lado, Stremmel (2015) sugiere utilizar información relevante de las entidades bancarias para estimar el ciclo financiero. En ese sentido, estamos incluyendo variables como liquidez bancaria<sup>3</sup> y monto del encaje bancario<sup>4</sup>. Adicionalmente hemos considerando otras variables relevantes como el nivel de activos de las empresas bancarias y la palanca global. Otro indicador a tener en cuenta es el monto del circulante ya que mide la cantidad de dinero en efectivo que mantienen las personas en su poder. Esta variable es importante en periodos de crisis pues las personas tienden a demandar más efectivo mientras el PBI se desacelera (Ramírez y Vásquez, 2014). El coeficiente de dolarización también es un indicador importante del mercado financiero porque permite conocer la magnitud de una crisis cambiaria si el país mantiene un elevado nivel de deuda en moneda extranjera. Finalmente, se eligieron otras series que pueden impactar sobre las decisiones de ahorro e inversión de los consumidores peruanos como las tasas de interés activas y pasivas y monto del ahorro total en el Sistema Financiero.

En la tabla 3.2 se encuentran las principales estadísticas de dichas series analizadas:

---

<sup>1</sup>El EMBIG es la abreviatura de *Emerging Market Bond Index* el cual mide el diferencial de rendimientos de los bonos soberanos de Perú y los bonos libre de riesgo (se asumen que son los bonos soberanos de Estados Unidos). Esta diferencia es una medida del riesgo país: a mayor valor de la diferencia significa que los bonos de Perú tienen mayor riesgo, es decir tienen riesgo de pago.

<sup>2</sup>El Valor de Fondo de las AFPs mide el valor del portafolio que tienen invertido las AFPs en el Perú y en el extranjero.

<sup>3</sup>De acuerdo al glosario de términos del Banco Central de Reservas, la liquidez bancaria mide el dinero que poseen los bancos para afrontar sus operaciones diarias y posibles crisis financieras.

<sup>4</sup>El encaje es la cantidad de dinero proveniente de los pasivos que los bancos están obligados a mantener en las cuentas del Banco Central de Reservas. Es una medida regulatoria que tiene por objetivo reducir la expansión del crédito en el Sistema Financiero.

Cuadro 3.2: Estadísticos resumen

Variable	Mínimo	Q1	Q2	Q3	Máximo	Promedio	Desv.Estd.
actnt	40,968.92	44,798.89	101,803.85	215,313.59	340,943.08	132,627.71	88,870.17
clnt	3,457.89	6,445.12	18,468.51	37,784.99	67,803.70	22,878.83	16,963.53
aho	41,650.07	66,280.73	160,245.14	276,031.72	497,445.26	190,464.66	133,160.95
encj	923.27	1,241.62	4,600.09	10,602.89	25,364.27	6,643.16	5,746.79
cred	50,534.76	55,502.26	119,363.61	270,358.80	423,853.94	163,457.03	115,290.09
igbvl	1,134.93	2,906.37	14,129.00	18,453.00	23,612.02	11,819.12	7,528.08
spp	5,753.64	23,200.96	67,196.11	118,960.31	174,617.51	74,689.18	52,355.62
tc	2.55	2.90	3.26	3.44	3.62	3.19	0.29
liq	37,864.63	43,343.22	91,925.68	180,155.99	316,320.06	117,111.30	76,355.94
dol	37.97	48.01	54.62	73.13	79.91	58.60	13.30
rin	28,302.85	37,514.75	99,889.79	191,231.66	264,324.01	117,127.22	74,824.28
plca	6.27	7.14	7.58	8.24	15.78	8.51	2.60
tamn	12.29	16.24	19.79	23.88	37.12	20.74	5.32
tpmn	1.06	2.33	2.57	3.34	13.63	3.57	2.63
tamex	6.11	7.77	8.71	10.45	17.08	9.43	2.32
tpmex	0.28	0.62	0.98	1.99	5.37	1.51	1.34
embig	103.95	154.00	195.00	372.92	816.27	281.90	182.63

### 3.3. Plan de análisis de datos

En la primera etapa de la investigación se recoge la información a través de la página web del BCRP. Una vez que se obtengan los datos, serán exportados al software R para el análisis estadístico exploratorio. En una primera fase, el análisis de datos consiste en un análisis exploratorio de las series de tiempo a través del análisis de componentes principales y de la clasificación jerárquica factorial. De acuerdo a la clasificación de las series, se estudiará cada grupo encontrado con el análisis de componentes principales. Luego, se utilizará el análisis evolutivo para determinar la estabilidad o variabilidad de la estructura de correlación estudiada y se detectarán, si existen, los periodos en los cuales la estructura sea más inestable.

Es importante mencionar que como se quiere interpretar los resultados en términos de ciclo financiero del Perú, todo este estudio es orientado a poder interpretar los resultados conseguidos, sean los factores o variables representativas como estimaciones del ciclo financiero. Por lo tanto, el estudio tiene que considerarse evolutivo, en el sentido que, bajo los resultados conseguidos para las análisis, se podrán considerar en seguida otros métodos para lograr una mejor identificación del ciclo financiero.

## Capítulo 4

### Marco Teórico

#### 4.1. Análisis de Componentes principales (PCA)

El análisis de componentes principales (PCA) es el método de análisis exploratorio más conocido y útil al tratar con datos cuantitativos, ya que ayuda a delinear las características más relevantes del conjunto de datos analizados, teniendo en cuenta la estructura de correlación entre las variables consideradas y la detección de valores atípicos dentro de las unidades. De hecho, PCA es una herramienta para estructurar la información contenida en una tabla de datos  $\mathbf{X}$  con  $n$  filas o unidades y  $p$  columnas o variables, donde con  $n \gg p$ , en una jerarquía de importancia, desde la más relevante hasta la menos relevante.

##### 4.1.1. Inercia

Suponemos que la información que se encuentra en la tabla de datos  $\mathbf{X}$  se mide por su *inercia*. La inercia de un punto con respecto a otro es su distancia al cuadrado ponderada, de modo que la *inercia de una nube de puntos*  $\mathcal{N}$  con respecto a un punto  $P$  es la suma de sus distancias cuadradas a ella y ponderada.

Por lo general, representamos las  $n$  unidades de fila  $\mathbf{x}_i$  de  $\mathbf{X}$  en un *espacio geométrico de variables afines*  $\mathbb{R}^p$ , generado por  $p$  columnas-variables  $\mathbf{X}_j$  de  $\mathbf{X}$ , formando una base ortogonal. En la correspondiente nube de puntos  $\mathcal{N}_X$  las distancias entre puntos-unidades pueden interpretarse como disimilitudes, en el sentido que se estará asumiendo que los puntos próximos corresponden a unidades con valores parecidos de las variables observadas, mientras que los puntos lejanos se refieren a valores muy diferentes. Por consecuencia, la distancia se interpreta como una medida de disimilitud. Así, la inercia de la tabla corresponde a la inercia con respecto al baricentro de la nube  $\mathcal{N}_X$ , ya que, como los datos se suponen estandarizados, esto se asume como origen. De hecho, es fácil de demostrar que dicha inercia es mínima con respecto a cualquier otro punto. De manera análoga, las  $p$  variables pueden representarse en el *espacio de unidades*  $\mathbb{R}^n$  generado por las unidades de fila. Su nube  $\mathcal{N}'_X$  generalmente está formada por vectores que comienzan desde el origen y los cosenos de sus ángulos pueden interpretarse como correlaciones. La misma inercia resulta para la nube de los extremos de los  $p$  vectores-variables (de longitud 1) sobre la hipersfera unitaria y por tanto la inercia con respecto al origen de esta nube, igual a la de  $\mathcal{N}_X$ , vale  $p$ . La inercia sobre cada eje se corresponde a la suma de cuadrados de las correlaciones variables-factor. La representación de las variables en un plano se resulta como proyección sobre un círculo conocido como el círculo de correlación de los vectores unitarios que le corresponden.

*PCA* tiene como objetivo identificar *ejes factoriales* ortogonales, a lo largo de los cuales las coordenadas de *unidades* son *componentes principales*, que dividen de manera óptima la inercia en componentes ortogonales de cantidad decreciente. La optimalidad consiste en maximizar la inercia mientras se extraen los ejes. De esta forma, se lleva al investigador a estudiar e interpretar los ejes factoriales partiendo de los de mayor inercia, por consiguiente, los más informativos, hasta los menos relevantes.

#### 4.1.2. Racionalidad del PCA

Los conceptos básicos de *PCA* resultan de los siguientes teoremas.

**Teorema 4.1.1** (Descomposición en valores singulares). *Cualquier matriz  $n \times p$  real  $\mathbf{X}$  puede descomponerse como*

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$$

donde:

1.  $\mathbf{U}$  es una matriz ortogonal real  $n \times n$ , es decir, tal que  $\mathbf{U}'\mathbf{U} = \mathbf{I}$ ;
2.  $\mathbf{\Sigma}$  es una matriz real diagonal no negativa  $n \times p$ , cuyas entradas diagonales  $\sigma_i$  se denominan valores singulares de  $\mathbf{X}$ ;
3.  $\mathbf{V}$  es una matriz ortogonal real  $p \times p$ .

*Demostración.* La prueba se puede encontrar en Golub y Van Loan (2013, pag 76). □

**Teorema 4.1.2** (Descomposición en valores y vectores propios). *Cualquier matriz simétrica  $p \times p$  real  $\mathbf{M}$  puede descomponerse como*

$$\mathbf{M} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'$$

donde:

1.  $\mathbf{V}$  es una  $p \times p$  ortogonal real matriz de vectores propios de  $\mathbf{M}$ , tal que  $\mathbf{V}'\mathbf{V} = \mathbf{I}$ ;
2.  $\mathbf{\Lambda}$  es una matriz real diagonal no negativa  $p \times p$ , cuyas entradas diagonales  $\lambda$  se denominan valores propios de  $\mathbf{M}$ .

*Demostración del teorema 4.1.2.* Aplicando el teorema 4.1.1 a  $\mathbf{M}$ : ya que es simétrico, entonces  $\mathbf{M} = \mathbf{M}'$  y consecuentemente  $\mathbf{M}'\mathbf{M}$  y  $\mathbf{M}\mathbf{M}'$  son iguales, por lo tanto  $\mathbf{U} = \mathbf{V}$  y  $\mathbf{M} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'$ , con matriz diagonal  $\mathbf{\Lambda}$ . Entonces  $\mathbf{M}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}$  y  $\mathbf{\Lambda}$  son valores propios de  $\mathbf{M}$ . □

Es fácil mostrar una relación especial que vincula ambos teoremas en el caso especial en el que  $\mathbf{M} = \mathbf{X}'\mathbf{X}$ . De hecho, dada cualquier matriz  $\mathbf{X}$ , tanto  $\mathbf{X}'\mathbf{X}$  como  $\mathbf{X}\mathbf{X}'$  son simétricas y sus descomposiciones propias resultan del teorema, como sigue:

$$\begin{aligned}\mathbf{X}'\mathbf{X} &= \mathbf{V}\mathbf{\Sigma}\mathbf{U}'\mathbf{U}\mathbf{\Sigma}\mathbf{V}' = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}' = \mathbf{V}\mathbf{\Lambda}\mathbf{V}' = \mathbf{M} \\ \mathbf{X}\mathbf{X}' &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}'\mathbf{V}\mathbf{\Sigma}\mathbf{U}' = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}' = \mathbf{U}\mathbf{\Lambda}\mathbf{U}' = \mathbf{N},\end{aligned}$$

dadas las condiciones en  $\mathbf{U}$  y  $\mathbf{V}$ . Por lo tanto, los valores singulares  $\Sigma$  de  $\mathbf{X}$  son las raíces cuadradas de los valores propios  $\Lambda$  de ambos  $\mathbf{X}'\mathbf{X} = \mathbf{M}$  y  $\mathbf{X}\mathbf{X}' = \mathbf{N}$ ,  $\mathbf{V}$  es la matriz de vectores propios de  $\mathbf{M} = \mathbf{X}'\mathbf{X}$  y  $\mathbf{U}$  es uno de los vectores propios de  $\mathbf{N} = \mathbf{X}\mathbf{X}'$ . Tenga en cuenta también que, de  $\mathbf{X}'\mathbf{X}\mathbf{V} = \mathbf{V}\Sigma^2$ , al multiplicar previamente por  $\mathbf{X}$ , resulta  $\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{V} = \mathbf{X}\mathbf{V}\Sigma^2$ . Por lo tanto,  $\mathbf{X}\mathbf{V}$  es una matriz de autovectores de  $\mathbf{X}\mathbf{X}'$  y recíprocamente  $\mathbf{X}'\mathbf{U}$  es una matriz de autovectores de  $\mathbf{X}'\mathbf{X}$ . Notar también que, dado que  $\mathbf{X}\mathbf{V} = \mathbf{V}\Sigma$  y  $\mathbf{X}'\mathbf{U} = \mathbf{U}\Sigma$ , se deduce que las normas de estos vectores propios son los valores propios correspondientes y, por tanto,

$$\mathbf{U} = \Sigma^{-1}\mathbf{X}\mathbf{V}, \quad \mathbf{V}' = \Sigma^{-1}\mathbf{X}'\mathbf{U}.$$

En  $\mathbb{R}^p$  la matriz  $\mathbf{V}$  representa el cambio de base de las variables originales a la nueva base ortogonal compuesta por los autovectores. Así, los elementos  $v_{j\alpha}$  de una columna son los componentes del vector propio unitario  $\mathbf{v}_\alpha$  en el conjunto de generadores formado por las variables originales, sus cuadrados suman 1,  $\sum_j v_{j\alpha}^2 = 1$ , y se multiplican por los valores de cada unidad de cada variable para obtener la coordenada de la unidad en este eje. Así, un aumento unitario de esos valores se transforma en un aumento dado por los componentes. Por esta razón se consideran *contribuciones* de una variable a la formación de un eje. Pueden ocurrir dos situaciones extremas: todas las variables o características contribuyen por igual (por alrededor de  $1/p$  cada uno) a la formación del eje, de modo que resume una parte común de todos ellos; o una variable contribuye exclusivamente al eje.

Tal como se señaló anteriormente, los componentes principales son las coordenadas de las unidades en los ejes factoriales: estas se obtienen transformando la matriz original  $\mathbf{X}$  mediante la matriz de vectores propios:  $\mathbf{C} = \mathbf{X}\mathbf{V}$ . Note que esta transformación da implícitamente la inercia de los componentes principales igual a los valores propios correspondientes, ya que  $\mathbf{X}\mathbf{V} = \mathbf{V}\Lambda$  y  $\mathbf{V}$  es unitaria. Como las variables no cambian, pero la base sí y es ortogonal, su distancia al centroide de las variables se descompone de acuerdo con las coordenadas de los distintos ejes. Como la base es ortogonal, la distancia al cuadrado de cada variable al centroide es la suma de las coordenadas al cuadrado  $d_i^2 = \sum_\alpha c_{i\alpha}^2$ . Por lo tanto, las razones  $q_{i\alpha} = c_{i\alpha}^2/d_i^2$  suman 1 y representa la parte de inercia del punto a lo largo de cada eje. Esto puede entenderse como una *calidad de representación* de una variable en el eje y puede acumularse con las de otros ejes para dar la calidad de representación de una unidad en un plano, un espacio, etc.

Si ahora configuramos en  $\mathbb{R}^n$  la matriz  $\mathbf{U}$  también representa un cambio de base, de las unidades originales a unidades nuevas. Así, los elementos cuadrados  $u_{i\alpha}$  suman 1 y pueden considerarse como *contribuciones*, esta vez de las unidades a la formación de los factores. Aquí también pueden ocurrir casos extremos, desde contribuciones equilibradas en las que todas las unidades contribuyen en alrededor de  $1/n$  al factor hasta el caso en que una unidad contribuye más exclusivamente a dicho factor. En este último caso, cabe preguntarse si se desea mantener o no esta unidad junto con las demás, dada su evidente diferencia.

De manera análoga a las unidades, para obtener las coordenadas de las variables, las obtenemos multiplicando  $\mathbf{X}'\mathbf{U}$ . Tenga en cuenta que, dado que ambas columnas de matrices están centradas, las coordenadas son un múltiplo de la covarianza entre variables y factores.

De esta manera, la inercia de una característica o variable se descompone en componentes ortogonales, siendo cada uno su covarianza con cada eje. Así, el cuadrado de las covarianzas representan una calidad de representación de una característica en un factor y pueden acumularse en un plano, un espacio, etc.

*PCA* realizado de esta manera sólo tiene sentido si las variables involucradas pueden sumarse, es decir, si representan la misma entidad y se miden de la misma manera, porque es una cantidad común (la inercia expresada por la variación de las medidas de las variables) que se dividirán en componentes principales. Considerar que la misma cantidad expresada con diferentes unidades de medida puede dar resultados bastante diferentes. Como se mencionó anteriormente, utilizar las variables en sus unidades originales podría sesgar los resultados del *PCA*. Por esta razón se sugiere utilizar el *PCA* a las variables estandarizadas. En ese sentido, hay dos formas de analizar el *PCA*, uno sobre las variables originales sin transformar el cual utiliza la matriz de covarianzas, y otra forma de analizar el *PCA* es sobre los datos estandarizados el cual se basa en la matriz de correlación de las variables.

*PCA* también puede verse como la solución de un problema de optimización. La más conocida es la búsqueda de direcciones de las cuales se maximiza la inercia de la nube de puntos que representan las unidades en el espacio  $\mathbb{R}^p$ . Suponemos que  $\mathbb{R}^p$  generado por las variables originales en los que la nube de  $n$  unidades-puntos está en la posición dada por sus valores tomados para cada variable. Se puede demostrar que la inercia de la nube es igual a la suma de la inercia a lo largo de cada variable, que es la traza de la matriz de covarianza o correlación entre las variables. En este marco, *PCA* corresponde a la búsqueda iterativa de las direcciones (ortogonales a las encontradas anteriormente) a lo largo de las cuales la inercia de la nube es máxima. La solución de este problema es equivalente al siguiente teorema, que establece la importancia decreciente de los componentes principales, una vez clasificados en orden decreciente del valor propio correspondiente.

**Teorema 4.1.3** (Eckart y Young). *Sea  $\mathbf{X}$  una matriz real  $n \times p$  con rango  $K$  y sea  $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$  su descomposición en valores singulares  $\sigma_k$ , ordenados en orden decreciente. Entonces el rango  $k < K$  de la matriz  $\mathbf{B}$  tal que  $\|\mathbf{X} - \mathbf{B}\|^2 = \min_{\mathbf{C}}(\|\mathbf{X} - \mathbf{C}\|^2)$  es dado por:*

$$\mathbf{B} = \mathbf{U}_{n,k}\mathbf{\Sigma}_k\mathbf{V}'_{k,p} \quad (4.1)$$

donde  $\mathbf{U}_{n,k}$  está formado por las primeras  $k$  columnas de  $\mathbf{U}$ ,  $\mathbf{V}'_{k,p}$  está formado por las primeras  $k$  filas de  $\mathbf{V}'$ , y  $\mathbf{\Sigma}_k$  es una matriz diagonal formada por los primeros  $k$  elementos diagonales de  $\mathbf{\Sigma}$ .

En consecuencia, se justifica el hábito habitual de estudiar las componentes principales uno a uno, partiendo del valor propio mayor. De hecho, (4.1) establece que la matriz  $\mathbf{X}$  puede reconstruirse usando como máximo  $\min(p, n) = r$  matrices de rango uno que llamamos *capas*, que son:

$$x_{ij} = \sum_{\alpha=1}^r \sigma_{\alpha} u_{i\alpha} v_{j\alpha}.$$

Dado que tanto  $\mathbf{u}$  como  $\mathbf{v}$  tienen la norma 1, la inercia llevada por cada capa  $\alpha$  en la reconstrucción,  $\mathbf{X}_{\alpha} = \sigma_{\alpha} \mathbf{u}_{\alpha} \mathbf{v}'_{\alpha}$ , depende esencialmente de  $\sigma_{\alpha}$ . Por lo tanto, disminuye de la

primera a la última capa.

### 4.1.3. Reglas de detención

Surge una pregunta, si nos detenemos o no después de elegir algunos componentes. Esta pregunta puede plantearse si *PCA* se introduce como un método para reducir de manera óptima la dimensión de una tabla de datos, minimizando la inercia no considerada. De hecho, esta es la presentación más común - la búsqueda de ejes factoriales que maximicen la inercia de las unidades de nube proyectadas sobre ellos - pero dirige la atención del usuario a este aspecto en particular, que en el análisis exploratorio puede no ser el enfoque adecuado. De hecho, dónde detenerse depende en gran medida de la naturaleza del estudio bajo análisis: en un marco exploratorio no tiene sentido introducir una regla de detención fijando cualquier tipo de nivel, ya que el propósito de *PCA* es explorar los datos hasta el final de la información disponible. Por otro lado, uno puede desear saber si los resultados obtenidos son consistentes o son ruido aleatorio. Esta es la razón por la que decidimos aportar a nuestro análisis un par de estadísticas que puedan dar una idea de la aleatoriedad de los resultados obtenidos.

El primer método es el estadístico de *Brokenstick* (Barton y David, 1956). Se basa en la distribución esperada del tamaño (en orden decreciente) de  $p$  subintervalos obtenidos mediante una elección aleatoria de  $p - 1$  puntos de corte del intervalo real  $[0, 1]$  (Barton y David, 1956). Frontier (1976) argumenta que la inercia total de una tabla de datos, si los valores propios fueran aleatorios, se distribuiría de acuerdo con ellos de la misma manera, es decir, se *desgloza* en inercia a lo largo de los componentes principales como una barra cortada en  $p$  pedazos aleatorios. Por lo tanto, se espera que la longitud absoluta de cada pieza en relación al total (se supone que es igual a  $p$ , el número de valores propios), ordenados en orden decreciente, sea:

$$l_k = \frac{1}{p} \sum_{i=k}^p \frac{1}{i}$$

Por lo tanto, en orden descendente, se puede tener en cuenta cualquier valor propio  $\lambda_k > l_k$ , que sea mayor que la correspondiente pieza *desglozada* esperada. Notar que esta fórmula solo depende del número de variables.

El segundo método es el estadístico de *Rencher-Bartlett* (Rencher, 2002). En su forma original, el test de Bartlett para distribución esférica estima si la tabla de datos en cuestión merece ser analizada, comprobando la relevancia solamente del primer valor propio. Esta prueba es una generalización de la prueba de Bartlett, cuyo objetivo es comprobar si cada valor propio secuencial es significativamente diferente de los restantes. La prueba se basa en las estadísticas del  $k$ -ésimo valor propio,  $k = 1, \dots, p$ :

$$\chi_k^2 = \left( n - \frac{2p + 11}{6} \right) \left( k \log_e \bar{\lambda} - \sum_{i=k}^p \log_e \lambda_i \right), \text{ con } \bar{\lambda} = \frac{\sum_{i=k}^p \lambda_i}{(p - k + 1)},$$

que se distribuyen como chi-cuadrado con  $\frac{(p-k)(p-k+3)}{2}$  grados de libertad. Por lo tanto, las dimensiones se consideran significativas si la prueba de chi-cuadrado resulta significativa en un nivel  $\alpha$  predefinido. Nótese que, a diferencia del test de *Brokenstick* que tiende a retener

pocos componentes, el método de Rencher es criticado, ya que tiende a retener más componentes principales que las realmente necesarias (Jolliffe, 2002).

De hecho, ninguno de los métodos parece satisfactorio: uno es demasiado conservador y el otro demasiado relajado (Camiz y Pillar, 2018). Por lo tanto, preferimos ser flexibles y adoptar un enfoque de *buen sentido*, basado en algunas observaciones:

1. Dado que la inercia de cada variable es igual a 1, y los valores propios suman  $p$ , el valor propio en sí vale la inercia a lo largo del factor correspondiente medido en términos de variables resumidas.
2. Por tanto, los factores a tener en cuenta deberían explicar una proporción razonable de inercia: en la literatura se sugiere limitar la atención a aquellos componentes cuyos valores propios son al menos mayores al promedio ( $= 1$ ), aunque Jolliffe (2002) sugiere 0.7.
3. Si se pueden interpretar los factores, se revelan nuevos conocimientos sobre el contenido de los datos y merecen ser tomados en cuenta, independientemente de su inercia.
4. Algunos factores pueden estar muy influenciados por un pequeño número de unidades: esto debe tenerse en cuenta para explicar el comportamiento de estas unidades.
5. Algunos factores pueden representar *ruido* no estructurado, error o variación individual sin ningún interés para el estudio: esto puede determinarse verificando la distribución aleatoria de las unidades y la pérdida de correlaciones con las variables.
6. La naturaleza exploratoria del estudio sugiere utilizar todas estas observaciones como una forma de explicar la tabla de datos y no como una herramienta para reducir artificialmente la dimensión de la tabla.

#### 4.2. Clasificación Factorial Jerárquica (HFC)

Una atención especial merece el PCA por pares de variables: en este caso es fácil demostrar que el primer componente principal sintetiza lo que las dos variables tienen en común y el segundo mide sus diferencias. Estas características llevaron a Denimal (2007) a desarrollar su *Clasificación Factorial Jerárquica* (HFC), propuesta por Camiz (2021) para clasificar series de tiempo de precios de acciones brasileñas. El método tiene como objetivo combinar la clasificación de variables con los métodos factoriales similar al análisis exploratorio de PCA en un solo procedimiento, de modo que dé como resultado al mismo tiempo una jerarquía y un conjunto de planos principales asociados a los nodos de la jerarquía. La asociación entre las variables se basa en su covarianza recíproca y, para cada nodo, el método proporciona dos planos principales donde se pueden representar variables y unidades, respectivamente. Dado que el método se basa en el mismo espacio geométrico que el PCA, los dos componentes principales resultantes de cada nodo se pueden representar como elementos ilustrativos en los planos principales del PCA. Esto permite una *interoperabilidad* entre los dos métodos. Se puede aplicar HFC a variables estandarizadas cuantitativas. El método crea una jerarquía cuyos nodos son grupos de variables que se agregan progresivamente y un índice de jerarquía que mide cuánto las variables del grupo formado son diferentes. A cada grupo se pretende asociar una variable representativa que resuma todas las características del grupo. Luego,

cortando la jerarquía a un nivel dado del índice se resulta una partición compuesta para varias ramas del dendrograma que se resultan. El método considerado funciona de la siguiente manera:

1. Al principio, cada variable se considera como un grupo compuesto por ella misma y por tanto se le representa. Luego el algoritmo recursivo se basa en los siguientes pasos. Tener en cuenta que el algoritmo agrupa las variables y cada grupo tendrá una variable representativa.
2. Se comparan todos los pares de grupos existentes, a través de su variable representativo: cada par de variables representativas se somete a un PCA no-normalizado, es decir, el PCA de su matriz de covarianza  $2 \times 2$ . Tener en cuenta que los dos valores propios suman la inercia de las dos variables representativas: es fácil entender que el primero lleva la cantidad que comparten y el segundo la cantidad de sus diferencias, respectivamente.
3. Entre todos estos pares, se selecciona el par cuyo segundo valor propio de PCA es mínimo. Esto significa que se busca el par de variables de cuya diferencia es mínima.
4. Los dos grupos de variables correspondientes al par seleccionado se fusionan y el grupo resultante forma un nuevo nodo de la jerarquía.
5. Dado el nuevo nodo, el primer componente principal del PCA de este nodo se elige como su variable representativa.
6. El primer valor propio es igual a la inercia del primer componente principal, que es una parte de la inercia total de las variables en el nodo. Notar que esta inercia puede ser muy grande, ya que el PCA se basa en la covarianza, dependiendo tanto del número como de la inercia de las variables reunidas en el grupo.
7. El segundo componente principal representa la cantidad de diferencias entre las variables que componen el nodo, las cuales se esperan que sean las menores posibles.
8. Se elige el segundo valor propio como índice de jerarquía del nodo.
9. A cada nodo de la jerarquía se le asocian dos planos principales: en el primero se representan las covarianzas / correlaciones entre las variables pertenecientes al nodo y los dos componentes principales como coseno de ángulos, mientras que en el segundo se representa la dispersión de las unidades - visto sólo por estas variables. Por el significado de los componentes principales, se puede realizar una partición fácilmente asociada (aunque no óptima) de las unidades en cuatro grupos, de acuerdo con la concordancia-discordancia de sus valores sobre las variables representativas de los grupos agregados. De esta manera, las relaciones entre variables representativas de diferentes nodos pueden usarse para comprender similitudes y diferencias entre los correspondientes grupos de variables.
10. Si las variables son  $p$ , los pasos 2...9 se repiten  $p - 1$  veces, obteniendo una clasificación factorial jerárquica completa de las variables.
11. Para obtener una partición adecuada, se puede cortar el dendrograma en un nivel dado del índice. De hecho, es importante evitar la creación de grupos que contengan variables independientes. Para esta tarea, asumimos que el índice de jerarquía no puede superar el 0.7, lo contrario de lo que recomienda Jolliffe(2002) para que un factor sea considerado relevante.

Cabe señalar que el primer componente principal emitido en cada paso es un promedio ponderado de las variables que forman el nodo: si sus correlaciones con estas variables tienen todas el mismo signo, puede ser considerado como positivo y tiene una posición de baricentro ponderada dentro ellos, justificando así su uso como representante de ellos. En caso de que sean de signo opuesto, la variable representativa se refiere a un *dipolo* de variables, que identifican la misma variable representativa con significado opuesto. En el caso de las series de tiempo, la variable representativa es un promedio ponderado de los que forman el grupo y, como tal, su patrón a lo largo del tiempo puede verse como un proxy del patrón de los demás variables; por otro lado, las series en diferencias pueden describir su variación a lo largo del tiempo.

El resultado es que los dos componentes principales del último nodo construido de HFC suelen estar muy correlacionados con los dos primeros componentes principales del PCA realizado en todo el conjunto de datos. Por tanto, su interpretación puede ser bastante similar. De hecho, ambos métodos se basan en el mismo espacio geométrico de PCA; por lo tanto, las variables representativas de todos los nodos pueden representarse como elementos suplementarios en los planos principales del PCA ordinario, lo que permite una interoperabilidad entre los dos métodos. Notar que las variables representativas de los diferentes nodos pueden no ser ortogonales; así su interpretación puede referirse al grupo correspondiente de variables.

### 4.3. Análisis Evolutivo (EPCA)

La idea del PCA evolutivo (EPCA) es bastante simple: significa definir una ventana móvil de un número determinado de observaciones a lo largo del tiempo y realizar un PCA en cada ventana. Esto proporciona una enorme cantidad de información que puede resultar difícil de estudiar. Por tanto, se consideran de interés dos salidas gráficas para un primer análisis: la serie temporal de los primeros valores propios emitidos por el PCA y la serie temporal de las correlaciones entre las variables y las primeras componentes principales. De esta manera se pueden detectar gráficamente los años en los que ocurren algunos cambios muy importantes: primero, observando variaciones importantes en los tamaños de los valores propios ya que el tamaño del primer valor propio indica mayor o menor homogeneidad entre las variables en cada periodo; segundo, al observar cambios importantes en las correlaciones, es decir, en el significado de los componentes principales. Se sabe que para realizar un PCA se requiere tener más observaciones que variables. Por lo tanto, para definir la ventana móvil debe ser un intervalo de tiempo más grande que el número de variables. Otro punto que merece ser tomado en cuenta es el hecho de que dos PCA realizadas en ventanas posteriores diferirían solo en una unidad: la primera observación de la primera ventana que se elimina y la última observación de la segunda ventana que ingresa. De hecho, la variación entre dos ventanas subsiguientes dependería de solo dos veces las unidades  $n - 1$  distantes entre sí, y eso contribuye solo a  $1/n$  para el cálculo de correlación.

Aún baja, esta influencia puede ser demasiado, causando variaciones locales fuertes. Por esto se ha pensado a una ponderación de las unidades, la misma en cada ventana, dando mayor pesos a las observaciones centrales y menor a las extremas. De esta manera, el cálculo de la correlación en cada ventana se basa más sobre los valores centrales, cuya variación local tiene más sentido, y menos sobre los extremos, que se van intercambiando y que podrían ser

mucho más diferentes. La experiencia muestra que de esta manera las variaciones resultan más suavizadas. Esta suavidad se puede controlar utilizando un factor  $k$  multiplicativo. De hecho, consideramos la secuencia  $i = 1, \dots, m$  y calculamos valores  $v_i$  utilizando una función gaussiana. Como el promedio es  $am = \frac{m+1}{2}$  y el desvío estándar es  $s = \frac{m-1}{\sqrt{2}}$  los valores  $v_i = e^{-\frac{1}{2}\left(\frac{(i-am)*k}{s}\right)^2}$ ,  $i = 1, \dots, m$  y son la densidad de la función gaussiana en dichos puntos. Estos son todos positivos y iguales por pareja con misma distancia al promedio. Es evidente que el valor  $k$  creciendo reduce el peso de los extremos, así contribuyendo a determinar la suavización. Al final  $w_i = \frac{v_i}{\sum_i v_i}$  son los pesos que se estarán utilizando. En este trabajo, después de una experimentación con valores variados, se eligió el parámetro  $k = 3$  ya que resultó en la mejor suavización, y dichos pesos se mantuvieron iguales para todas las ventanas  $nm + 1 = 240$ , de modo que, mientras se avanza, la fecha de entrada tendría un peso muy bajo, luego sube hasta la mitad y luego disminuye hasta salir de la ventana. De esta forma, la contribución a la inercia fue mayor para las fechas centrales.

Es importante mencionar, además, que durante el proceso iterativo del cálculo de las componentes principales se puede observar que el signo de algunas componente varía de una ventana a otra complicando la representación de las correlaciones. Por tal motivo, el código que usaremos intenta corregir este problema eligiendo en cada paso el signo positivo de la mayoría de las correlaciones, pero con evidentes problemas en las salidas.

Finalmente, con base en los resultados emitidos por el análisis evolutivo se puede dividir el periodo total en intervalos en los que la estructura de los datos sea suficientemente homogénea. Esto se puede lograr aplicando el algoritmo *Calinski-Harabasz* a la serie de tiempo de los primeros valores propios emitidos por EPCA. El algoritmo identifica como puntos de corte en la serie los intervalos entre dos fechas posteriores minimizando la inercia dentro de los grupos formados. Por lo tanto, si se construyen varias particiones con un número creciente de grupos, las mejores corresponden a los máximos locales del ratio  $\frac{BSS/(k-1)}{WSS/(n-k)}$ , donde  $BSS$  es la suma de cuadrado entre grupos,  $WSS$  es la suma de cuadrado dentro de cada grupo y  $k$  es el número de grupos que se van a formar. Variando progresivamente  $k$  se eligen como grupos los  $k$  correspondientes a máximos locales, si hay. Lamentablemente no siempre se encuentran máximos, así que el método no siempre se puede aplicar. En este caso, se busca las mayores diferencias entre inercias de grupos en número.

## Capítulo 5

### Resultados

#### 5.1. Análisis de componentes principales

Como parte del análisis exploratorio se realizó un análisis de componentes principales para las 17 variables durante el periodo enero 1999 a noviembre 2020. Como se observa en la siguiente tabla el primer valor propio tiene un porcentaje de inercia de 72%, el segundo de 12.9% y el tercero de 7.1%. Estos tres primeros valores propios acumulan un 92% de la variabilidad de los datos. De acuerdo al test de Brokenstick, un componente se retiene si su valor propio asociado es mayor que el valor dado por la distribución del test Brokenstick. Según los resultados de la tabla 5.1 sólo se elegiría el primer valor propio. En el caso del test de Rencher-Bartlett, se elegirían 16 componentes principales con un nivel de significancia de 5%.

Teniendo en cuenta estos resultados obtenidos y la sugerencia de Jolliffe (2002) de guardar sólo las componentes con valores propios mayor a 0.7, analizaremos los tres primeros componentes pues concentran una elevada participación de la varianza de los datos (92%).

Cuadro 5.1: Resultados del PCA al conjunto de datos

Nº CP	Eigenvalue	%Inercia	%Acumulado	Brokenstick	Rencher	g.l	p-value
1	12.25	72.06	72.06	3.44	13,897.75	152	0.00
2	2.19	12.88	84.94	2.44	9,572.57	135	0.00
3	1.21	7.12	92.05	1.94	7,962.71	119	0.00
4	0.64	3.79	95.84	1.61	6,422.67	104	0.00
5	0.31	1.85	97.69	1.36	5,001.91	90	0.00
6	0.16	0.95	98.65	1.16	3,892.76	77	0.00
7	0.08	0.45	99.10	0.99	3,049.07	65	0.00
8	0.05	0.27	99.36	0.85	2,590.21	54	0.00
9	0.04	0.21	99.57	0.72	2,309.57	44	0.00
10	0.03	0.18	99.75	0.61	2,016.22	35	0.00
11	0.02	0.13	99.87	0.51	1,617.56	27	0.00
12	0.01	0.06	99.93	0.42	1,103.61	20	0.00
13	0.01	0.04	99.97	0.34	787.76	14	0.00
14	0.00	0.02	99.99	0.26	384.41	9	0.00
15	0.00	0.01	100.00	0.19	170.58	5	0.00
16	0.00	0.00	100.00	0.12	35.41	2	0.00
17	0.00	0.00	100.00	0.06	0.00	0	1.00

**Nota:** g.l: Grados de libertad.

### 5.1.1. Análisis del primer componente

En el gráfico 5.1 se muestra las correlaciones de las variables con el primer y segundo componente. Se observa que la variable **tc** tiene una correlación negativa con el primer componente, pero tiene una correlación menor al resto de variables las cuales se correlacionan negativamente con dicho componente. En el caso de la variable **plca** tiene una correlación positiva con el primer componente principal, pero es menor al resto de variables que a su vez se correlacionan positivamente con este componente.

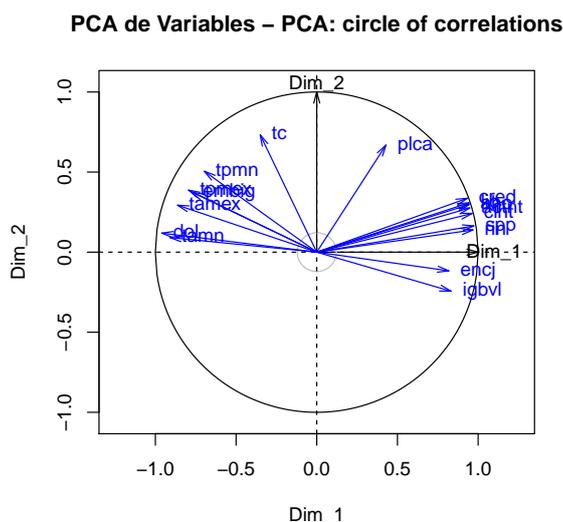


Gráfico 5.1: Correlación entre el primer y segundo componente

Como se muestra en el cuadro 5.2, la mayoría de las variables tienen una contribución similar o superior al promedio ( $5,9\% = 100/17$ ) excepto las variables **tc**, **plca**, **tpmn** y **embig**. Así mismo, se observa que la correlación de cada variable  $x_i$  es alta (mayor a 0.7 en valor absoluto) excepto para las variables anteriormente mencionadas. Finalmente el indicador de calidad (*quality*) nos indica que todas las variables están correctamente representadas por el primer componente exceptuando **tc** y **plca**.

Las series de tiempo elegidas para el análisis presentan una tendencia positiva o negativa. El primer componente obtenido sintetiza dichas tendencias opuestas. Como se observa en el gráfico 5.2, dicha tendencia común es representada como una tendencia creciente a lo largo del tiempo. Esto se explica porque la suma de las contribuciones de las variables con tendencia positiva (*aho*, *spp*, *rin*, *clnt*, *cred*, *actnt*, *liq*) es de 52%, mientras que la suma de las contribuciones de las variables con tendencia negativa (*tamn*, *tamex*, *tpmn*, *tpmex*) es de 22%, es decir las variables con tendencia positiva tienen la mayor explicación en el primer componente.

En el gráfico 5.2 se evidencia un quiebre alrededor del 2003 producto del quiebre en las tasa activas en moneda nacional y extranjera. Alrededor del año 2008 también se observa otro quiebre debido a que se produjo la crisis financiera. Finalmente, en el 2015 se observa que la serie se mantuvo constante debido a que las tasas activas en moneda nacional y extranjera también se mantuvieron constantes.

Cuadro 5.2: Resultados del primer componente

id	Eigenvector	%Contribución	%Correlación	Corr. Múltiple	Calidad	Calidad Ac.
actnt	0.27	7.34	0.95	0.95	0.90	0.90
clnt	0.28	7.57	0.96	0.96	0.93	0.93
aho	0.27	7.27	0.94	0.94	0.89	0.89
encj	0.23	5.47	0.82	0.82	0.67	0.67
cred	0.27	7.11	0.93	0.93	0.87	0.87
igbvl	0.24	5.66	0.83	0.83	0.69	0.69
spp	0.28	7.73	0.97	0.97	0.95	0.95
tc	-0.10	1.00	-0.35	0.35	0.12	0.12
liq	0.27	7.27	0.94	0.94	0.89	0.89
dol	-0.27	7.54	-0.96	0.96	0.92	0.92
rin	0.28	7.65	0.97	0.97	0.94	0.94
plca	0.12	1.50	0.43	0.43	0.18	0.18
tamn	-0.26	6.75	-0.91	0.91	0.83	0.83
tpmn	-0.20	3.97	-0.70	0.70	0.49	0.49
tamex	-0.25	6.08	-0.86	0.86	0.75	0.75
tpmex	-0.23	5.16	-0.80	0.80	0.63	0.63
embig	-0.22	4.93	-0.78	0.78	0.60	0.60

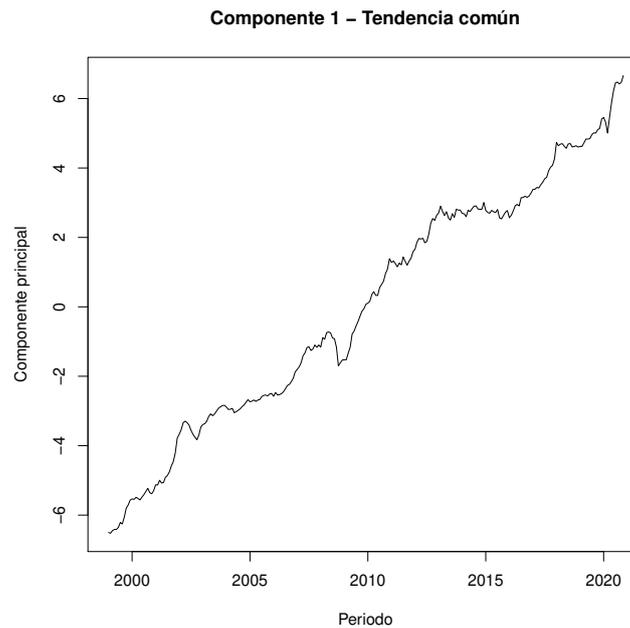


Gráfico 5.2: Componente Principal N° 1: Tendencia Común

### 5.1.2. Análisis del segundo componente

En el gráfico de correlaciones 5.1 se observa que las variables **tc**, **plca** y **tpnm** tienen una alta correlación positiva con el segundo componente principal.

Como se muestra en el cuadro 5.3, sólo las variables *tc*, *plca* y *tpnm* tienen una contribución significativa para explicar el segundo componente alcanzando una contribución total de 56.6%. Así mismo, el indicador de calidad (*quality*) nos indica que todas las otras variables están débilmente representadas pues tienen un bajo indicador de calidad excepto las tres variables mencionadas anteriormente.

Cuadro 5.3: Resultados del segundo componente

id	Eigenvector	%Contribución	%Correlación	Corr. Múltiple	Calidad	Calidad Ac.
actnt	0.19	3.47	0.28	0.99	0.08	0.98
clnt	0.16	2.67	0.24	0.99	0.06	0.99
aho	0.20	4.02	0.30	0.99	0.09	0.98
encj	-0.08	0.62	-0.12	0.83	0.01	0.68
cred	0.23	5.17	0.34	0.99	0.11	0.98
igbvl	-0.17	2.73	-0.24	0.87	0.06	0.75
spp	0.11	1.22	0.16	0.99	0.03	0.97
tc	0.50	24.50	0.73	0.81	0.54	0.66
liq	0.21	4.32	0.31	0.99	0.10	0.99
dol	0.08	0.65	0.12	0.97	0.01	0.94
rin	0.09	0.88	0.14	0.98	0.02	0.96
plca	0.45	20.44	0.67	0.80	0.45	0.63
tamn	0.06	0.38	0.09	0.91	0.01	0.84
tpmn	0.34	11.67	0.51	0.86	0.26	0.74
tamex	0.20	3.93	0.29	0.91	0.09	0.83
tpmex	0.26	6.86	0.39	0.88	0.15	0.78
embig	0.25	6.47	0.38	0.86	0.14	0.75

El segundo componente está representado principalmente por las variables tipo de cambio, el ratio de palanca global de los bancos y la tasa pasiva en moneda nacional.

Como se mencionó anteriormente, el incremento del ratio de capital global en el 2018 se debió a un cambio regulatorio en el cual se solicitó a los bancos un incremento del ratio para tener mayor cobertura de liquidez en caso de crisis. Esta variable no podría ser retirada porque está capturando la fortaleza financiera de una entidad bancaria ante una crisis de liquidez. Por otro lado, el tipo de cambio es una variable financiera que está asociado a fluctuaciones macroeconómicas principalmente externas y tiende a subir ante factores negativos como por ejemplo una caída de las principales economías del mundo hace que los inversionistas refugien su dinero en dólares y esto hace aumentar el tipo de cambio. Finalmente, la tasa de interés pasiva en moneda nacional es un factor que influye en las decisiones de las personas y/o empresas para ahorrar. En el gráfico 5.3 se muestra el segundo componente principal. Según lo expuesto en el párrafo anterior este componente puede asociarse a la respuesta al riesgo que tienen las empresas o personas ante factores externos o internos. La duración que tiene este componente es de 20 años el cual es mayor a la duración del ciclo económico (5 años en promedio). Por tal motivo, podríamos considerar este segundo componente principal como un componente del ciclo financiero de mediano plazo.

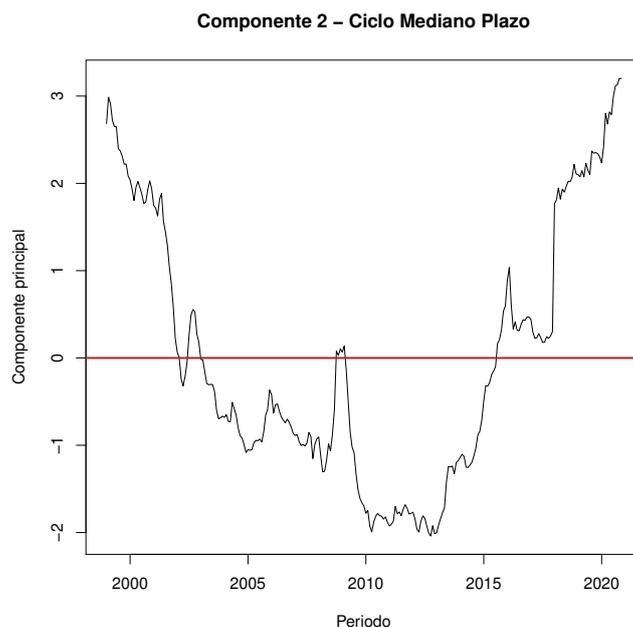


Gráfico 5.3: Componente Principal N° 2: Ciclo Financiero - Mediano Plazo

Durante los años 2001-2002 se observa un salto de la serie por encima de cero debido a una mayor volatilidad en el indicador del Embig Perú (*embig*). En esos años el Perú salía de una crisis política luego del gobierno de Alberto Fujimori. Por otro lado, durante los años 2008-2009 también hay un pico en la serie producto de la crisis financiera global que afectó el riesgo país. Alrededor del año 2016 se evidenció un salto en la serie ocasionado por el cambio en los niveles de ahorros en el Sistema Financiero (*aho*) y en el Embig Perú (*embig*). Finalmente, en el año 2018 se presenta un salto atípico producto del *cambio regulatorio* en el indicador de la palanca global que fue incrementado por la SBS.

Para una mejor interpretación del segundo componente, se añadió una línea horizontal en 0 (umbral) para identificar periodos de crecimiento o caída en la serie. Del gráfico se desprende que durante el periodo 2003-2015 este componente se encontraba por debajo del umbral indicando que las series también habían tenido un comportamiento decreciente en este periodo. Un tipo de cambio bajo significa que no hay escenarios de macroeconómicos negativos, mientras que una tasa pasiva decreciente implica que hay muchos ahorristas en el sistema financiero por lo cual las entidades bancarias pagarían menos por los depósitos. Finalmente, un indicador de palanca global bajo indica que el regulador no esperaba una crisis de liquidez por lo que permitió a los bancos mantener menos capital en caso de este escenario. En conclusión podríamos decir que este periodo donde la serie se encontraba por debajo del umbral cero fue un periodo de poca inestabilidad financiera en el Perú.

### 5.1.3. Análisis del tercer componente

En el gráfico 5.4 se observa que la variable **tc** tiene una alta correlación negativa con el tercer componente principal y es la variable que tiene mayor peso. La variable **plca** también tiene una correlación negativa con el tercer componente pero tiene una contribución de 0.06 %. El resto de variables tienen correlación positiva con el tercer componente. Finalmente, las variables que se encuentran en el cuadrante derecho indican una baja correlación con el tercer

componente.

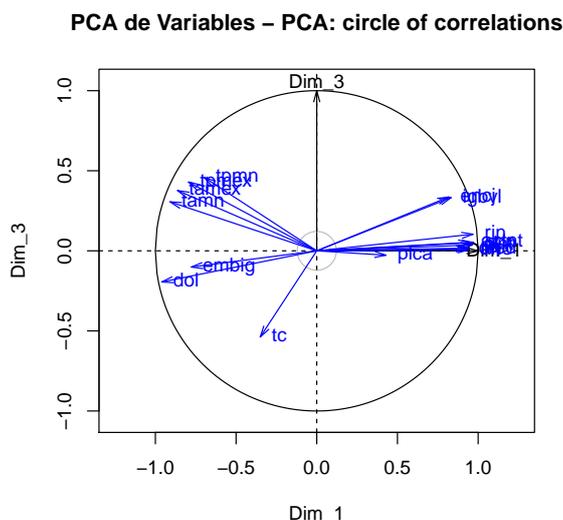


Gráfico 5.4: Correlación entre el primer y tercer componente

Como se muestra en el cuadro 5.4, el tercer componente está representado principalmente por las variables tipo de cambio (**tc**), la tasa pasiva en moneda nacional (**tpmn**) y moneda extranjera (**tpmex**); y la tasa activa en moneda extranjera (**tamex**) las cuales alcanzan una contribución total de 68.3%. Se observa que la calidad acumulada de estas variables es mayor al 95% y también la correlación múltiple es mayor al 97%.

Cuadro 5.4: Resultados del tercer componente

id	Eigenvector	%Contribución	%Correlación	Corr. Múltiple	Calidad	Calidad Ac.
actnt	0.05	0.23	0.05	0.99	0.00	0.98
clnt	0.03	0.11	0.04	0.99	0.00	0.99
aho	0.01	0.00	0.01	0.99	0.00	0.98
encj	0.31	9.35	0.34	0.89	0.11	0.80
cred	0.01	0.02	0.01	0.99	0.00	0.98
igbvl	0.30	9.12	0.33	0.93	0.11	0.86
spp	0.05	0.21	0.05	0.99	0.00	0.98
tc	-0.49	23.94	-0.54	0.97	0.29	0.95
liq	0.02	0.02	0.02	0.99	0.00	0.99
dol	-0.18	3.13	-0.19	0.99	0.04	0.98
rin	0.09	0.89	0.10	0.98	0.01	0.97
plca	-0.03	0.06	-0.03	0.80	0.00	0.63
tamn	0.28	7.74	0.31	0.96	0.09	0.93
tpmn	0.42	17.33	0.46	0.98	0.21	0.95
tamex	0.34	11.73	0.38	0.99	0.14	0.97
tpmex	0.39	15.26	0.43	0.98	0.19	0.97
embig	-0.09	0.85	-0.10	0.87	0.01	0.76

Como se muestra en el gráfico 5.5 el tercer componente principal puede indicar representar un ciclo financiero de corto plazo. Una característica de un ciclo es que debe fluctuar alrededor de un valor neutral de modo que puedan identificarse periodos de subida y caída en la serie.

Este ciclo de corto plazo está influenciado principalmente por variables de tasas de interés y el tipo de cambio. En el caso de las tasas de interés, estas fluctúan debido al movimiento de la tasa de interés de referencia que administra el Banco Central de Reservas del Perú (BCRP) el cual mensualmente modifica las tasas de acuerdo a la situación económica del País. En el caso del tipo de cambio, este componente estaría capturando las fluctuaciones de corto plazo de esta variable. Algunas de estas fluctuaciones son ocasionadas por la coyuntura política como por ejemplo elecciones presidenciales. Por estos motivos, se observa una mayor fluctuación en el tercer componente producto de los cambios en el corto plazo de las variables de mayor ponderación.

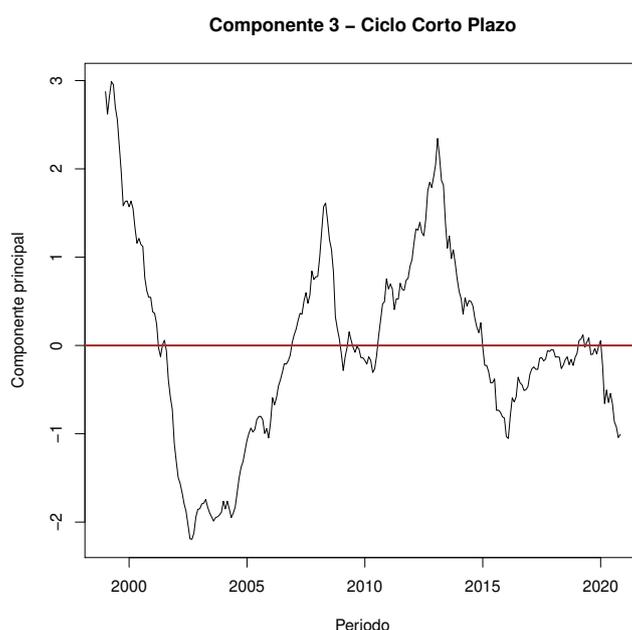


Gráfico 5.5: Componente Principal N° 3: Ciclo Financiero - Corto Plazo

Durante el periodo inicial de la muestra (1999) la serie presentó un nivel alto debido a que las tasas de interés (activas y pasivas) por esos años eran muy altas. Conformen pasaron los años las tasas de interés disminuyeron debido a la mayor competencia de los bancos en el mercado financiero. Así, la serie alcanza su mínimo valor en el 2004. Luego volvió a tener otro pico alrededor de los años 2008-2009 debido a la crisis financiera. Posteriormente, la serie tuvo un pico en el 2013 producto de una caída del tipo de cambio a su nivel más bajo en la muestra (2.5 PEN/USD). La serie volvió a tener una caída en el 2016 ocasionada por el incremento del tipo de cambio. Finalmente, es importante mencionar que el tercer componente principal estuvo por encima del umbral cero en periodos de 1999-2002, 2007-2008 y 2011-2015 los cuales coinciden con periodos de auge económico.

El signo o sentido que presenta el tercer componente en 5.5 es el esperado pues es un componente compuesto principalmente por las tasas de interés con una contribución acumulada de 52%. Podríamos indicar que este tercer componente es básicamente un ciclo de tasas de interés. Así, dicho ciclo de tasas de interés está estrechamente relacionado con los ciclos económicos: si una economía está *creciendo* con fuerza y aumentan las presiones inflacionarias, los bancos centrales *aumentarán* las tasas de interés para desacelerar la economía

y evitar la inflación. En cambio, si la economía estuviera en *recesión* con una inflación en descenso y un desempleo en aumento, los bancos centrales *disminuirían* las tasas de interés para estimular un aumento en la tasa de crecimiento económico.

Luego de observar los tres componentes obtenidos con la técnica de PCA, podemos concluir que el primer factor está relacionado con las tendencias de largo plazo de las series. El segundo factor empezó en un pico en el año 2000 y volvió alcanzar dicho pico durante el 2020, por lo cual nos indica que este ciclo ha tenido una duración de 20 años y, además, durante 10 años estuvo en periodos de contractivos (por debajo del umbral de 0). Finalmente, el tercer componente ha presentado varios ciclos con diferentes duraciones: el primero ciclo comprende los periodos Enero 1999 a Marzo 2008 con duración de 9.3 años, el segundo de Abril 2008 a Marzo 2013 con duración de 4.7 años y el último de Abril 2013 a Noviembre 2020 con duración de 6.8 años. En promedio, el tercer componente tiene una periodicidad de 6.9 años.

## 5.2. Clasificación Factorial Jerárquica

Como parte del análisis exploratorio, en esta sección se aplica la clasificación factorial jerárquica.

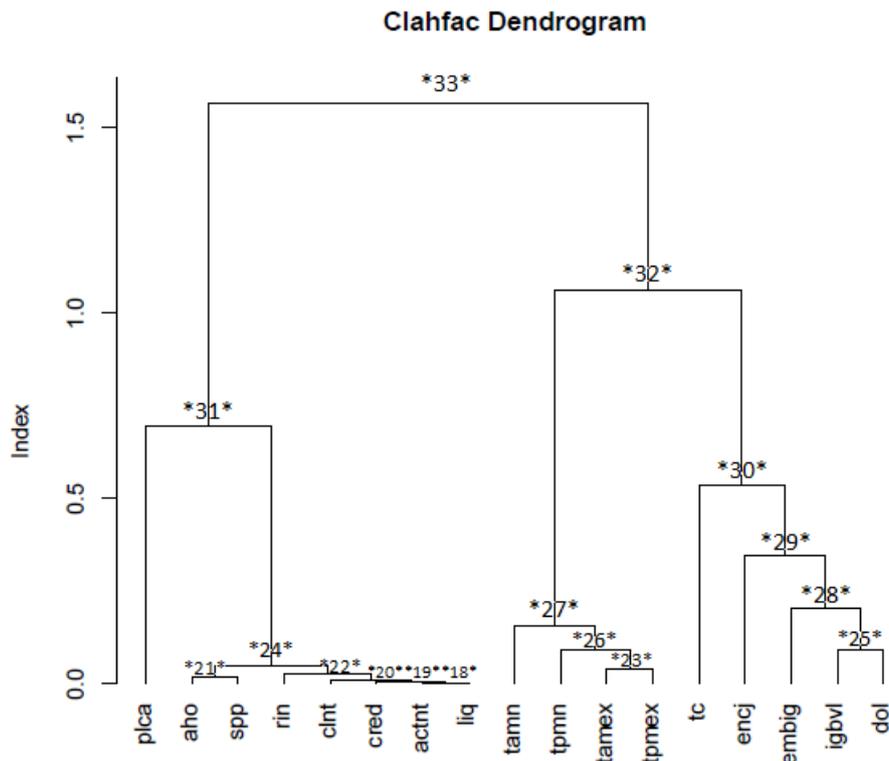


Gráfico 5.6: Dendrograma de clasificación jerárquica (eje vertical corresponde al índice de la jerarquía)

En la gráfica 5.6 se muestra el dendrograma que representa la jerarquía construida por el método HFC. Visualmente se observa que el algoritmo agrupa en diferentes categorías las variables. Para obtener una idea consistente sobre las particiones de los datos revisemos el cuadro 5.5, en el que se reportan los elementos básicos de la construcción del índice de jerarquía a través del método HFC. En este cuadro, empezando la lectura desde la izquierda,

se muestra los siguientes campos:

- *Nodo*: indica la etiqueta del nodo que se observa en el dendrograma.
- *ngr*: indica el número de grupo de la partición resultante.
- *N1* y *N2*: indica las variables y/o nodos que se han juntado en el paso anterior para formar el nuevo nodo. Por ejemplo, para formar el nodo \*25\* se han juntado las variables *igbvl* (6) y *dol* (10) <sup>1</sup>. En el caso del nodo \*30\* se han juntado la variable *tc* (8) y el nodo \*29\*. El nodo \*24\* se ha formado con la unión de los nodos \*21\* y \*22\*. Así, se pueden interpretar el resto de nodos.
- *NumVar*: indica la cantidad de variables que se encuentran en el nuevo nodo formado.
- *Vp1* y *Vp2*: indica los dos valores propios del nodo formado. El segundo es conocido como el índice de jerarquía.
- *% Inercia Local* (%In. Loc. 1 y 2): indica la repartición en cuotas de la inercia local (suma de las inercias de los nodos que se juntan) en las dos variables representativas. Por ejemplo, en el nodo \*24\* la inercia local del primer componente representa el 99.30 % que es igual al primer valor propio 6.89 dividido por la inercia del nodo \*24\* (igual a 6.93 = 6.89+0.05), mientras que la inercia local del segundo componente representa el 0.70 % (0.05/6.93).
- *% Inercia Global* (%In. Glb. 1 y 2) : indica la repartición en cuotas de la inercia total (suma de las inercias de cada variables) en las dos variables representativas. Por ejemplo, en el nodo \*24\* la inercia global del primer componente representa el 40.50 % que es igual al primer valor propio 6.89 dividido por la inercia total de los datos (igual a 17 debido a que son 17 variables analizadas), mientras que la inercia global del segundo componente representa el 0.29 % (0.05/17).
- *% Inercia Global Acumulada* (%In. Glb. Ac.2): indica la inercia acumulada de la inercia global del segundo componente (%In. Glb. 2). Es decir, esta midiendo el % acumulado que no se considera como similitud entre los datos. Nótese que la inercia total (100 %) se resulta como suma de la inercia global acumulada de los segundos componentes con la inercia global de la primera componente del último nodo. En este caso sería 71.15 % + 28.85 % = 100 %.

De acuerdo a la literatura, elegimos cortar el dendrograma antes del nodo \*31\* debido a que los nodos que se encuentran por encima de él tienen un índice de jerarquía mayor a 0.7. Por lo tanto, los grupos de la partición obtenida son los siguientes: \*27\* (compuesto por *tamn*, *tpmn*, *tamex* y *tpmex*), \*30\* (compuesto por *tc*, *encj*, *embig*, *igbvl* y *dol*) y \*31\* (compuesto por *plca*, *aho*, *spp*, *rin*, *clnt*, *cred*, *actnt* y *liq*).

Podemos resumir los grupos encontrados en la siguiente tabla

---

<sup>1</sup>Las etiquetas numérica que ha recibido cada variable son: 1 actnt, 2 clnt, 3 aho, 4 encj, 5 cred, 6 igbvl, 7 spp, 8 tc, 9 liq, 10 dol, 11 rin, 12 plca, 13 tamn, 14 tpmn, 15 tamex, 16 tpmex, 17 embig

Cuadro 5.5: dendrograma producido por el algoritmo de clasificación jerárquica

Nodo	ngr	N1	N2	Num Var	Vp1	Vp2 (Index)	% In. Loc. 1	% In. Loc. 2	% In. Glb. 1	% In. Glb. 2	% In. Glb. Ac 2
18	16	1	9	2	2.00	0.00	99.86	0.14	11.75	0.02	0.02
19	15	5	18	3	2.99	0.00	99.84	0.16	17.60	0.03	0.04
20	14	2	19	4	3.98	0.01	99.78	0.22	23.43	0.05	0.10
21	13	3	7	2	1.98	0.02	99.00	1.00	11.65	0.12	0.21
22	12	11	20	5	4.95	0.03	99.41	0.59	29.14	0.17	0.39
23	11	15	16	2	1.96	0.04	97.99	2.01	11.53	0.24	0.62
24	10	21	22	7	6.89	0.05	99.30	0.70	40.50	0.29	0.91
25	9	6	10	2	1.91	0.09	95.44	4.56	11.23	0.54	1.45
26	8	14	23	3	2.87	0.09	96.86	3.14	16.86	0.55	1.99
27	7	13	26	4	3.71	0.15	96.00	4.00	21.84	0.91	2.90
28	6	17	25	3	2.70	0.21	92.91	7.09	15.90	1.21	4.12
29	5	4	28	4	3.36	0.35	90.68	9.32	19.75	2.03	6.15
30	4	8	29	5	3.82	0.54	87.66	12.34	22.47	3.16	9.31
31	3	12	24	8	7.19	0.70	91.18	8.82	42.29	4.09	13.40
32	2	27	30	9	6.47	1.06	85.89	14.11	38.05	6.25	19.65
33	1	31	32	17	12.10	1.56	88.55	11.45	71.15	9.20	28.85

Grupo 1	Grupo 2	Grupo 3
aho	tamn	tc
spp	tpmn	encj
rin	tamex	embig
clnt	tpmex	igbvl
cred		dol
actnt		
liq		
plca		

- El grupo 1 está representado por variables con tendencia creciente junto con la serie que tuvo un quiebre estructural por cambios regulatorios (*plca*). La característica de la variable representativa se analizará en la siguiente sección.
- El grupo 2 está representado por variables de tasas de interés activas y pasivas. Estas series se caracterizan por presentar una tendencia decreciente. La característica de la variable representativa se analizará en la siguiente sección.
- El grupo 3 está representado por variables que presentan elevada volatilidad. La característica de la variable representativa se analizará en la siguiente sección.

### 5.2.1. Estudio de los nodos superiores

En esta sección se presentan los gráficos de correlaciones de los nodos superiores \*33\*, \*32\*, \*31\* y \*30\*:

#### Nodo \*33\*

Este nodo es el primer nivel del árbol de clasificación (Gráfico 5.6) que está compuesto por los

nodos \*31\* y \*32\*. En el gráfico 5.7 se muestra un círculo de correlaciones donde en el eje X se coloca la primera variable representativa del nodo \*33\*, que mide el resumen de las variables que conforman este nodo, y en el eje Y se coloca la segunda variable representativa del nodo \*33\*, que mide las diferencias de las variables que conforman este nodo. Los vectores de color azul muestran las correlaciones de las variables que conforman el nodo con cada uno de estos componentes. En este nivel son todas las 17 variables iniciales. También se muestra en color rojo las variables representativas de los nodos que se juntan: \*31A\* y \*31B\* son las dos variables representativas que miden la parte común y las diferencias, respectivamente, de las variables que forman el nodo \*31\*; y \*32A\* y \*32B\* son las dos variables representativas que miden la parte común y las diferencias, respectivamente, de las variables que forman el nodo \*32\*. Observando el gráfico 5.7 se ve que es muy parecido a la representación de las variables sobre el primer plano factorial de la PCA (gráfico 5.1). Por lo tanto, podemos concluir que la variable representativa del nodo \*33\* es equivalente al primer componente principal obtenido con las 17 series y por lo tanto representa la tendencia común de las series. En cambio, la segunda variable representativa del nodo \*33\* es equivalente al segundo componente principal obtenido con las 17 series y por lo tanto representa el ciclo de mediano plazo de las series.



de 1.06 lo cual indicaría que el nodo \*31\* es homogéneo y el nodo \*32\* no lo es. El detalle de ambos nodos se estudiarán más adelante.

En el gráfico de correlaciones 5.7 se observa que las variables *tc* y *plca* parecen formar grupos aislados. Por otro lado, se observa que las variables correlacionadas positivamente con la variable representativa del nodo \*33\* (\*33A\*) son *spp*, *rin*, *clnt*, *actnt*, *aho*, *liq*, *cred*, *igbvl*, *encj* y *plca*, mientras que el resto de variables presentan una correlación negativa. La magnitud de la correlación con la variable representativa (\*33A\*) se observa en la tabla 5.7:

Cuadro 5.7: Correlación de las variables del nodo \*33\* con sus variables representativas

Variable	*33A*	*33B*
spp	0.97	0.19
rin	0.96	0.19
clnt	0.96	0.27
actnt	0.94	0.31
aho	0.94	0.31
liq	0.94	0.34
cred	0.93	0.37
igbvl	0.84	-0.23
encj	0.82	-0.07
plca	0.43	0.56
tc	-0.37	0.66
tpmn	-0.71	0.56
embig	-0.78	0.36
tpmex	-0.80	0.40
tamex	-0.87	0.31
tamn	-0.91	0.12
dol	-0.96	0.09

También se ha calculado la correlación entre las variables representativas del nodo \*33\* (\*33A y \*33B\*) con los dos primeros componentes principales del nodo \*33\* en la tabla 5.8 y se ha encontrado que la variable representativa del nodo \*33\* (\*33A\*) tiene una correlación muy cercana a 1 con el primer componente principal del nodo \*33\*.

Cuadro 5.8: Correlación entre variables representativas y componentes principales - \*33\*

Correlación	Componentes Principales		Representativas	
	PCA 1 *33*	PCA 2 *33*	*33A*	*33B*
PCA 1 *33*	1.00			
PCA 2 *33*	0.00	1.00		
*33A*	1.00	-0.02	1.00	
*33B*	0.02	0.98	0.00	1.00

### Nodo \*32\*

Este nodo es el segundo nivel del árbol de clasificación (Gráfico 5.6) que está compuesto por los nodos \*27\* y \*30\*.

De acuerdo al índice de jerarquía, el nodo \*32\* no es homogéneo. Si revisamos el índice para los nodos que componen el nodo \*32\* se observa que el nodo \*27\* tiene un índice de jerarquía menor al del nodo \*30\*, por lo tanto el nodo \*27\* es más homogéneo que el nodo

\*30\*. Esto también se comprueba si observamos el gráfico de correlaciones 5.9 en el cual se observa que las variables de tasas activas y pasivas forman un grupo, y las cuatro variables restantes forman un dipolo.

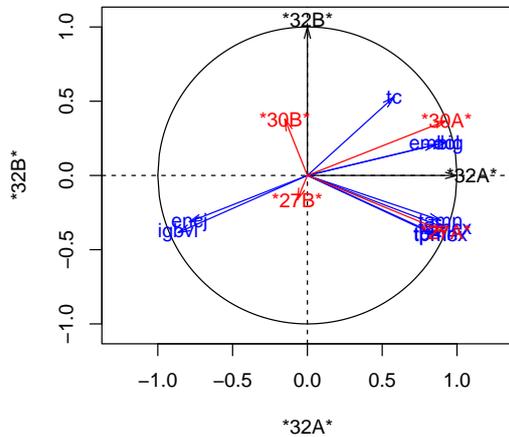


Gráfico 5.9: Correlaciones Nodo \*32\*

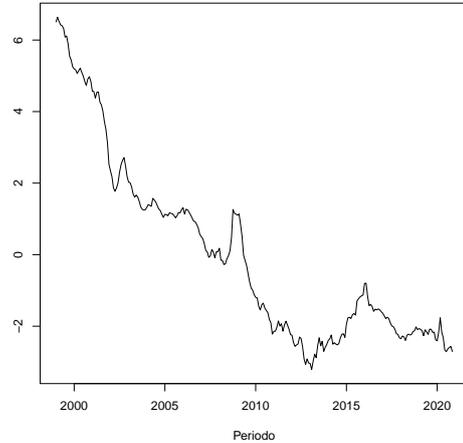


Gráfico 5.10: Variable representativa \*32\*

La variable representativa se muestra en 5.10 el cual muestra una tendencia negativa. Esta variable está midiendo la expansión del sector financiero, es decir a menor valor de la variable representativa significa que el mercado financiero está en periodo de auge: cuando las tasas activas disminuyen significa que existe mayor competencia entre los bancos, por lo tanto el mercado financiero se está expandiendo. Cuando el tipo de cambio también disminuye significa que los inversionistas observan que el mercado financiero peruano es sólido y pueden invertir dentro del país. De igual manera, un descenso del Embig Perú (*embig*) indica que la economía peruana es visto como un destino de inversión para capitales extranjeros. Un descenso del coeficiente de dolarización significaría que ante un riesgo cambiario el impacto sería leve indicando, además, que la mayor proporción de crédito otorgados en el país son en moneda local. Un incremento del índice de la bolsa de valores (*igbv1*) muestra las empresas que cotizan en el mercado de valores peruano están al alza debido a una mayor demanda de sus acciones por parte de otros inversionistas. Un incremento del encaje (*encj*) también sugiere que los bancos están otorgando mayores créditos producto de un crecimiento económico en el país.

En la tabla 5.9 se muestra los valores propios del nodo \*32\*. Como se puede observar, el nodo \*32\* no es homogéneo porque el primer componente principal sólo explica el 72.1% de la inercia de las variables que componen el nodo \*32\*; el segundo y tercer valor propio explican el 13.5% y 6.4% sugiriendo que el nodo \*32\* puede subdividirse en 2 o 3 grupos.

Cuadro 5.9: Resultados del PCA al nodo \*32\*

Nº CP	Eigenvalue	%Inercia	%Acumulado	Brokenstick	Rencher	g.l	p-value
1	6.49	72.09	72.09	2.83	3,658.46	44	0.00
2	1.21	13.47	85.55	1.83	1,749.09	35	0.00
3	0.57	6.36	91.91	1.33	1,149.05	27	0.00
4	0.40	4.41	96.33	1.00	779.49	20	0.00
5	0.17	1.91	98.24	0.75	302.32	14	0.00
6	0.05	0.51	98.75	0.55	21.70	9	0.01
7	0.05	0.50	99.25	0.38	17.30	5	0.00
8	0.04	0.45	99.70	0.24	10.46	2	0.01
9	0.03	0.30	100.00	0.11	0.00	0	1.00

**Nota:** g.l: Grados de libertad.

La magnitud de la correlación de las variables del nodo \*32\* con la variable representativa (\*32A\*) se observa en la tabla 5.10. Sólo hay dos variables que se correlacionan negativamente con la variable representativa del nodo \*32\*: *igbvl* y *encj*. El resto de variables tiene una correlación positiva. Debido a que la variable del tipo de cambio (*tc*) presenta una correlación positiva baja (0.58) también nos sugiere que el nodo \*32\* debe subdividirse en dos o más grupos. El estudio de las clases del nodo se analizarán más adelante.

Cuadro 5.10: Correlación de las variables del nodo \*32\* con sus variables representativas

Variable	*32A*	*32B*
dol	0.93	-0.23
tamex	0.92	0.35
tamn	0.89	0.30
tpmex	0.89	0.41
embig	0.86	-0.21
tpmn	0.86	0.40
tc	0.58	-0.53
encj	-0.79	0.31
igbvl	-0.86	0.38

También se ha calculado la correlación entre las variables representativas del nodo \*32\* (\*32A\* y \*32B\*) con los dos primeros componentes principales del nodo \*32\* en la tabla 5.11 y se ha encontrado que la variable representativa del nodo \*32\* (\*32A\*) tiene una correlación de 1 (negativa) con la primera componente y un correlación muy baja con la segunda componente principal.

Cuadro 5.11: Correlación entre variables representativas y componentes principales - \*32\*

Correlación	Componentes Principales		Representativas	
	PCA 1 *32*	PCA 2 *32*	*32A*	*32B*
PCA 1 *32*	1.00			
PCA 2 *32*	0.00	1.00		
*32A*	-1.00	0.01	1.00	
*32B*	-0.01	-0.97	0.00	1.00

### Nodo \*31\*

Este nodo es el tercer nivel del árbol de clasificación (Gráfico 5.6) que está compuesto por los nodos \*24\* y la variable *plca*.

En el gráfico 5.11 se muestra un círculo de correlaciones donde se observa claramente que la variable *plca* presenta un comportamiento distinto al resto de variables. Así mismo, si observamos el cuadro 5.5 notaremos que el nodo \*31\* se forma con la unión de la variable *plca* y el nodo \*24\*, y el segundo valor propio es 0.7 el cual es alto indicando que es un nodo no homogéneo. Sin embargo, el segundo valor propio del nodo \*24\* tiene un valor de 0.05 evidenciando que es un nodo homogéneo. En el gráfico 5.12 se muestra la variable representativa del nodo \*31\* donde se observa que hay un salto alrededor del 2018 producto del cambio regulatorio que afectó la variable *plca*. Por tal motivo, es necesario retirar esta variable y quedarse con el nodo \*24\* y la variable *plca* aislada.

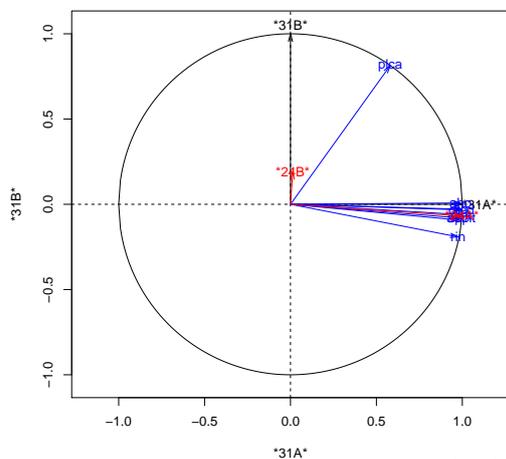


Gráfico 5.11: Correlaciones Nodo \*31\*

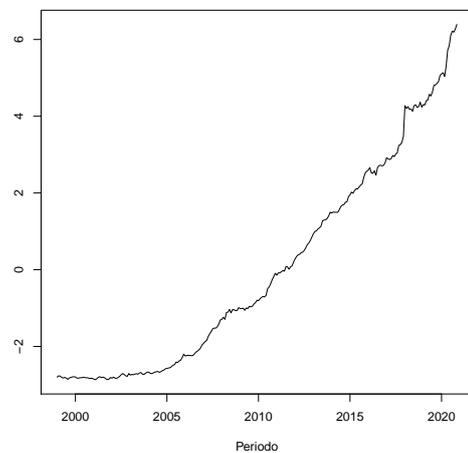


Gráfico 5.12: Variable representativa \*31\*

En la tabla 5.12 se muestra los valores propios del nodo \*31\*. Como se puede observar, el nodo \*31\* no es muy homogéneo porque el primer componente principal sólo explica el 89.9% de la inercia de las variables que componen el nodo \*31\*; el segundo valor propio explica el 9.0% sugiriendo que el nodo \*31\* puede subdividirse en 2 grupos. Más adelante analizaremos el nodo \*24\* donde se verificará que el primer valor propio de este nodo explica más de 98% confirmando que el nodo \*24\* es un nodo homogéneo.

Cuadro 5.12: Resultados del PCA al nodo \*31\*

Nº CP	Eigenvalue	%Inercia	%Acumulado	Brokenstick	Rencher	g.l	p-value
1	7.19	89.88	89.88	2.72	7,704.94	35	0.00
2	0.72	9.01	98.89	1.72	4,312.04	27	0.00
3	0.05	0.61	99.50	1.22	1,599.64	20	0.00
4	0.02	0.30	99.80	0.89	1,115.39	14	0.00
5	0.01	0.13	99.93	0.64	686.65	9	0.00
6	0.00	0.06	99.98	0.44	368.59	5	0.00
7	0.00	0.01	100.00	0.27	66.29	2	0.00
8	0.00	0.01	100.00	0.13	0.00	0	1.00

**Nota:** g.l: Grados de libertad.

La magnitud de la correlación de las variables del nodo \*31\* con la variable representativa (\*31A\*) se observa en la tabla 5.13. Todas las variables tienen una correlación positiva fuerte con la variable representativa excepto la variable *plca*. Este análisis refuerza la idea de que el nodo \*31\* debe separarse entre la variable *plca* y el nodo \*24\*.

Cuadro 5.13: Correlación de las variables del nodo \*31\* con sus variables representativas

Variable	*31A*	*31B*
liq	1.00	-0.03
cred	1.00	-0.03
clnt	0.99	-0.06
actnt	0.99	-0.08
aho	0.99	0.01
spp	0.98	-0.09
rin	0.98	-0.19
plca	0.58	0.81

También se ha calculado la correlación entre las variables representativas del nodo \*31\* (\*31A\* y \*31B\*) con los dos primeros componentes principales del nodo \*31\* en la tabla 5.14 y se ha encontrado que la variable representativa del nodo \*31\* (\*31A\*) tiene una correlación muy cercana a 1 con la primera componente pero la variable representativa \*32B\*, que mide las diferencias del nodo \*32\*, tiene una correlación cercana a 1 con el segundo componente principal del nodo \*32\*. Por lo tanto, este análisis nos indica que el nodo \*32\* debe separarse en dos grupos: la variable *plca* y el nodo \*24\*.

Cuadro 5.14: Correlación entre variables representativas y componentes principales - \*31\*

Correlación	Componentes Principales		Representativas	
	PCA 1 *31*	PCA 2 *31*	*31A*	*31B*
PCA 1 *31*	1.00			
PCA 2 *31*	0.00	1.00		
*31A*	1.00	0.00	1.00	
*31B*	0.00	1.00	0.00	1.00

### Nodo \*30\*

Este nodo es el cuarto nivel del árbol de clasificación (Gráfico 5.6) que está compuesto por los nodos \*29\* y la variable *tc*. En el gráfico 5.13 se muestra un círculo de correlaciones donde se observa un dipolo entre el conjunto de variables *dol* y *embig*; y *igbvl* y *encj*. Así mismo, se observa que la variable *tc* parece independiente del resto de variables. En el gráfico 5.14 se muestra la variable representativa de la clase \*30\*. Por otro lado, es necesario analizar adecuadamente si la variable *tc* debe formar parte del nodo \*30\* o debe analizarse por separado.

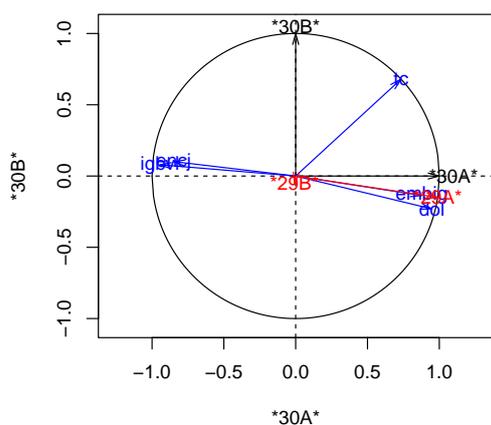


Gráfico 5.13: Correlaciones Nodo \*30\*

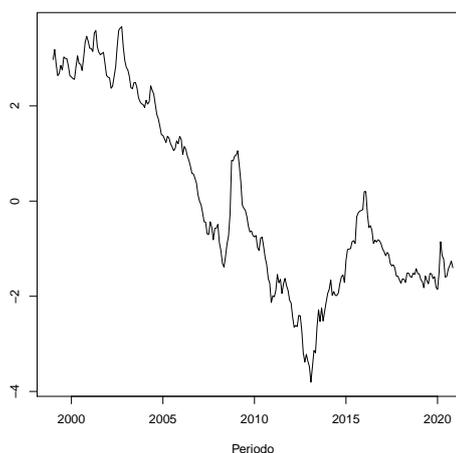


Gráfico 5.14: Variable representativa \*30\*

La variable representativa se muestra en 5.14 el cual muestra una tendencia negativa hasta el periodo 2013 y luego presenta una tendencia positiva. A menor valor de la variable representativa significa que el mercado financiero está en periodo de auge, mientras que cuando crece la variable el mercado financiero se contrae. Hasta el periodo 2013, el tipo de cambio (*tc*) y el coeficiente de dolarización (*dol*) mostraban una tendencia negativa. Tal como hemos indicado el caso del nodo \*33\*, la reducción del tipo de cambio significa que los inversionistas observan que el mercado financiero peruano como un mercado sólido y pueden invertir dentro del país. De igual manera, una disminución del Embig Perú (*embig*) indica que la economía peruana tiene menor riesgo para invertir. A partir del 2013, ambas variables mostraron una tendencia positiva. Así mismo, el índice bursátil (*igbvl*) presentó disminuciones en su nivel indicando que los inversionistas mostraban poco interés en las acciones de las empresas peruanas que cotizan en la bolsa de valores. También se observó un decremento del nivel de encaje bancario (*encj*) debido a las menores colocaciones bancarias. Actualmente, debido al contexto de la pandemia la variable representativa del nodo \*30\* se mantiene en niveles altos a comparación de su periodo más bajo en 2013 indicando que el mercado financiero peruano se encuentra contraído.

En la tabla 5.15 se muestra los valores propios del nodo \*30\*. Como se puede observar, el nodo \*30\* no es muy homogéneo porque el primer componente principal sólo explica el 76.4 % de la inercia de las variables que componen el nodo \*30\*; el segundo y tercer valor propio explican el 11.0 % y 8.5 % respectivamente, sugiriendo que el nodo \*30\* puede subdividirse en 2 o más grupos.

Cuadro 5.15: Resultados del PCA al nodo \*30\*

Nº CP	Eigenvalue	%Inercia	%Acumulado	Brokenstick	Rencher	g.l	p-value
1	3.82	76.41	76.41	2.28	1,290.32	14	0.00
2	0.55	11.03	87.44	1.28	370.38	9	0.00
3	0.42	8.47	95.91	0.78	266.10	5	0.00
4	0.15	3.08	98.99	0.45	77.22	2	0.00
5	0.05	1.01	100.00	0.20	0.00	0	1.00

**Nota:** g.l: Grados de libertad.

En la tabla 5.16 se muestra la correlación de las variables del nodo \*30\* con la variable representativa \*30A\*. La variable *tc* presenta una correlación positiva con la variable representativa de 0.73. Como magnitud, es la variable que tiene menos correlación con la variable representativa \*30A\*. Sin embargo, la variable *tc* tiene la más alta correlación con la segunda variable representativa \*30B\* la cual mide las diferencias del grupo. De esta manera, podemos decir que la variable *tc* debe analizarse por separado del resto de variables que componen el nodo \*30\*. El sentido económico es que el tipo de cambio (*tc*), a diferencia del resto de variables que componen el nodo \*30\*, es una variable que no se puede predecir con facilidad pues mide la aversión al riesgo que tienen los inversionistas sobre un país. En el Perú, el tipo de cambio fluctúa con mucha facilidad a pesar de la intervención monetaria que realiza el Banco Central de Reservas para estabilizar su nivel ante un escenario económico y/o político negativo. Por este motivo, es mejor separar las variables del nodo \*29\* y el *tc*.

Cuadro 5.16: Correlación de las variables del nodo \*30\* con sus variables representativas

Variable	*30A*	*30B*
dol	0.95	-0.23
embig	0.88	-0.14
tc	0.73	0.68
encj	-0.85	0.10
igbvl	-0.95	0.08

También se ha calculado la correlación entre las variables representativas del nodo \*30\* (\*30A\* y \*30B\*) con los dos primeros componentes principales del nodo \*30\* en la tabla 5.17 y se ha encontrado que la variable representativa del nodo \*30\* (\*30A\*) tiene una correlación negativa muy cercana a 1 con la primera componente. De igual manera, la variable representativa \*30B\*, que mide las diferencias del nodo \*30\*, tiene una correlación negativa cercana a 1 con el segundo componente principal del nodo \*30\*. Por lo tanto, este análisis nos indica que el nodo \*30\* debe separarse en dos grupos: la variable *tc* y el nodo \*29\*.

Cuadro 5.17: Correlación entre variables representativas y componentes principales - \*30\*

Correlación	Componentes Principales		Representativas	
	PCA 1 *30*	PCA 2 *30*	*30A*	*30B*
PCA 1 *30*	1.00			
PCA 2 *30*	0.00	1.00		
*30A*	-1.00	0.00	1.00	
*30B*	0.00	-1.00	0.00	1.00

### 5.2.2. Estudio de las clases

En esta sección se presentan los gráficos de correlaciones de las clases \*29\*, \*27\* y \*24\*:

#### Clase \*29\*

Esta clase está compuesto por series con mucha volatilidad como el riesgo país (medido con el *embig*) y el índice de precio de acciones (*igbvl*). En el gráfico 5.16 se muestra un círculo de correlaciones donde las variables *embig* y *dol* tienen correlación positiva con la primera variable representativa, mientras que las variables *encj* y *igbvl* tienen correlación negativa.

Debido al dipolo que se forma con las variables, se elaboró un gráfico cambiando el signo de las variables correlacionadas positivamente y negativamente con la variable representativa y se comparó con la tendencia de la variable representativa. Como se observa en el gráfico 5.15 la variable representativa tiene una tendencia similar a las variables correladas positivamente (*embig* y *dol*). Sin embargo, el sentido económico mejor nos indica que cuando el mercado financiero peruano muestra signos de expansión se debe a una disminución del coeficiente de dolarización (*dol*) y del riesgo país (*embig*) pues los inversionistas extranjeros consideran que el mercado financiero tiene bajo riesgo y menor probabilidad de una crisis cambiaria. Pero cuando el índice bursátil (*igbvl*) disminuye se debe a la menor demanda de las acciones de las empresas peruanas que cotizan en la Bolsa de Valores de Lima y esto se traduce en una reducción del mercado financiero peruano. De igual manera, cuando se reduce el volumen de créditos otorgado por las entidades bancarias también se reduce el encaje bancario generando (*encj*) generando un impacto negativo en el mercado financiero peruano. Por estas razones, se considerará que el signo de la variable representativa del nodo 29 sea con el signo cambiando, tal como se muestra en el gráfico 5.17.

Al igual que la explicación del nodo \*30\*, esta variable representativa muestra una tendencia negativa hasta el periodo 2013 y luego presenta una tendencia positiva. Hasta el periodo 2013, el mercado financiero peruano mostraba signos de expansión debido a que el coeficiente de dolarización (*dol*) estaba disminuyendo, así como el riesgo país (*embig*). Sin embargo, a partir del 2013 el índice bursátil (*igbvl*) disminuyó por la menor cotizaciones de las empresas peruanas. De igual manera, el volumen de créditos se redujo generando menores niveles de encaje bancario (*encj*). Por estos motivos, luego del 2013 el mercado financiero peruano ha mostrado signos de contracción financiera. Por otro lado, es importante mencionar que la fluctuación de la variable representativa del nodo \*30\* y \*29\* son similares indicando que el retiro de la variable *tc* es apropiado y no altera la interpretación de los resultados.

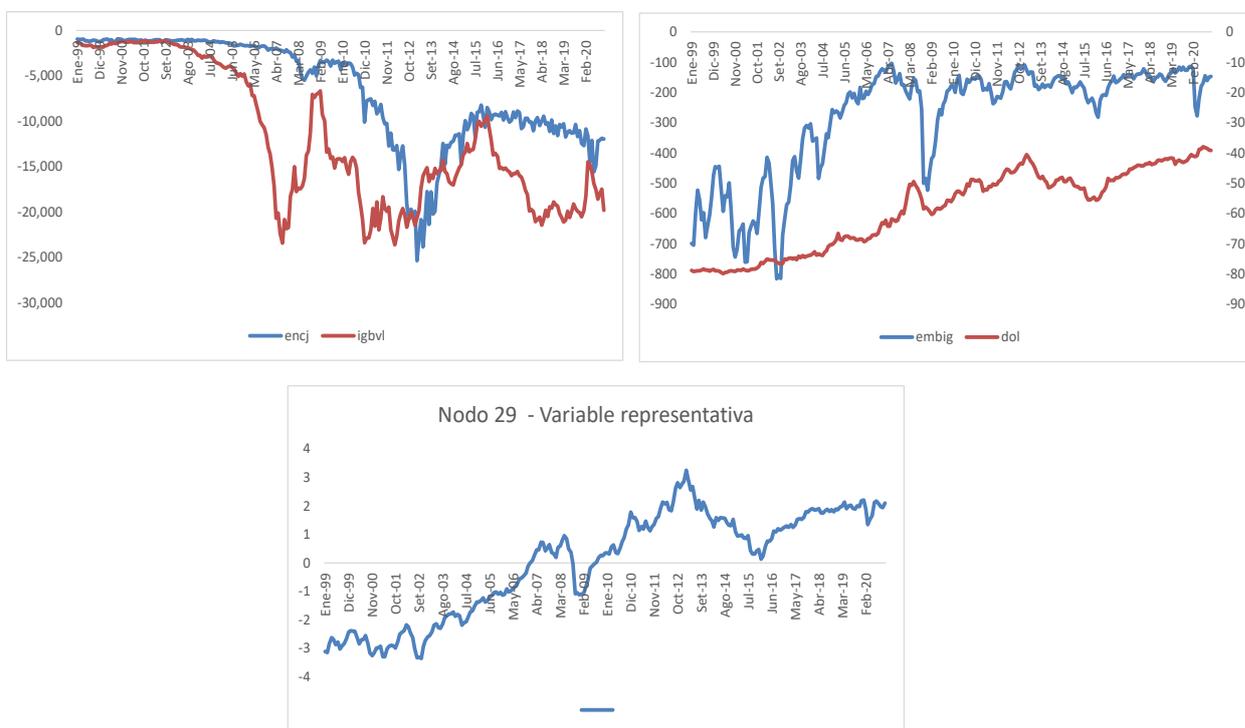


Gráfico 5.15: Variable representativa y sus componentes

En el gráfico 5.17 se muestra la variable representativa de esta clase la cual es similar al 5.14 indicando que la separación de la variable *tc* del nodo \*30\* no altera los resultados.

En la tabla 5.18, se muestra los resultados del análisis de PCA sobre la clase \*29\*. Se observa que el primer componente principal tiene un valor de inercia de 84.0%. Aunque esta participación es baja, es un grupo más homogéneo que la clase \*30\*.

Cuadro 5.18: Resultados del PCA al nodo \*29\*

Nº CP	Eigenvalue	%Inercia	%Acumulado	Brokenstick	Rencher	g.l	p-value
1	3.36	83.97	83.97	2.08	1,110.10	9	0.00
2	0.42	10.58	94.55	1.08	222.28	5	0.00
3	0.15	3.86	98.41	0.58	49.74	2	0.00
4	0.06	1.59	100.00	0.25	0.00	0	1.00

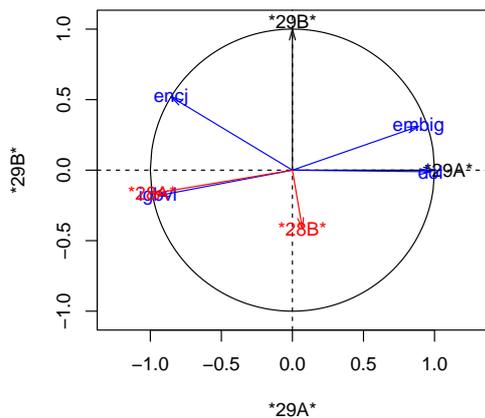


Gráfico 5.16: Correlaciones Nodo \*29\*

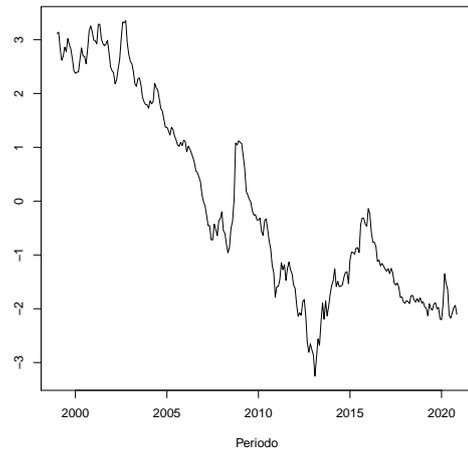


Gráfico 5.17: Variable representativa \*29\*

**Nota:** g.l: Grados de libertad.

En la tabla 5.19 se muestra la correlación de las variables de la clase \*29\* con la variable representativa de \*29\*. Todas las variables tienen una correlación por encima de 80 %, lo cual evidencia que es un grupo más homogéneo que el nodo \*30\*.

Cuadro 5.19: Correlación de las variables del nodo \*29\* con sus variables representativas

Variable	*29A*	*29B*
dol	0.97	0.01
embig	0.89	-0.31
encj	-0.85	-0.52
igbvl	-0.95	0.18

También se ha calculado la correlación entre las variables representativas de la clase \*29\* (\*29\*A y \*29B\*) con los dos primeros componentes principales de \*29\* en la tabla 5.20 y se ha encontrado que la variable representativa del nodo \*29\* (\*29A\*) tiene una correlación positiva muy cercana a 1 con la primera componente. Por lo tanto, este análisis nos indica que el nodo \*29\* debe considerarse como un nodo homogéneo.

Cuadro 5.20: Correlación entre variables representativas y componentes principales - \*29\*

Correlación	Componentes Principales		Representativas	
	PCA 1 *29*	PCA 2 *29*	*29A*	*29B*
PCA 1 *29*	1.00			
PCA 2 *29*	0.00	1.00		
*29A*	1.00	0.00	1.00	
*29B*	0.00	-0.96	0.00	1.00

Estos análisis nos sugieren que todas las variables de la clase \*29\* forman un grupo homogéneo y la variable representativa de la clase \*29\* los representa adecuadamente.

Por otro lado, en el gráfico 5.18 se muestra un comparativo del tercer componente obtenido de todas las series (gráfico 5.5) y de las variables representativas de las clases \*30\* y \*29\*.

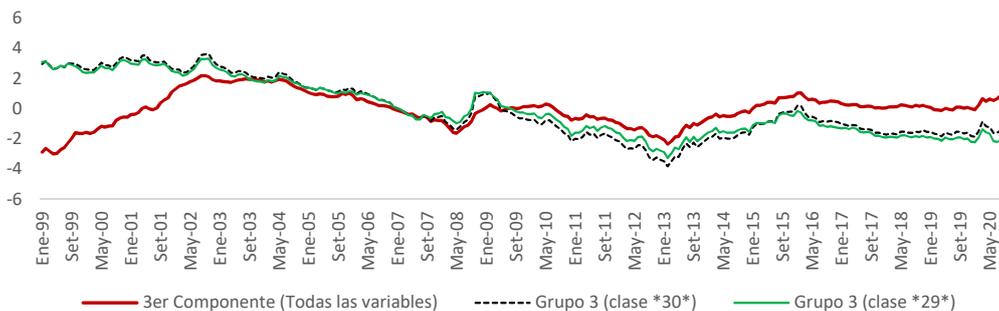


Gráfico 5.18: Componentes Principales - Series Volátiles

La correlación entre estas tres variables se muestra en la tabla 5.21:

Cuadro 5.21: Correlación entre variables representativas del nodo \*29\*, \*30\* y la 3era componente principal obtenidas con todas las variables

Correlación	Grupo 3 (clase *30*)	Grupo 3 (clase *29*)	3er Componente
Grupo 3 (clase *30*)	1.00		
Grupo 3 (clase *29*)	0.99	1.00	
3er Componente	0.33	0.26	1.00

### Clase \*27\*

Esta clase es el quinto nivel del árbol de clasificación (Gráfico 5.6) que está compuesto por los nodos \*26\* y la variable *tamn*. Esta clase está compuesto solamente por las tasas pasivas y activas. En el gráfico 5.19 se muestra un círculo de correlaciones donde todas las variables tienen correlación positiva con la variable representativa. En el gráfico 5.20 se muestra la tendencia de la variable representativa de la clase \*27\*. La tendencia es negativa es producto de que en los últimos años se ha incrementado la competencia bancaria por captar mayores clientes generando que las tasas de interés disminuyen a lo largo del tiempo.

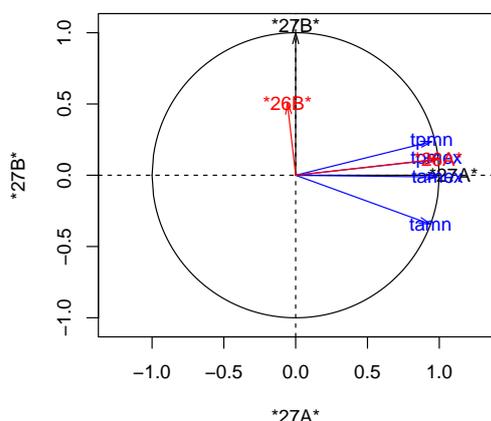


Gráfico 5.19: Correlaciones Nodo \*27\*

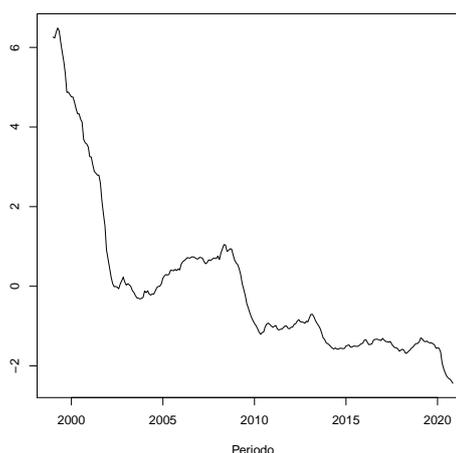


Gráfico 5.20: Variable representativa \*27\*

En la tabla 5.22 se muestra los valores propios del nodo \*27\*. Como se puede observar, el nodo \*27\* es muy homogéneo porque el primer componente principal explica el 92.8 % de la inercia de las variables.

Cuadro 5.22: Resultados del PCA al nodo \*27\*

Nº CP	Eigenvalue	%Inercia	%Acumulado	Brokenstick	Rencher	g.l	p-value
1	3.71	92.81	92.81	2.08	1,728.10	9	0.00
2	0.20	4.99	97.80	1.08	240.94	5	0.00
3	0.06	1.44	99.24	0.58	26.27	2	0.00
4	0.03	0.76	100.00	0.25	0.00	0	1.00

**Nota:** g.l: Grados de libertad.

En la tabla 5.23 se muestra la correlación de las variables del nodo \*27\* con la variable representativa \*27A\*. Se observa que todas las variables tienen una alta correlación (positiva) mayor a 90 %. Este análisis indica que el nodo \*27\* es un nodo homogéneo.

Cuadro 5.23: Correlación de las variables del nodo \*27\* con sus variables representativas

Variable	*27A*	*27B*
tamn	0.94	0.34
tpmn	0.95	-0.24
tpmex	0.98	-0.11
tamex	0.99	0.01

También se ha calculado la correlación entre las variables representativas del nodo \*27\* (\*27A y \*27B\*) con los dos primeros componentes principales del nodo \*27\* en la tabla 5.24 y se ha encontrado que la variable representativa del nodo \*27\* (\*27A\*) tiene una correlación negativa muy cercana a 1 con la primera componente. Por lo tanto, este análisis nos indica también que el nodo \*27\* es un nodo homogéneo y no debe separarse.

Cuadro 5.24: Correlación entre variables representativas y componentes principales - \*27\*

Correlación	Componentes Principales		Representativas	
	PCA 1 *27*	PCA 2 *27*	*27A*	*27B*
PCA 1 *27*	1.00			
PCA 2 *27*	0.00	1.00		
*27A*	1.00	0.00	1.00	
*27B*	0.00	0.95	0.00	1.00

### Clase \*24\*

Esta clase es el sexto nivel del árbol de clasificación (Gráfico 5.6) que está compuesto por los nodos \*21\* y \*22\*. En el gráfico 5.21 se muestra un círculo de correlaciones donde se observa que todas las variables están altamente correlacionadas. En el gráfico 5.22 se muestra la variable representativa la cual con respecto a la variable representativa del nodo \*31\* ya no muestra un cambio abrupto alrededor del 2018, siendo que este cambio solo se encontraba en la variable *plca*. La variable representativa tiene una pendiente positiva debido a que las series que componen esta variable también tienen una tendencia positiva. Sin embargo, se observa que hasta el año 2005 la variable representativa tiene un crecimiento casi nulo, pero

posteriormente comienza a crecer a mayores tasas.

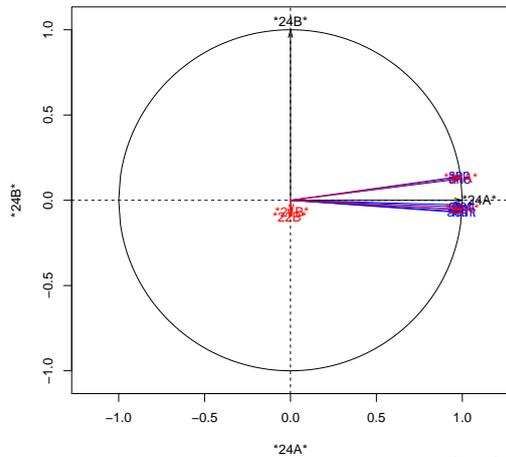


Gráfico 5.21: Correlaciones Nodo \*24\*

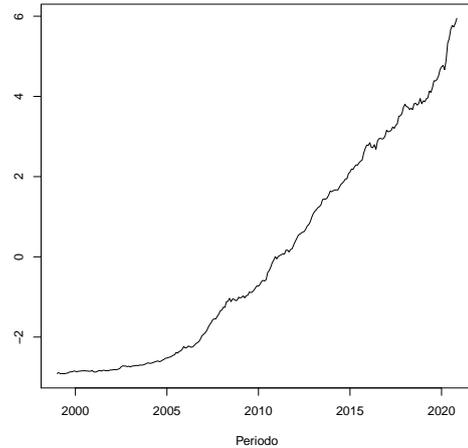


Gráfico 5.22: Variable representativa \*24\*

En la tabla 5.25 se muestra los valores propios del nodo \*24\*. Como se puede observar, el nodo \*24\* es muy homogéneo porque el primer componente principal explica el 98.4 % de la inercia de las variables que componen el nodo \*24\*.

Cuadro 5.25: Resultados del PCA al nodo \*24\*

Nº CP	Eigenvalue	%Inercia	%Acumulado	Brokenstick	Rencher	g.l	p-value
1	6.89	98.36	98.36	2.59	7,379.03	27	0.00
2	0.05	0.74	99.11	1.59	1,731.57	20	0.00
3	0.05	0.65	99.76	1.09	1,444.85	14	0.00
4	0.01	0.15	99.90	0.76	664.10	9	0.00
5	0.01	0.07	99.98	0.51	394.27	5	0.00
6	0.00	0.02	99.99	0.31	70.14	2	0.00
7	0.00	0.01	100.00	0.14	0.00	0	1.00

**Nota:** g.l: Grados de libertad.

En la tabla 5.26 se muestra la correlación de las variables del nodo \*24\* con la variable representativa \*24A\*. Se observa que todas las variables tiene una correlación cercana a 1 con la variable representativa \*24A\*. Por lo tanto podemos decir que el nodo \*24\* es un nodo homogéneo y que no requiere que se separe sus variables.

Cuadro 5.26: Correlación de las variables del nodo \*24\* con sus variables representativas

Variable	*30A*	*30B*
actnt	1.00	-0.07
clnt	1.00	-0.04
liq	1.00	-0.06
cred	0.99	-0.03
aho	0.99	0.13
rin	0.99	-0.07
spp	0.99	0.14

También se ha calculado la correlación entre las variables representativas del nodo \*24\* (\*24\*A y \*24B\*) con los dos primeros componentes principales del nodo \*24\* en la tabla 5.27 y se ha encontrado que la variable representativa del nodo \*24\* (\*24A\*) tiene una correlación muy cercana a 1 con la primera componente. Por lo tanto, este análisis nos indica que el nodo \*24\* debe considerarse con un nodo unidimensional.

Cuadro 5.27: Correlación entre variables representativas y componentes principales - \*24\*

Correlación	Componentes Principales		Representativas	
	PCA 1 *24*	PCA 2 *24*	*24A*	*24B*
PCA 1 *24*	1.00			
PCA 2 *24*	0.00	1.00		
*24A*	1.00	0.00	1.00	
*24B*	0.00	0.87	0.00	1.00

### 5.3. Síntesis de los resultados

En base a los resultados obtenidos en el análisis anterior, concluimos que las variables deben clasificarse en 5 grupos. Esta clasificación nos asegura que exista homogeneidad dentro de cada grupo. Solo hay dos grupos conformado por una sola variable, *tc* y *plca* respectivamente.

- Grupo 1 (clase \*24\*): Este grupo está conformado por las series con tendencia positiva. La variable representativa de este grupo también presenta una tendencia positiva (ver gráfico 5.22) y mide la tendencia de largo plazo del ciclo financiero. Económicamente son los fenómenos que ocurren en el ciclo financiero que tienen larga duración en el sistema financiero. Por ejemplo, la variable de ahorro (*aho*) mide las decisiones de ahorro de las personas para un gastar o invertir en el futuro. Así mismo, el monto que invierten las AFP (*spp*) es destinado a activos financieros que otorgan un rendimiento de largo plazo pues está asociado al tiempo de jubilación de las personas. En el caso de las reservas internacionales RIN (*rin*), éstas tienen por objetivo afrontar choques macroeconómicos negativos de liquidez. Dado que estos choques no existen con frecuencia en el Perú, las RIN se han acumulado a lo largo del tiempo y se pueden usar para ante un escenario adverso. El circulante (*clnt*) es una variable financiera agregada que mide la cantidad de billetes y monedas que emite el Banco Central y que se encuentran en poder del público. Por ello, el monto del circulante sigue creciendo debido a la preferencia que tienen las personas por mantener efectivo y realizar sus transacciones. El volumen de créditos (*cred*) también muestra una tendencia creciente debido a la postura que tienen los bancos de incrementar las ofertas crediticia y ganar participación de mercado. Esta situación trae consigo, además, que los bancos incremente sus activos internos netos (*actnt*). Finalmente, la liquidez de las empresas bancarias (*liq*) indican la disposición de dinero que tienen los bancos para afrontar sus operaciones y posibles crisis financieras. Dado que no han existido muchas crisis financieras en el Perú, esta variable sigue aumentando y es una señal positiva para el resto de inversionista porque los bancos tienen la capacidad de afrontar periodos macroeconómicos adversos.
- Grupo 2 (clase \*27\*): Este grupo está conformado por series con tendencia negativa

como las tasas activas y pasivas en el sistema financiero. Este grupo de variable son importantes en el sistema financiero pues determina las decisiones de ahorros de las familias y el volumen de créditos que desembolsan los bancos. La tendencia negativa se explica debido a que a inicios de 1999 existían pocos competidores dentro de la banca peruana, por lo tanto para captar ahorros las entidades bancarias ofrecían pagar tasas de ahorros altas. Así mismo, para desembolsar créditos los bancos podían cobrar tasas activas altas debido a la escasez de competidores. Sin embargo, conforme han transcurrido los años se han incorporado nuevas instituciones financieras en el mercado peruano generando que los bancos reduzcan sus tasas activas y consecuentemente sus tasas pasivas.

- Grupo 3 (clase \*29\*): Este grupo está conformado por las series con volatilidad financiera. La volatilidad del índice bursátil peruano (*igbvl*) puede ser ocasionado por el la volatilidad en el sector económico al que pertenecen las empresas. Por ejemplo, las ventas de las empresas minera dependen de la cotización internacional de los precios de los *commodities*. Por otro lado, la variación en el EMBIG Perú (*embig*) también tiene su origen en los mercados internacionales pues un incremento en este indicador significa que los inversionista tienen mayor al riesgo para invertir en el Perú. El monto de encaje es un monto de dinero que los bancos destinan como ahorros en el Banco Central de Reserva para controlar la expansión de los créditos. El indicador de encaje es volátil pues todos los días este monto varía dependiendo del volumen de créditos otorgados por los bancos. Finalmente, el coeficiente de dolarización (*dol*) mide el volumen de activos que posee el Sistema Financiero en moneda extranjera. Esta variable permite dimensionar el impacto que puede tener una crisis cambiaria en el país.
- Grupo 4: Es un grupo conformado por la del tipo de cambio (*tc*). De acuerdo a los análisis realizado en la secciones previas, el tipo de cambio es una variable que no se puede predecir con facilidad y que varía diariamente en los mercados financieros internacionales. A pesar que actualmente el Banco Central de Reserva interviene en el mercado cambiario para reducir su volatilidad, es poco probable alcanzar un nivel deseable para una economía pues su valor depende de las decisiones que toman los agentes económicos internacionales.
- Grupo 5: Es un grupo conformado por la palanca global de las empresas bancarias (*plca*). Este indicador considera el patrimonio como porcentaje de los activos. Regulatoriamente, hasta el 2018 las instituciones financieras debían mantener un ratio mínimo de 9.5 %, sin embargo luego del 2018 se produjo un cambio regulatorio el cual indicaba que el ratio mínimo debía ser 15 %.

En el gráfico 5.23 se muestran las variables representativas de los grupos 1, 2 y 3 respectivamente.

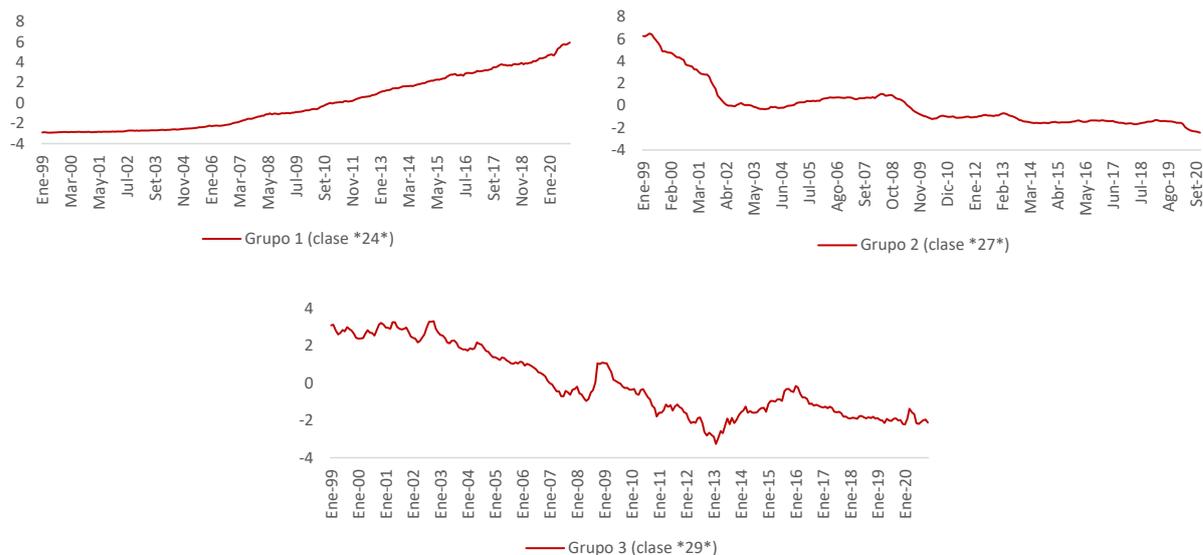


Gráfico 5.23: Variables representativas grupo 1, 2 y 3

En el gráfico 5.24 se muestra las variables  $tc$  y  $plca$  las cuales conforman el grupo 4 y 5 respectivamente.



Gráfico 5.24: Variables  $tc$  y  $plca$

Es importante mencionar que existe una semejanza entre la variable representativa del nodo \*29\* y el tipo de cambio ( $tc$ ), sin embargo se decidió separar  $tc$  del nodo \*29\* porque es una variable difícil de predecir a diferencia del resto de variables que conforman el nodo \*29\*. Además, la variable representativa del nodo \*30\* y \*29\* tienen una alta correlación (98.9 %) indicando que el retiro de la variable  $tc$  del nodo \*30\* no altera la interpretación de los resultados.

En la tabla 5.28 se muestra la correlación entre las variables representativas de los 3 grupos formados y las dos series aisladas:

Cuadro 5.28: Correlación entre los 5 grupos

Correlación	<i>plca</i>	<i>tc</i>	<i>Grupo1</i>	<i>Grupo2</i>	<i>Grupo3</i>
<i>plca</i>	1				
<i>tc</i>	0.29	1			
<i>Grupo1</i>	0.52	-0.17	1		
<i>Grupo2</i>	-0.17	0.33	-0.71	1	
<i>Grupo3</i>	-0.29	0.62	-0.83	0.75	1

## 5.4. Análisis Evolutivo

El análisis evolutivo es una metodología innovadora que tiene por objetivo estudiar si la estructura de correlación varía a lo largo del tiempo y cómo está variando. En la sección 4.3 se describió la metodología. En un análisis previo, con una ventana móvil de 24 meses, se ha observado una variabilidad excesiva, en particular en el caso de los 3 grupos, donde se esperaba una mayor homogeneidad. Esto se ha interpretado como fluctuaciones de corto plazo de cada variable, las cuales no necesariamente se encuentran correlacionadas. Por este motivo, se propone un análisis con una ventana de 48 meses, donde se obtiene una mejor homogeneidad. En este último análisis, los parámetros elegidos para el desarrollo del EPCA fue una ventana móvil de 48 meses y 3 desviaciones estándar. Como la PCA ha mostrado diferencias importantes entre las variables empleadas y la HFC ha sugerido cinco grupos de variables, en esta sección ejecutaremos el análisis evolutivo con las variables representativas de los grupo 1, 2, 3 y las dos variables aisladas (*tc* y *plca*).

### 5.4.1. Análisis evolutivo sobre las 5 variables representativas

Se aplicó la metodología de análisis evolutivo (EPCA) a los 3 grupos definidos anteriormente junto con las dos variables aisladas (*tc* y *plca*). Como se mencionó anteriormente, todos los análisis evolutivos se han realizado considerando una ventana móvil de 48 meses y 3 desviaciones estándar para calcular los pesos. En el gráfico 5.25 se muestran los tres primeros valores propios del evolutivo.

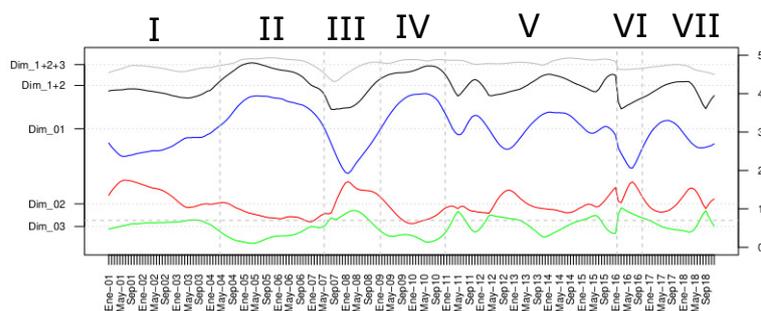


Gráfico 5.25: Evolutivo de los componentes principales - Variables Representativas

El algoritmo de Fisher para identificar los puntos de corte, a través del método de Calinski-Harabász, se muestra en el gráfico 5.26.

n	Within ssq	Indice Calinski-Harabasz	Puntos finales												
			1	2	3	4	5	6	7	8	9	10	11	12	
1	60.65	0.000	0	0	0	0	0	0	0	0	0	0	0	0	0
2	50.90	20.479	34	216	0	0	0	0	0	0	0	0	0	0	0
3	38.93	39.600	41	74	216	0	0	0	0	0	0	0	0	0	0
4	30.50	52.379	40	78	94	216	0	0	0	0	0	0	0	0	0
5	20.16	84.725	40	77	97	121	216	0	0	0	0	0	0	0	0
6	17.11	89.063	40	77	97	120	180	216	0	0	0	0	0	0	0
7	14.52	94.883	40	77	97	120	181	190	216	0	0	0	0	0	0
8	12.73	97.884	25	43	77	97	120	181	190	216	0	0	0	0	0
9	11.30	100.484	25	43	77	94	101	120	181	190	216	0	0	0	0
10	9.98	104.601	25	43	74	80	93	100	120	181	190	216	0	0	0
11	8.68	111.553	25	43	74	80	93	100	120	181	191	205	216	0	0
12	7.40	122.335	25	43	74	80	93	100	120	137	148	181	190	216	0

Gráfico 5.26: Algoritmo de Fisher - 5 variables representativas

Debido a que el algoritmo de Fisher no encuentra un máximo local se ha elegido considerar siete grupos de partición de la muestra (I, II, III, IV, V, VI y VII) ya que con dichos grupos se alcanza una varianza más regular. En el gráfico 5.25 se muestra los siete periodos indicados en la parte superior del gráfico. Así mismo, del gráfico es evidente que existe una correlación negativa entre las series de los dos primeros valores propios del evolutivo, mientras que el tercer valor propio resulta poco significativo ya que en pocos periodos se encuentra por encima de 0.7. Por el contrario, la serie del segundo valor propio sólo en dos ocasiones se encuentra por debajo de 0.7, así que puede considerarse como una serie relevante. La serie del primer valor propio tiene variaciones relevantes: en dos períodos (II y IV) se encuentra arriba de su promedio; en el periodo V la serie está alrededor de este promedio indicando que se trata de un periodo de mayor homogeneidad, mientras que en los demás periodos la homogeneidad es menor.

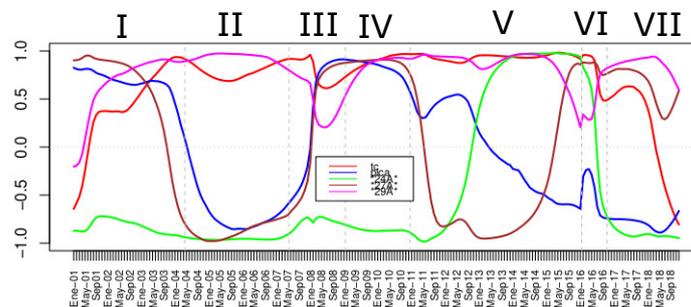


Gráfico 5.27: Correlación 1era componente principal - 5 variables representativas

En el grafico 5.27 se encuentran las series de correlaciones de las 5 variables representativas con la series de la primera componente principal. Se observa un patrón variado, con \*24A\* en oposición con las demás variables excepto en los últimos periodos y casi siempre opuesta a las variables \*29A\* y *tc*. Por otro lado, las variables *plca* y \*27A\* varían su signo de su correlación a lo largo del tiempo. Se puede sintetizar el patrón diciendo que los primeros factores representan la oposición entre *tc* y \*29A\* con \*24A\*, con un patrón entre *tc* y \*29A\* bastante parecido (correspondiente a la correlación que se había detectado en el análisis de PCA general).

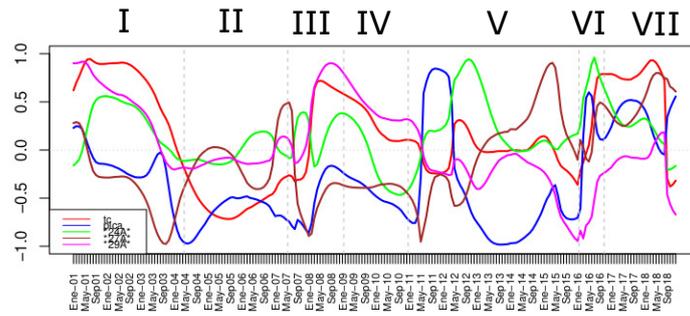


Gráfico 5.28: Correlación 2da componente principal - 5 variables representativas

En el gráfico 5.28 se muestra las series de las correlaciones entre las series y la segunda componente principal. Se nota un patrón irregular de difícil interpretación, excepto en el periodo II y VII ya que son casi siempre todas las series son decrecientes y crecientes, respectivamente. La dificultad de interpretación es aún mayor en los períodos menos homogéneos.

En las siguientes secciones se muestra el análisis evolutivo de los tres grupos retirando las variables *plca* y *tc* debido a que la primera tiene un comportamiento diferente a las demás, mientras que la otra variable al ser retirada del nodo \*30\* resulta en un grupo más homogéneo.

#### 5.4.2. Análisis evolutivo del primer grupo

El análisis evolutivo sobre las variables del primer grupo se encuentra en el gráfico 5.29 mostrando la tabla de cortes:

n	Within ssq	Indice Calinski-Harabasz	Puntos finales												
			1	2	3	4	5	6	7	8	9	10	11	12	
1	156.05	0.000	0	0	0	0	0	0	0	0	0	0	0	0	0
2	61.65	163.847	36	216	0	0	0	0	0	0	0	0	0	0	0
3	27.38	333.596	36	184	216	0	0	0	0	0	0	0	0	0	0
4	22.76	310.473	36	108	182	216	0	0	0	0	0	0	0	0	0
5	12.85	470.297	37	81	104	182	216	0	0	0	0	0	0	0	0
6	9.59	534.400	35	42	80	104	182	216	0	0	0	0	0	0	0
7	7.21	616.632	35	42	80	105	176	187	216	0	0	0	0	0	0
8	6.31	616.520	35	42	76	84	104	176	187	216	0	0	0	0	0
9	5.38	643.562	34	38	44	76	84	104	176	187	216	0	0	0	0
10	4.66	669.811	34	38	44	76	85	102	108	176	187	216	0	0	0
11	4.02	704.189	15	33	37	44	76	85	102	108	176	187	216	0	0
12	3.50	742.033	15	33	37	44	76	85	102	109	173	182	189	216	0

Gráfico 5.29: Algoritmo de Fisher - Grupo 1

En base a los resultados de 5.29, se eligieron 7 grupos, correspondientes al máximo valor según el criterio de Calinski-Harabasz.

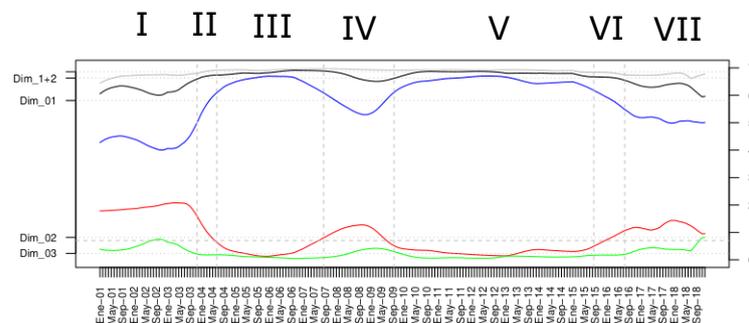


Gráfico 5.30: Evolutivo de los componentes principales - Grupo 1

En el gráfico 5.30 se encuentra una fuerte homogeneidad, excepto en el periodo I y VII donde los segundos valores propios son mayores a 0.7 pero menores que los primeros valores propios.

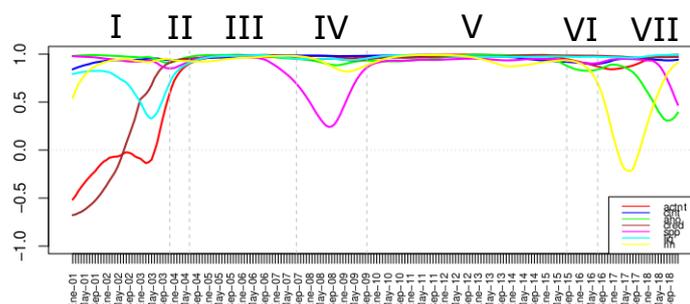


Gráfico 5.31: Correlación 1era componente principal - Grupo 1

En el gráfico 5.31, se encuentra una mayor homogeneidad debido a las altas correlaciones, excepto que en el periodo I las variables *actnt* y *cred* están inicialmente en oposición y al final de dicho periodo la variable *liq* está menos correlacionado. Posteriormente, en el periodo IV se reduce la correlación de *spp* y en el último periodo *rin* se pone en oposición, así como las variables *aho* y *spp*.

En el gráfico 5.32 se encuentran la correlación de la series con la segunda componente principal.

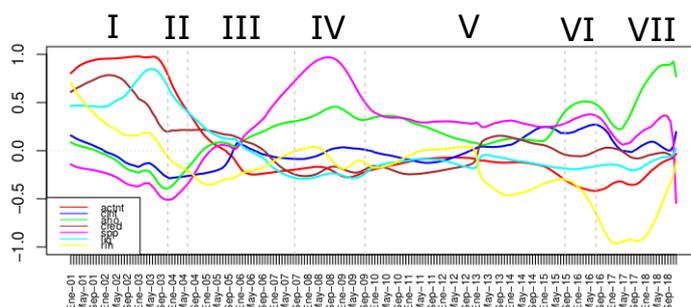


Gráfico 5.32: Correlación 2da componente principal - Grupo 1

En general se observa que las correlaciones son muy bajas excepto en el periodo I y VII, aunque con algunas series opuesta; y en el periodo IV donde la variable *spp* presenta una fuerte correlación con la segunda componente.

Los gráficos 5.31 y 5.32 indican que las series se han mantenido estables durante el periodo de muestra analizado excepto durante el periodo 2007-2008 el cual está caracterizado por la crisis financiera en esos años. Esto indica que las series son susceptibles a la fluctuación de los mercados financieros. Al igual que Stremmel (2015) se encontró que las variables financieras de las entidades bancarias son relevantes para explicar las fluctuaciones del ciclo financiero, principalmente los cambios de largo plazo. Así mismo, otra variable importante del primer grupo es el circulante. De acuerdo a Ramírez y Vásquez (2014) esta variable es importante en periodos de crisis pues las personas tienden a demandar más efectivo mientras el PBI se desacelera. Las reservas internacionales (RIN) forman parte del primer grupo. Esta variable es usada por el BCRP para afrontar choques de liquidez internacional en crisis financieras.

En el gráfico 5.33 se muestra la comparación entre la primera componente principal, primera variable representativa, y promedio global. La primera componente y la variable representativa se superponen.

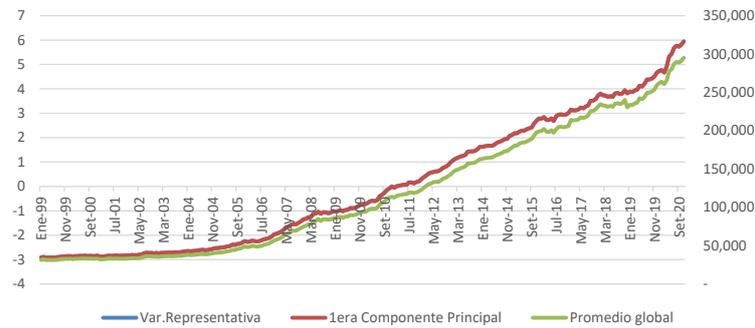


Gráfico 5.33: Comparación grupo 1

Por otro lado, en el cuadro 5.29 se muestra la comparación entre las tres componentes. Como se observa, la correlación es muy alta, cercana a 1 indicando que los tres grupos son idénticos.

Cuadro 5.29: Correlación entre variables representativas, componentes principales y variable global

Correlación	Var.Repr.	1era Comp.Princ.	Prom.Global
Var.Repr.	1		
1era Comp.Princ.	1	1	
Prom.Global	0.9997	0.9997	1

Debido a que el promedio global de las variables que componen el primer grupo está altamente correlacionada con la primera componente principal y la variable representativa podemos considerar el promedio global como una serie de tiempo que representa los periodos financieros del primer grupo. Esta serie presenta una tendencia positiva porque está compuesta por series con tendencia positiva. Los cambios de las series están representados por el promedio global. Como se observa en el gráfico 5.34, la serie presenta cambios a inicios del año 2020 producto de la crisis sanitaria ocasionada por el Covid-19. Conforme transcurrieron los meses, la serie recuperó su tendencia creciente.

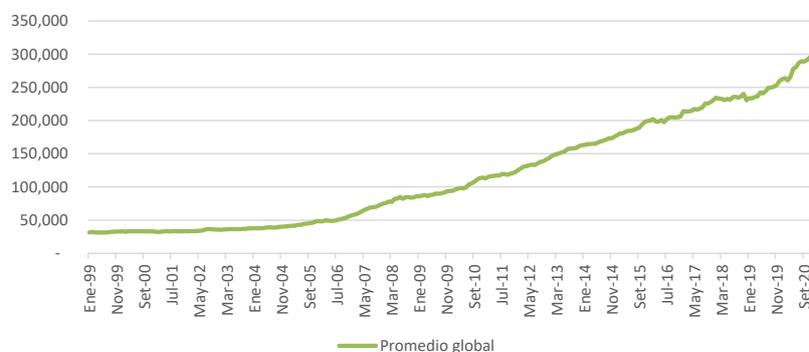


Gráfico 5.34: Promedio Global grupo 1

### 5.4.3. Análisis evolutivo del segundo grupo

Los cortes del análisis evolutivo se muestran en el gráfico 5.35:

n	Within ssq	Indice Calinski-Harabasz	Puntos finales												
			1	2	3	4	5	6	7	8	9	10	11	12	
1	88.72	0.000	0	0	0	0	0	0	0	0	0	0	0	0	0
2	52.84	72.675	120	216	0	0	0	0	0	0	0	0	0	0	0
3	42.46	77.364	22	121	216	0	0	0	0	0	0	0	0	0	0
4	29.55	106.123	23	88	119	216	0	0	0	0	0	0	0	0	0
5	23.47	117.324	22	74	85	119	216	0	0	0	0	0	0	0	0
6	20.31	117.876	23	54	74	85	119	216	0	0	0	0	0	0	0
7	18.13	116.216	23	54	74	85	119	195	216	0	0	0	0	0	0
8	14.05	138.241	22	74	85	120	145	171	192	216	0	0	0	0	0
9	10.89	164.433	23	54	74	85	120	145	171	192	216	0	0	0	0
10	9.57	170.290	23	54	74	83	88	120	145	171	192	216	0	0	0
11	8.23	182.312	23	54	74	83	88	117	122	145	171	192	216	0	0
12	7.31	189.304	22	37	53	74	83	88	117	122	145	171	192	216	216

Gráfico 5.35: Algoritmo de Fisher - Grupo 2

En base a los resultados de 5.35, el valor máximo del indicador se alcanza con 6 grupos los cuales se representan en el siguiente gráfico 5.36 con los tres primeros valores propios del análisis evolutivo.

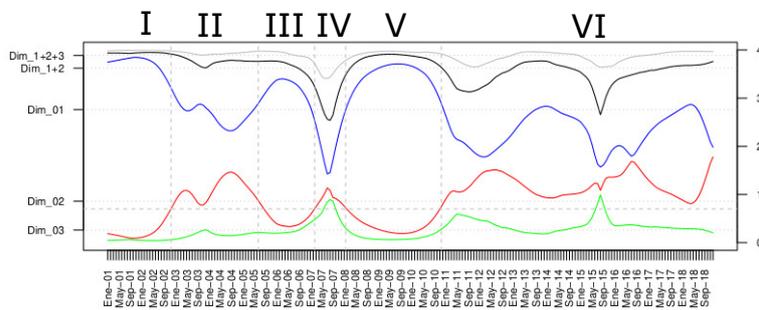


Gráfico 5.36: Evolutivo de los componentes principales - Grupo 2

En el gráfico 5.36 se observa que la tercera componente principal está casi siempre por debajo de 0.7, la segunda tiene valores relevantes en los periodos II, IV y VI, siendo los periodos I III y V los más homogéneos.

En el grafico siguiente se encuentran las series de correlaciones con las primeras componentes.

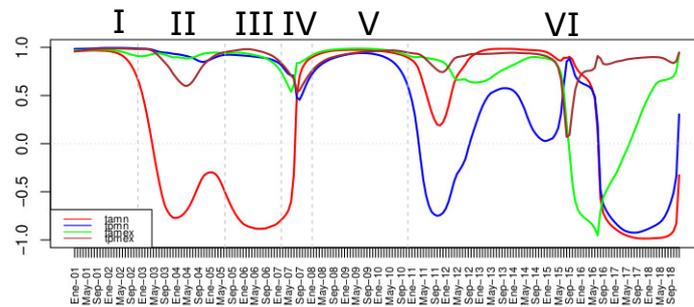


Gráfico 5.37: Correlación 1era componente principal - Grupo 2

En el grafico 5.37 se observa una estabilidad del sentido de las componentes, excepto que en los periodos II y III donde la variable *tamn* se pone en oposición con las demás, mientras que *tpmn* se pone en oposición con las demás en algunas fechas del periodo VI. Finalmente, en la segunda parte del periodo VI todas las series se van mezclando.

En el siguiente gráfico se encuentran las correlaciones de las series con las segundas componentes principales.

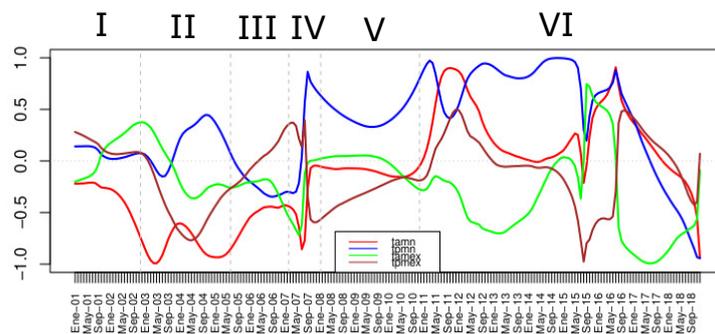


Gráfico 5.38: Correlación 2da componente principal - Grupo 2

En el gráfico 5.38 se observa una gran inestabilidad y en particular dos instantes del periodo IV y VI, con variaciones muy fuertes que sugieren una observación cuidadosa de los datos.

Los gráficos 5.37 y 5.38 indican que las series han presentado mayor fluctuación durante los periodos 2002-2006 y 2011-2020. En el primer periodo mencionado, hubo una mayor entrada de entidades financieras lo que generó una competencia de tasas para captar clientes. En cuanto al último periodo, la volatilidad de las series es producto de la entrada de capitales extranjeros lo que ocasionó mayor oferta de crédito en la economía reduciendo las tasas de interés. Al igual que Krznar y Matheson (2017), las tasas de interés son series importantes para medir el ciclo financiero de una economía. De acuerdo a estos autores los ciclos generados por las tasas de interés son considerados como ciclos de mediano plazo. Los resultados encontrados también están en línea con Ma y Zhang (2016) los cuales consideran que las tasas de interés juegan un papel clave en la economía y en el proceso del ciclo financiero. Las tasas de interés de mercado son el resultado de la interacción de la oferta y la demanda de crédito. Representan tanto el precio de la liquidez para las empresas como las preferencias por el consumo presente frente al futuro por parte de los consumidores y ahorradores.

En el gráfico 5.39 se muestra la comparación entre la primera componente principal, primera variable representativa, y promedio global. La primera componente y la variable representativa se superponen.



Gráfico 5.39: Comparación grupo 2

Por otro lado, en el cuadro 5.30 se muestra la comparación entre las tres componentes. Como se observa, la correlación es muy alta, cercana a 1 indicando que los tres grupos son idénticos.

Cuadro 5.30: Correlación entre variables representativas, componentes principales y variable global

Correlación	Var.Repr.	1era Comp.Princ.	Prom.Global
Var.Repr.	1		
1era Comp.Princ.	1	1	
Prom.Global	0.995	0.995	1

Debido a que el promedio global de las variables que componen el segundo grupo está altamente correlacionada con la primera componente principal y la variable representativa podemos considerar el promedio global como una serie de tiempo que representa los periodos financieros del segundo grupo. Esta serie presenta una tendencia negativa debido a la reducción de las tasas de interés en el mercado financiero peruano. El promedio global se muestra en el gráfico 5.40.



Gráfico 5.40: Promedio Global grupo 2

#### 5.4.4. Análisis evolutivo del tercer grupo

Los cortes del análisis evolutivo se muestran en el gráfico 5.41:

n	Within ssq	Indice Calinski-Harabasz	Puntos finales											
			1	2	3	4	5	6	7	8	9	10	11	12
1	51.22	0.000	0	0	0	0	0	0	0	0	0	0	0	0
2	35.72	46.417	20	216	0	0	0	0	0	0	0	0	0	0
3	27.22	62.591	27	68	216	0	0	0	0	0	0	0	0	0
4	19.17	88.620	27	70	101	216	0	0	0	0	0	0	0	0
5	13.46	118.397	27	70	101	208	216	0	0	0	0	0	0	0
6	10.19	140.893	27	70	100	155	208	216	0	0	0	0	0	0
7	6.96	189.757	15	31	70	100	155	208	216	0	0	0	0	0
8	4.62	262.238	15	31	70	101	124	154	208	216	0	0	0	0
9	3.90	279.256	13	24	33	70	101	124	154	208	216	0	0	0
10	3.22	306.997	13	25	34	66	71	101	124	154	208	216	0	0
11	2.83	319.179	13	25	34	66	71	101	124	154	205	209	216	0
12	2.51	329.544	12	20	29	36	66	71	101	124	154	205	209	216

Gráfico 5.41: Algoritmo de Fisher - Grupo 3

En este análisis no se ha obtenido un máximo local, por lo que se eligieron 7 clases que parecen bastante diferenciados.

En el siguiente gráfico 5.42 se encuentran las series de los primeros tres valores propios del tercer grupo. Se observa una fuerte homogeneidad, excepto en los periodos I, III y VII, con los terceros valores propios casi siempre por debajo de 0.7.

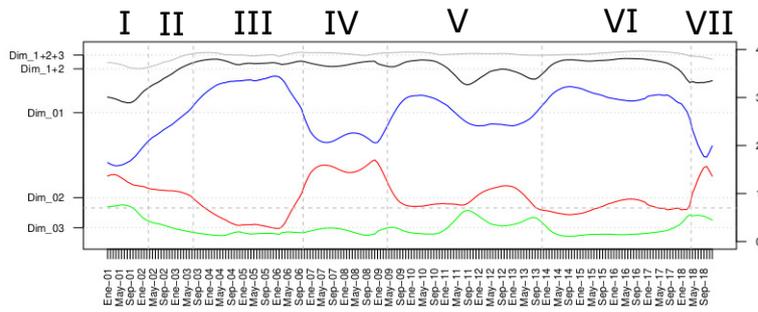


Gráfico 5.42: Evolutivo de los componentes principales - Grupo 3

En el gráfico 5.43 se observa una oposición regular de *encj* y *igbvl* con *dol* y *embig* con un cambio de signo entre los períodos I, II, III, V y VI, mientras que en los períodos IV y VII hay un cruzamiento de todas las series. Se puede decir que este grupo es bastante homogéneo, excluyendo estos dos períodos IV y VII.

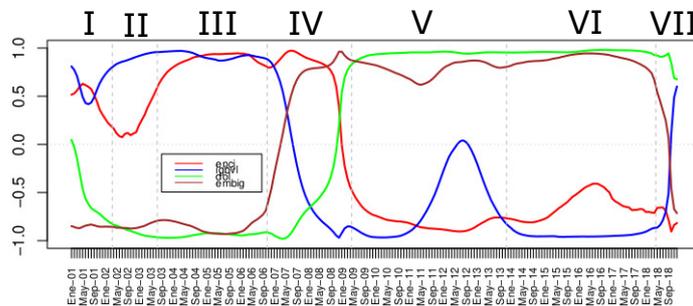


Gráfico 5.43: Correlación 1era componente principal - Grupo 3

En el siguiente gráfico 5.44 se observa el patrón de las series de correlaciones con las segundas componentes principales:

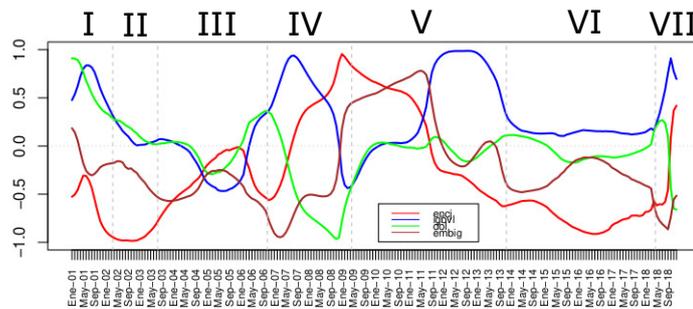


Gráfico 5.44: Correlación 2da componente principal - Grupo 3

Este gráfico 5.44 también resulta difícil la interpretación, ya que en cada período se encuentran estructuras muy diferentes, así como también se observa cambios de signos.

Los gráficos 5.43 y 5.44 muestran que las series que componen el tercer grupo presentan alta volatilidad. Esto tiene sentido debido a que las series como el EMBIG Perú (*embig*) y el Índice General de la Bolsa de Valores (*igbvl*) se determinan en los mercados financieros. Al igual que Shen et al. (2018) se determinó que el índice bursátil *igbvl* es relevante para explicar fluctuaciones del ciclo financiero. Esto va en línea con lo expuesto por Claessens et al. (2011) quienes consideran que para medir el ciclo financiero de una economía es importante tener en

cuenta el precio de las acciones. A diferencia Ramos (2020), el precio de las acciones forma parte del ciclo financiero de corto plazo. Ramos (2020) estimó un ciclo de mediano plazo considerando entre sus variables financieras *igbv* y *tc*.

En el gráfico 5.45 se muestra la comparación entre la primera componente principal, primera variable representativa, y promedio global. Como se indicó en la sección de estudio de clases, la tendencia de la variable representativa de este grupo es decreciente. Así mismo, la primera componente principal y el promedio global de las series también presentan una relación decreciente con este componente.



Gráfico 5.45: Comparación grupo 3

Por otro lado, en el cuadro 5.31 se muestra la comparación entre las tres componentes.

Cuadro 5.31: Correlación entre variables representativas, componentes principales y variable global

Correlación	Var.Repr.	1era Comp.Princ.	Prom.Global
Var.Repr.	1		
1era Comp.Princ.	1	1	
Prom.Global	0.9757	0.9764	1

Debido a que el promedio global de las variables que componen el tercer grupo está altamente correlacionada con la primera componente principal y la variable representativa podemos considerar el promedio global como una serie de tiempo que representa los periodos financieros del tercer grupo. Esta serie presenta una volatilidad debido a que está compuesto por variables que se determinan en el mercado financiero peruano. El promedio global se muestra en el gráfico 5.46.



Gráfico 5.46: Promedio Global grupo 3

Los análisis de los tres grupos muestran una cierta homogeneidad a lo largo del tiempo que justifica su agregación. Por otro lado, de los análisis mostrados podemos sintetizar que se observó diferentes patrones evidentes en el análisis de las 5 variables representativas de las clases.

Finalmente, en el análisis de los tres grupos se observó diferentes grados de homogeneidad. Esto nos lleva a averiguar y comparar las series en los períodos mayormente diferentes, para interpretar mejor los resultados proporcionados. En particular, una inspección en los períodos de mínima homogeneidad parece ser necesaria.

## Capítulo 6

### Conclusiones

En base a los resultados presentados en el capítulo 5, podemos concluir que las variables representativas de cada uno de los tres grupo representan adecuadamente el conjunto de las series que las conforman.

Acerca de la primera variable representativa del primer grupo podemos concluir que dicha componente identifica el componente de larga duración del ciclo financiero. Este componente mide la evolución del sistema financiero a lo largo del tiempo a través de los cambios en los patronos de ahorro, acumulación de reservas internacionales por parte del Banco Central para enfrentar una posible crisis de liquidez de externa, el valor de los fondos de pensiones, el nivel de liquidez del sistema financiero, la cantidad de dinero que circula en la economía, el volumen de créditos y los activos de las empresas bancarias. Todos estos indicadores deben monitorearse pues muestran alertas del sistema financiero que pueden generar efectos adversos en la economía en el largo plazo. Sobre este punto debe tener en cuenta que hay investigaciones que señalan que una caída del sector financiero afecta al sector real elevando los niveles de desempleo, aumento de inflación y reducción de la producción (Ma y Zhang, 2016).

Con respecto a la primera variable representativa del segundo grupo podemos concluir que dicho componente mide la evolución de las tasas activas y pasivas en el sistema financiero. El comportamiento de las tasas no suele fluctuar en el corto plazo, es decir los bancos no modifican sus tasas cada 2 o 3 meses pues generaría cambios en la conducta de sus clientes que podría llevarle a pérdidas crediticias. Por tal motivo, consideramos que la variable representativa del segundo grupo es un componente de mediano plazo el cual mide el comportamiento de los bancos para captar clientes a través de la modificación de sus tasas activas y/o pasivas. El cambio de las tasas de interés activas o pasivas podría indicar que los bancos se encuentran en una guerra de precios para captar mayores clientes. Esto podría beneficiar a los clientes en un corto plazo pero podría generar un endeudamiento en el mediano plazo y consecuentemente llevar a una crisis financiera como se observó en el 2007-2008 en Estados Unidos. Este componente puede ser regulado a través de la fijación de tasas máximas otorgadas por el Banco Central.

Finalmente, sobre la primera variable representativa del tercer grupo podemos concluir que es un componente de corto plazo del ciclo financiero. Este componente podemos asociarlo con la volatilidad que tiene el mercado financiero peruano, es decir con el riesgo de las empresas locales o el riesgo país que ven los inversionistas para invertir en el país.

Concluyendo, se puede decir que las variables representativas de los tres grupos repre-

sentan aspectos diferentes de la economía peruana. A pesar de estar correlacionados (revisar cuadro 5.28), se diferencian por la duración: corto, mediano y largo plazo. Sin embargo, las otras dos variables independientes (*plca* y *tc*) están muy poco correlacionados. Esta conclusión demuestra que la técnica de clasificación jerárquica (HFC) tiene ventajas sobre el análisis de componentes principales (PCA), ya que proporciona variables efectivamente representativas de grupos, mientras la PCA puede proporcionar factores cuya interpretación puede ser de difícil interpretación, ya que puede sintetizar solo parte de las variables.

Por otro lado, el análisis evolutivo nos permitió encontrar patrones diferentes en cada uno de los tres grupos. En el caso del primer grupo el cual está compuesto por series temporales de tendencia creciente se encontró que las variables financieras de las entidades bancarias son relevantes para explicar las fluctuaciones del ciclo financiero, principalmente los cambios de largo plazo. En el caso del segundo grupo, se corroboró que las tasas de interés son variables importantes para medir el ciclo financiero de una economía. De acuerdo a la literatura los ciclos financieros generados por las tasas de interés son considerados como ciclos de mediano plazo. En el tercer grupo se encontró que las variables como el precio de las acciones son importantes para explicar fluctuaciones del ciclo financiero en el corto plazo. Esto está en línea con lo expuesto por Ramos (2020) quién encuentra que el ciclo financiero peruano tiene dos componentes, uno de mediano plazo y otro de corto plazo.

+

## Referencias

- Adarov, A. (2018). *Financial Cycles Around the World* (Vienna Institute for International Economic Studies No. 145). Recuperado del sitio web del Instituto de Viena para Estudios Económicos Internacionales: [//wiiw.ac.at/financial-cycles-around-the-world-p-4459.html](http://wiiw.ac.at/financial-cycles-around-the-world-p-4459.html).
- Barton, D. y David, F. (1956). Some notes on ordered random intervals. *Journal of the Royal Statistical Society. Series B (Methodological)*, 8(1)\*/ 79-94.
- Billio, M. y Petronevich, A. (2017). *Dynamical Interaction between Financial and Business Cycles* (University Ca' Foscari of Venice, Dept. of Economics Research Paper Series No. 24). Recuperado del sitio web del SSRN: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3054438](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3054438).
- Borio, C. (2012). *The financial cycle and macroeconomics: What have we learnt?* (BIS Working Paper No. 395). Recuperado del sitio web de Bank for International Settlements: <https://www.bis.org/publ/work395.htm>
- Cagliarini, A., y Price, F. (2017). Exploring the Link between the Macroeconomic and Financial Cycles. En J. Hambur y J. Simon (Eds.), *Monetary policy and financial stability in a world of low interest rates* (pp. 7-50). Recuperado de <https://www.rba.gov.au/>
- Camiz, S. (2021). Exploratory Classification of Time-Series. In: B. Adıgüzel Mercangöz (Ed.) *Handbook of Research on Emerging Theories, Models, and Applications of Financial Econometrics*. Cham (Switzerland), Springer Nature: 1-29. DOI: 978-3-030-54108-8.
- Camiz, S. y Pillar, V.D. (2018). Identifying the Informational/Signal Dimension in Principal Component Analysis. *Mathematics*, 6(11) 269: 1-16
- Choy, M. y Chang, G. (2014). Medidas macroprudenciales aplicadas en el Perú. *Revista Estudios Económicos BCRP*, 27, 25-50.
- Claessens, S., Kose, A., y Terrones, M. (2011). Financial Cycles: What? How? When?, *NBER International Seminar on Macroeconomics*, 7(1), 303-344.
- De la Cuba, M. (2009). La crisis financiera internacional y los canales de transmisión. *Revista Moneda*, 139, 4-9.
- Denimal, J. (2007). *Classification hiérarchique optimisée d'un tableau de mesures*. *Revue de Statistique Appliquée* 148 (2): 29-61.

- Drehmann, M., Borio, C., y Tsatsaronis, K. (2012). *Characterising the Financial Cycle: Don't Lose Sight of the Medium Term!* (BIS Working Paper No. 380). Recuperado del sitio web de Bank for International Settlements: <https://www.bis.org/publ/work380.htm>
- El-Baz, O. (2018). The Synchronization of Financial and Business Cycles in Saudi Arabia. *Scholedge International Journal of Management Development*, 5(4), 32-47.
- Frontier, S. (1976). Étude de la décroissance des valeurs propres dans une analyse en composantes principales: Comparaison avec le modèle du bton brisé. *Journal of Experimental Marine Biology and Ecology*, 25(1):67-75.
- Golub, G. H. and Van Loan, C. F. (2013). *Matrix computations*. John Hopkins University Press, Baltimore (MA).
- Harding, D., y Pagan, A. (2002). Dissecting the cycle: a methodological investigation. *Journal of Monetary Economics*, 49(2), 365-381.
- Holló, D., Kremer, M., y Lo Duca, M. (2012). *CISS - A composite indicator of systemic stress in the financial system* (ECB Working Paper No. 1426). Recuperado del sitio web de European Central Bank: <https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1426.pdf?ffc22accac5952054016de83c244ee3a>
- Jolliffe, I. (2002). *Principal Component Analysis*, 2nd ed.; Springer: Berlin/Heidelberg, Germany.
- Karfakis, C., y Karfaki, E. (2018). Is the financial cycle a leading indicator of real output during expansions and contractions? A quantile analysis for Greece. *The Journal of Economic Asymmetries*, 18, 1-10.
- Krznar, I. y Matheson, T. (2017). Financial and Business Cycles in Brazil. *IMF Working Paper*, 17, 1-29.
- Lahura, E., Chang, G. y Salazar, O (2013). *Identificación de Episodios de Auge Crediticio: Una propuesta Metodológica con Fundamentos Económicos* (Banco Central de Reservas del Perú Working Paper No. 11). Recuperado del sitio web del SSRN: <https://ideas.repec.org/p/rbp/wpaper/2013-011.html>
- Ma, Y., y Zhang, J. (2016). Financial Cycle, Business Cycle and Monetary Policy. *International Journal of Finance Economics*, 21, 502-527. doi: 10.1002/ijfe.1566
- Menden, C. y Proaño, C. (2017). *Dissecting the financial cycle with dynamic factor models*. (Instituto de Política Macroeconómica IMK Working Papers No. 183). Recuperado del sitio web del Instituto de Política Macroeconómica: [https://www.boeckler.de/imk\\_5279.htm?produkt=HBS-006651&chunk=1&jahr=2017](https://www.boeckler.de/imk_5279.htm?produkt=HBS-006651&chunk=1&jahr=2017)
- Pérez, F. y Vilchez, D. (2018). Ciclos financieros en el Perú. *Revista Moneda*, 176, 4-9.

- Plašil, M., Seidler, J., y Hlaváč, P. (2016). A New Measure of the Financial Cycle: Application to the Czech Republic. *Eastern European Economics*, 54(4), 296-318.
- Ramírez, J. y Vásquez, J. (2014). Circulante y PBI en el Perú: ¿relación de causalidad? (enero 2000 - diciembre 2013). *Revista Moneda del Banco Central de Reservas del Perú*, 158, 11-15.
- Ramos, R. (2020). *Relación de causalidad entre el ciclo financiero y el ciclo económico del Perú entre el periodo 2000-2019* (tesis de maestría, Pontificia Universidad Católica del Perú, Lima, Perú). <https://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/16034>
- Rencher, A. C. (2002). *Methods of multivariate analysis*. Wiley Interscience, New York.
- Rojas, Y. (2010). Objetivos macroprudenciales e instrumentos de política monetaria. *Revista Moneda*, 146, 4-7.
- Rünstler, G. y Vlekke, M. (2016), Business, housing and credit cycles. *Journal of Applied Econometrics*, forthcoming.
- Schüller, Y., Hiebert, P., y Peltonen, T. (2015). *Characterising the financial cycle: a multivariate and time-varying approach* (ECB Working Paper No. 1846). Recuperado del sitio web de European Central Bank: <https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1846.en.pdf?e355531829891515a8371faa50e8a271>
- Shen, C., Shi, J. y Wu, M. (2018). Creating financial cycles in China and interaction with business cycles on the chinese economy. *Emerging Markets Finance Trade*, 54, 2897-2908.
- Smets, F. (2013). Financial stability and monetary policy: how closely interlinked?, *Sveriges Riksbank Economic Review*, 3, 121-160.
- Stremmel H. (2015). Capturing the financial cycle in Europe. *Working Paper European Central Bank*, 1811, 1-26.
- Strohsal, T., Proaño, C., y Wolters, J. (2017). *Characterizing the Financial Cycle: Evidence from a Frequency Domain Analysis*. (Instituto de Política Macroeconómica IMK Working Papers No. 189). Recuperado del sitio web del Instituto de Política Macroeconómica: [https://www.boeckler.de/imk\\_5279.htm?produkt=HBS-006768&chunk=1&jahr=2017#](https://www.boeckler.de/imk_5279.htm?produkt=HBS-006768&chunk=1&jahr=2017#)