



**Indicador sintético coincidente de la actividad  
económica de la Región del Libertador General  
Bernardo O'Higgins, Chile**

**Tesis para optar al grado de “Maestría en Economía Aplicada” de la  
Universidad Austral, Argentina**

***Autor: Carlos Isla Riffo.***

***Director: Facundo Sigal, Estadístico. MS en Estadística aplicada.***

## Contenido

Resumen .....	2
Antecedentes.....	2
Modelización.....	4
Datos .....	11
Pre procesamiento de la data .....	15
Estandarización.....	15
Análisis de Componentes Principales (PCA).....	16
Ajuste estacional .....	19
Resultados.....	24
Evaluación de Residuos .....	25
Comparación PIB regional.....	30
Reflexiones finales.....	31
Bibliografía.....	33

## **Resumen**

Este estudio propone el desarrollo de un indicador mensual de actividad económica para la Región del Libertador General Bernardo O'Higgins en Chile, aplicando una metodología basada en la técnica de Stock y Watson (1989). Esta técnica, que utiliza el Filtro de Kalman, ha sido ampliamente adoptada en la economía aplicada para la estimación de componentes no observables de una serie de variables macroeconómicas

A pesar de la importancia de tener una medida precisa de la actividad económica a nivel regional, actualmente Chile carece de un indicador de este tipo, dado que el Índice Mensual de Actividad Económica (IMACEC) regional ha sido descontinuado. La implementación de este indicador llenaría un vacío crítico y proporcionaría una herramienta valiosa para el análisis y la toma de decisiones en política económica local.

La técnica de Stock y Watson, complementada con el uso del Filtro de Kalman, es una metodología robusta y confiable que ha sido adoptada por instituciones internacionales de renombre, incluyendo la Reserva Federal de los Estados Unidos, el Banco Central Europeo y varios bancos centrales en países emergentes. Su aplicabilidad a un contexto regional en Chile permitirá un seguimiento más preciso y eficiente de la actividad económica, beneficiando a los formuladores de políticas, a los economistas y a los investigadores en su análisis de la economía regional.

## **Antecedentes**

La necesidad de disponer de indicadores económicos a nivel regional es un problema bien reconocido en economía, sobre todo cuando la formulación de políticas necesita responder a desequilibrios y dinámicas regionales (Feldkircher, 2015). La carencia de un indicador mensual de actividad económica a nivel regional en Chile, plantea la necesidad de desarrollar un nuevo instrumento que pueda proporcionar tales indicadores de manera oportuna y precisa a nivel local. El Índice de Actividad Económica Regional (INACER)

calculado por el Banco Central de cumplió esta labor hasta el 2018, año en que fue descontinuado.

Durante su existencia, el Índice de Actividad Económica Regional (INACER) del Banco Central jugó un papel esencial en la formulación de políticas y en la toma de decisiones relacionadas con el desarrollo económico regional. Su discontinuación en 2018 dejó un vacío en términos de información oportuna y precisa que podría utilizarse para comprender y abordar los desafíos específicos de las regiones.

Técnicas como el análisis de series temporales, la regresión y otras formas de modelización estadística han proporcionado durante mucho tiempo a los economistas herramientas valiosas para entender las dinámicas del mercado. Sin embargo, la modelización de factores, y en particular el uso del Filtro de Kalman, se ha popularizado en bancos centrales (Federal Reserve Board (FRB) de Estados Unidos 2020, Banco de Inglaterra 2003) y otras instituciones económicas en todo el mundo para realizar análisis y proyecciones económicas. Este enfoque aprovecha la información contenida en grandes conjuntos de series temporales para resumir las dinámicas económicas subyacentes a través de un pequeño número de factores (Stock & Watson, 2002; Bai & Ng, 2003). Estos factores, que son variables latentes no observables, se extraen para capturar la variabilidad común entre las series de tiempo observables (Bai & Ng, 2002).

Una parte integral de este enfoque es el Filtro de Kalman. Esta técnica de estimación recursiva permite el desglose de series temporales en sus componentes latentes y observables, especialmente en contextos donde el ruido y las perturbaciones son considerables (Kalman, 1960). Es especialmente relevante cuando se trata de actualizar indicadores económicos, ya que el Filtro de Kalman está diseñado para incorporar nuevas observaciones a medida que se disponen, garantizando que los análisis y pronósticos estén alineados con la información más reciente.

Si bien es cierto que métodos como el Análisis de Componentes Principales (ACP) y el de máxima verosimilitud han sido herramientas comunes para la estimación en modelos de factores dinámicos, en este estudio optamos por el Filtro de Kalman. La razón principal de

esta elección se centra en la versatilidad y adaptabilidad del Filtro de Kalman para integrar y actualizar información en tiempo real, especialmente cuando se trata de indicadores económicos. Esta técnica no solo es robusta en presencia de ruido, sino que ofrece una representación más precisa y oportuna de las dinámicas económicas latentes, haciéndola idónea para nuestro propósito.

La implementación de esta metodología requiere una cuidadosa selección de las series de tiempo a incluir en el análisis. Idealmente, estas series deberían ser representativas de la economía regional, estar disponibles a tiempo y ser actualizables regularmente. Además, deberían ser susceptibles de ser influenciadas por los mismos factores económicos subyacentes. Mientras que los indicadores nacionales ofrecen una perspectiva general de la economía de un país, los indicadores regionales, como el que buscamos desarrollar para la Región de O'Higgins, proporcionan una visión más matizada y detallada. Estos indicadores pueden desvelar desequilibrios y oportunidades específicas, lo que es fundamental para políticas y estrategias dirigidas.

Un caso destacable de la aplicación de esta metodología es el Indicador Sintético de Actividad Económica (ISAE) desarrollado por el Gobierno de la Provincia de Santa Fe en Argentina. Este indicador ha demostrado ser efectivo y útil en el análisis y la predicción de la actividad económica a nivel regional (Gobierno de Santa Fe, 2014). Un estudio reciente y extendido del tema puede encontrarse en Sigal, Camusso y Navarro (2022).

En este estudio, nos inspiramos en el ISAE y en la literatura de modelos de factores dinámicos para desarrollar un indicador económico para la Región de O'Higgins. La meta es proporcionar a los encargados de formular políticas una herramienta útil y precisa que les permita evaluar la salud económica de la región y tomar decisiones informadas.

## **Modelización**

La técnica del análisis de factores dinámicos se centra en la idea de que un vasto conjunto de series temporales puede ser sintetizado y representado por un número limitado de

factores comunes. Estos factores, intrínsecamente asociados con movimientos sistemáticos y amplios de la economía, se convierten en elementos centrales en el estudio y modelación de fenómenos macroeconómicos. La identificación y extracción de factores latentes se realiza sobre la base de la varianza compartida entre las series temporales. Este proceso busca encontrar un conjunto mínimo de factores no observables que puedan explicar la correlación existente entre las series. Intuitivamente, un factor latente representa patrones comunes que emergen al observar múltiples series a lo largo del tiempo. El principio de máxima varianza, por ejemplo, implica seleccionar factores que capturen la mayor cantidad de variabilidad conjunta de las series temporales, garantizando así que estos factores encapsulen la información esencial de las dinámicas macroeconómicas en cuestión.

Formalmente, un modelo de factores dinámicos puede ser representado por la ecuación:

$$Y_{it} = \lambda_i F_t + \varepsilon_{it} \quad (1)$$

Donde

- $i = 1, \dots, N$  representa cada serie y  $t = 1, \dots, T$  el tiempo.
- $Y_{it}$  es la  $i$ -ésima serie temporal en el tiempo  $t$ .
- $F_t$  es un vector de  $r$  factores comunes.
- $\lambda_i$  es un vector de cargas factoriales que mide la sensibilidad de la serie  $i$ -ésima a los factores comunes.
- $\varepsilon_{it}$  es un término de error idiosincrásico que representa la variabilidad específica de la serie que no es capturada por los factores comunes.

Las cargas factoriales  $\lambda_i$  desempeñan un papel crucial en la relación entre factores y series temporales. Cada  $\lambda_i$  mide la sensibilidad o respuesta de una serie temporal particular a los cambios en un factor latente. En otras palabras, nos indica cuánto varía una serie temporal en particular cuando un factor latente cambia en una unidad. Por lo tanto, un valor alto de  $\lambda_i$  sugiere que la serie está fuertemente influenciada por el factor, mientras que un valor cercano a cero implica que la serie es en gran medida independiente de ese factor.

La Ecuación (1) que define la relación entre las series temporales y los factores latentes puede ser expresada como  $Y_t = \Lambda F_t + \varepsilon_t$ , donde  $Y_t$  es el vector de observaciones,  $\Lambda$  es la matriz de factores de carga,  $F_t$  es el vector de factores latentes y  $\varepsilon_t$  es el término de error.

Una característica distintiva de los modelos de factores dinámicos es la asunción de que los factores comunes  $F_t$  evolucionan siguiendo una dinámica de vectores auto regresivos. En la formulación de los modelos de factores dinámicos, se asume que los factores comunes  $F_t$  siguen una dinámica de vectores autorregresivos. Matemáticamente, esta evolución puede ser descrita como:

$$F_t = \phi_1 F_{t-1} + \dots + \phi_p F_{t-p} + \eta_t \quad (2)$$

En esta formulación:

- $\phi_j, j = 1, \dots, p$ , son matrices de coeficientes autorregresivos de orden  $p$  (AR( $p$ )).
- $\eta_t$  es un vector de perturbaciones que se asume que es ruido blanco.

Los términos de error idiosincrásicos  $\varepsilon_{it}$  también se asumen como ruido blanco.

La representación en vectores auto regresivos de los factores comunes,  $F_t$ , indica una interacción entre estos a lo largo del tiempo, siendo posible capturar la transmisión de shocks entre ellos al considerar la matriz de covarianza de las perturbaciones  $\eta_t$ . Esta matriz, denotada como  $\Sigma_\eta = E[\eta_t \eta_t']$ , refleja cómo estas perturbaciones interactúan entre sí, donde  $E[.]$  denota la expectativa matemática.

La principal tarea en la estimación de modelos de factores dinámicos es obtener los factores comunes  $F_t$  y las cargas factoriales  $\lambda_i$  a partir de los datos observados  $Y_{it}$ . En esencia, el Filtro de Kalman realiza predicciones de los factores comunes tomando como base las observaciones previas y, a medida que se dispone de nueva información, actualiza estas predicciones. El proceso de filtrado se puede expresar mediante las siguientes ecuaciones:

$$F_{t|t-1} = \phi_1 F_{t-1|t-1} + \dots + \phi_p F_{t-p|t-p} \quad (3)$$

$$P_{t|t-1} = \phi_1 P_{t-1|t-1} \phi_1' + \dots + \phi_p F_{t-p|t-p} \phi_p' + Q \quad (4)$$

donde

- $F_{t|t-1}$  es la predicción de los factores en el tiempo  $t$  basada en las observaciones hasta el tiempo  $t - 1$
- $P_{t|t-1}$  es la matriz de covarianza de los errores de tal predicción
- $Q$  es la matriz de covarianza de las perturbaciones  $\eta_t$ .

Una vez que se han obtenido las predicciones de los factores y sus errores de predicción, se pueden actualizar estas predicciones utilizando las observaciones actuales en orden de optimizar la estimación de los factores. Este proceso de actualización se puede expresar mediante las siguientes ecuaciones:

$$F_{t|t} = F_{t|t-1} + P_{t|t-1} \lambda_i' (\lambda_i P_{t|t-1} \lambda_i' + \Sigma_i)^{-1} (Y_{it} - \lambda_i F_{t|t-1}) \quad (5)$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1} \lambda_i' (\lambda_i P_{t|t-1} \lambda_i' + \Sigma_i)^{-1} \lambda_i P_{t|t-1} \quad (6)$$

Donde

- $F_{t|t}$  es la actualización de la predicción de los factores basada en la observación actual
- $P_{t|t}$  es la matriz de covarianza de los errores de predicción actualizados
- $\Sigma_i$  es la matriz de covarianza de los errores idiosincrásicos.

En la estimación de modelos de factores dinámicos, la estructura de covarianza  $\Sigma_i$  juega un papel vital. Específicamente, esta matriz no solo capta la variabilidad idiosincrática de las series, sino que también refleja la volatilidad y las correlaciones potenciales entre las mismas. Si ciertas series tienen errores altamente correlacionados, esto puede señalar dinámicas compartidas que no son capturadas por los factores comunes. Explorar y justificar la estructura de  $\Sigma_i$  asegura que el modelo captura adecuadamente todas las

dinámicas relevantes. Matemáticamente, las entradas diagonales de  $\Sigma_i$  representan las varianzas de los errores idiosincrásicos, mientras que las entradas fuera de la diagonal representan las covarianzas.

Las ecuaciones (3) a (6) son esenciales para entender la lógica operativa del Filtro de Kalman en el contexto de la estimación de modelos de factores dinámicos. Estas ecuaciones iteran a través del tiempo, refinando las estimaciones a medida que se procesan las observaciones.

El Filtro de Kalman, en su esencia, no solo proporciona una forma recursiva y eficiente de estimar los estados latentes de un modelo de espacio de estados, sino que también está estrechamente relacionado con el concepto de Estimación Máxima Verosimilitud (MLE). Durante el proceso de filtrado, se calcula la densidad condicional de las observaciones  $Y_t$  dado todo el conocimiento pasado, denotada  $p(Y_t|Y_{t-1}, \theta)$ , donde  $\theta$  son los parámetros del modelo. La verosimilitud del modelo completo, dada una muestra de datos, es el producto de estas densidades condicionales:

$$L(\theta; Y_1, \dots, Y_T) = \prod_{t=1}^T p(Y_t|Y_{t-1}, \theta) \quad (7)$$

Maximizar esta función con respecto a  $\theta$  nos da los estimadores de máxima verosimilitud de los parámetros del modelo. Por tanto, el proceso de filtrado con el Filtro de Kalman proporciona las herramientas necesarias para llevar a cabo la estimación MLE en modelos de espacio de estados.

Además del filtrado, se aplica una técnica de suavizado basada en la utilización de toda la serie temporal. El suavizado de Kalman, matemáticamente, puede ser representado por:

$$F_{t|T} = F_{t|t} + J_t(F_{t+1|T} - F_{t+1|t}) \quad (8)$$

$$J_t = P_{t|t} \Phi_t' P_{t+1|t}^{-1} \quad (9)$$

Donde,

- $F_{t|T}$  es la estimación suavizada del factor en el tiempo  $t$  dada toda la muestra hasta  $T$
- $J_t$  es la matriz de ganancia de suavizado.

Esta técnica utiliza información tanto anterior como posterior a  $t$  para proporcionar estimaciones óptimas de los factores pasados.

Una etapa esencial para el correcto funcionamiento del Filtro de Kalman es su inicialización. Específicamente, se requiere establecer valores iniciales para  $F_{1|0}$  y  $P_{1|0}$ . Estos pueden ser determinados mediante diferentes criterios, pero un enfoque común es utilizar el método de los momentos. Matemáticamente, esto se puede representar como:

$$F_{1|0} = E[F_1] \quad (10)$$

$$P_{1|0} = Var[F_1] \quad (11)$$

Las condiciones iniciales del Filtro de Kalman, esenciales para su correcta operación, se establecieron utilizando una inicialización difusa<sup>1</sup>, denotada como:

$$F_{1|0} = 0 \quad (12)$$

$$P_{1|0} = a \times I \quad (13)$$

donde  $F_{1|0}$  es la predicción inicial del factor y  $P_{1|0}$  es la matriz de covarianza de los errores de predicción inicial, siendo  $I$  la matriz identidad y  $a$  a un valor grande. Este enfoque minimiza la influencia de las condiciones iniciales sobre las estimaciones, permitiendo que los datos guíen la identificación de las dinámicas subyacentes.

---

<sup>1</sup> Estas condiciones iniciales son consistentes con las configuraciones predeterminadas del paquete statsmodels utilizado para la estimación.

El modelo descrito anteriormente se puede interpretar como un modelo de espacio de estados. Un modelo de espacio de estados es una representación que nos permite entender y analizar procesos estocásticos en términos de variables observables y no observables. En otras palabras, bajo esta óptica, se hace una distinción entre las variables que pueden ser directamente observadas (como series temporales económicas) y aquellas que no pueden ser observadas directamente pero que influyen en las observadas. Estas últimas, en nuestro contexto, corresponden a los factores comunes.

En este enfoque, las ecuaciones de observación y transición son fundamentales para describir la relación entre las variables observables y no observables, así como para comprender cómo evolucionan en el tiempo. La ecuación de observación establece cómo las variables observables se relacionan con los componentes no observables a través de una matriz de carga. Por otro lado, la ecuación de transición describe la evolución de los componentes no observables a lo largo del tiempo, utilizando una matriz de transición y una matriz de ruido.

El uso del filtro de Kalman en este contexto nos permite estimar los componentes no observables y obtener una visión más precisa de la dinámica económica. Mediante la combinación de las ecuaciones de observación y transición, y la aplicación del filtro de Kalman, podemos descomponer las series observadas en sus componentes subyacentes y analizar su comportamiento de manera más detallada.

Dentro del modelo de espacio de estados, podemos identificar dos ecuaciones fundamentales que describen el comportamiento del sistema. La ecuación de observación, previamente mencionada como (1), establece la relación entre las variables observables y no observables. A través de la matriz de carga  $Z_t$ , que relaciona estas variables, y los términos de error  $\varepsilon_t$ , se captura la variabilidad de las observaciones que no es explicada por los factores comunes.

Por otro lado, la ecuación de transición, que corresponde a la ecuación (2), describe la evolución de los componentes no observables a lo largo del tiempo. Mediante la matriz de transición  $T_t$  y la matriz de ruido  $Q_t$ , se establece cómo los factores comunes cambian y se

relacionan entre sí. Los términos de error  $\eta_t$  representan la variabilidad no explicada por los factores comunes en la evolución de los componentes no observables.

Es importante destacar que las ecuaciones (3) y (4), previamente presentadas, son equivalentes a las ecuaciones de predicción y actualización del filtro de Kalman en el contexto de un modelo de espacio de estados. En la ecuación (3), la predicción del estado se realiza a través de la combinación lineal de los estados anteriores, mientras que en la ecuación (4), se actualiza la matriz de covarianza del estado teniendo en cuenta las matrices de transición y las perturbaciones.

En el siguiente apartado, se presentarán detalladamente las variables seleccionadas para nuestro estudio, considerando el modelo propuesto y haciendo mención a la diversidad de fuentes, excepciones de periodicidad y otros aspectos relevantes.

## **Datos**

La selección de variables en este estudio se ha realizado con meticulosa consideración, con el objetivo de captar una representación comprensiva de los fenómenos económicos bajo análisis. Estas variables han sido obtenidas de fuentes institucionales reconocidas, como el Banco Central y el Instituto Nacional de Estadísticas (INE), asegurando la integridad y la relevancia de los datos.

La combinación de estas variables ofrece una visión integral y coherente de la economía de la Región de O'Higgins, reflejando tanto tendencias generales como interacciones complejas. Su selección fue también motivada por la disponibilidad y su disposición mensual, con cierta regularidad y suficiencia de data. En específico las series son las siguientes:

#### 1) Consumo de Hogares e IPSFL

El consumo de hogares e IPSFL refleja la suma global de bienes y servicios adquiridos o consumidos por las familias y por las Instituciones Privadas sin Fines de Lucro al Servicio de los Hogares.

- Periodicidad: Trimestral
- Unidad de medida: Millones de Pesos, Volumen a precios del año anterior encadenado.
- Fuente: Banco Central

#### 2) PIB

El Producto Interno Bruto (PIB) representa el valor total de todos los bienes y servicios producidos en la región durante un período específico.

- Periodicidad: Trimestral
- Unidad de medida: Millones de Pesos, Volumen a precios del año anterior encadenado
- Fuente: Banco Centra

#### 3) Empleo (Desocupación)

La tasa de desocupación es el porcentaje de la fuerza laboral que está desempleada y buscando activamente trabajo.

- Periodicidad: Mensual
- Unidad de medida: Porcentaje (%)
- Fuente: Encuesta nacional de Empleo ENE, INE

#### 4) Número de Constituciones en Registro de Empresas y Sociedades

Cantidad de nuevas entidades legales, como empresas y sociedades, que han sido formalmente registradas en la región en un período determinado.

- Periodicidad: Mensual
- Unidad de medida: Millones de Pesos, Volumen a precios del año anterior encadenado
- Fuente: Servicio de impuestos internos S.I.I.

#### 5) Saldos en Cuentas Corrientes y Vista

Refleja la suma consolidada de fondos en cuentas corrientes y cuentas de ahorro a la vista en bancos comerciales en Chile, sirviendo como indicador de la liquidez en el sistema bancario.

- Periodicidad: Mensual
- Unidad de medida: Millones de Pesos
- Fuente: Comisión para el mercado financiero CMF

#### 6) Porcentaje de Deuda (consumo, vivienda y comercial)

Este indicador proporciona una descomposición de la cartera de préstamos de los bancos comerciales en Chile, en función de las categorías de consumo, vivienda y comercial, expresado como un porcentaje del total de deuda.

- Periodicidad: Mensual
- Unidad de medida: Porcentaje (%)
- Fuente: Comisión para el mercado financiero CMF

#### 7) Índice de Producción Manufacturera (IPMan)

Este índice mide la evolución de la actividad productiva de la industria manufacturera desde el punto de vista de la oferta<sup>2</sup>.

- Periodicidad: Mensual
- Unidad de medida: Índice
- Fuente: Instituto Nacional de Estadísticas

---

<sup>2</sup> Se compone de índices de producción de las siguientes divisiones: Elaboración de productos alimenticios, bebidas alcohólicas y no alcohólicas, productos de tabaco, producción y fabricación de productos de madera y corcho; fabricación de artículos de paja y de materiales trenzables, fabricación de papel y productos de papel, impresión y reproducción de grabaciones, fabricación de coque y productos de la refinación del petróleo, fabricación de sustancias y productos químicos, fabricación de productos farmacéuticos, sustancias químicas medicinales y productos botánicos de uso farmacéutico, fabricación de productos de caucho y de plástico, fabricación de otros productos minerales no metálicos, fabricación de metales comunes, fabricación de productos elaborados de metal excepto maquinaria y equipo, fabricación de equipo eléctrico, fabricación de maquinaria y equipo n.c.p, fabricación de vehículos automotores, remolques y semirremolques, fabricación de otros tipos de equipo de transporte y fabricación de muebles.

## 8) Electricidad

### a) Generación eléctrica

Cantidad de producción real, medida en megavatios hora (MWh), procedente de las distintas centrales generadoras que se reportan al coordinador eléctrico nacional.

- Periodicidad: Mensual
- Unidad de medida: MWh
- Fuente: Coordinador Eléctrico Nacional

## 9) Superficie Autorizada solicitudes de edificación

Refleja la superficie total, en metros cuadrados, correspondiente a las solicitudes de edificación (tanto habitacional como no habitacional) que han sido autorizadas en un mes determinado.

- Periodicidad: Mensual
- Unidad de medida: m<sup>2</sup>
- Fuente: Instituto Nacional de Estadísticas, INE

## 10) Índice de Ventas de Supermercados (ISUP)

A precios constantes, con Base promedio año 2014=100, este índice mide la evolución de la actividad comercial a través de las ventas netas totales de los establecimientos que operen dentro del territorio nacional y cuenten con tres o más cajas instaladas.

- Periodicidad: Mensual
- Unidad de medida: Índice
- Fuente: Instituto Nacional de Estadísticas, INE.

## 11) Tasa de ocupación en habitaciones

Indica el nivel de ocupación de las habitaciones disponibles en establecimientos de alojamiento, reflejando la demanda turística y su variabilidad

- Periodicidad: Mensual.
- Unidad de medida: Porcentaje (%).
- Fuente: Encuesta Mensual de Alojamiento turístico (EMAT), Instituto Nacional de Estadísticas, INE.

## 12) Exportaciones

Valor total de las exportaciones medidas en términos FOB ("Free on Board"), es decir, cuando están embarcadas y listas para su destino, expresado en millones de dólares estadounidenses.

- Periodicidad: Mensual.
- Unidad de medida: FOB en MMUS\$.
- Fuente: Instituto Nacional de Estadísticas, INE – Servicio Nacional de Aduanas SNA.

## **Pre procesamiento de la data**

### **Estandarización**

La preparación adecuada de los datos es un paso crucial para garantizar la calidad y relevancia de cualquier análisis subsiguiente, y aún más cuando se busca sintetizar información macroeconómica mediante indicadores compuestos. Para construir el Indicador Sintético de Actividad Económica en la Región de O'Higgins, adoptamos un enfoque indirecto que prioriza la precisión y la robustez de la metodología.

Este enfoque, antes de sumergirnos en la aplicación de un modelo factorial dinámico y el Filtro de Kalman, exige un procesamiento y ajuste meticuloso de las series de datos. Para ello fue fundamental asegurar la estacionariedad de las series. Para ello, recurrimos a la transformación en diferencias logarítmicas, una práctica estándar en la literatura económica. Esta transformación no solo estabiliza las series volátiles, sino que también las hace cointegradas, un requisito vital para la aplicación de técnicas como la de Stock y Watson (1989).

La estandarización de las series también es crucial, dada su heterogeneidad. Para garantizar que todas las series tengan el mismo peso en el análisis, se les resta su media y se dividen por su desviación estándar. Esta normalización pone de relieve las

fluctuaciones relativas en lugar de las fluctuaciones absolutas, lo que facilita la interpretación económica de los resultados.

## **Análisis de Componentes Principales (PCA)**

Posteriormente, siguiendo el modelo ISAE, optamos por el Análisis de Componentes Principales (PCA). Esta elección busca descomponer y simplificar la estructura del modelo, enfocándonos en aquellas variables con una contribución destacada a la variabilidad total. El PCA revela la estructura subyacente del conjunto de datos a través de la descomposición de la matriz de covarianza  $S$  de  $X$ . Matemáticamente, buscamos los vectores  $a$  a tales que:

$$a = \operatorname{argmax}_a a^T S a \quad \text{sujeto a } \|a\| = 1.$$

Los eigenvectores de  $S$  proporcionan una base ortogonal para el espacio de datos, y los datos transformados  $Z$  se obtienen mediante  $Z = Xa$ . Los eigenvectores de  $S$  proporcionan una base ortogonal para el espacio de datos, y la proyección de los datos originales sobre estos eigenvectores genera los componentes principales.

Las cargas factoriales, a menudo referidas como "pesos", son fundamentales para comprender la relación entre cada variable y el componente principal correspondiente. Matemáticamente, una carga factorial es simplemente el coeficiente de correlación entre la variable original y el componente. Estas cargas son el resultado del producto de los eigenvectores de la matriz de covarianza  $S$  y la raíz cuadrada de sus eigenvalores correspondientes:

$$\text{Carga}_{ij} = a_{ij} \sqrt{\lambda_j}$$

donde  $a_{ij}$  es el elemento de la matriz de eigenvectores asociado a la variable  $i$  y al componente  $j$ , y  $\lambda_j$  es el eigenvalor correspondiente al componente  $j$ . Una carga factorial alta en magnitud indica que la variable tiene una fuerte relación con el componente principal.

Por lo tanto, al examinar las cargas factoriales de las variables para un componente específico, podemos discernir qué variables están más fuertemente asociadas con ese componente y, por lo tanto, tienen mayor relevancia en la descripción de la variabilidad que el componente captura.

Los resultados mostrados en la Tabla 1, indican que los primeros cinco componentes tienen auto valores mayores a 1, acorde al criterio de Kaiser (1960, o en una aplicación más reciente, Bai & Ng 2002). Estos componentes representan una variabilidad acumulada del 67.8%, con los auto valores 1.810272, 1.668156, 1.445335, 1.321023 y 1.199842 explicando, respectivamente, el 16.5%, 15.2%, 13.1%, 12.0% y 10.9% de la variabilidad total.

Tabla 1: Auto valores y varianza PCA

<b>Auto valor</b>	<b>Diferencia</b>	<b>Proporción</b>
1.810272	0.142116	0.164570
1.668156	0.222821	0.151651
1.445335	0.124312	0.131394
1.321023	0.121181	0.120093
1.199842	0.392230	0.109077
0.807612	0.016880	0.073419
0.790732	0.135079	0.071885
0.655652	0.037085	0.059605
0.618568	0.214221	0.056233
0.404346	0.125884	0.036759
0.278462		0.025315

El método de Stock y Watson se basa en la identificación de un único factor que sintetiza la información común contenida en un conjunto de series temporales. En este sentido, y considerando el análisis previo, el primer componente principal emerge como la elección óptima, no solo por su auto valor predominante, sino por su capacidad para encapsular la estructura dominante del conjunto de datos. En la tabla 2 es posible observar que las series "exportaciones", "consumo" y "desocupación" son las que más influyen en PC1, con cargas factoriales de 0.3593, 0.3711 y -0.4089, respectivamente. Esta decisión está respaldada por el diferencial de 0.142116 entre los dos primeros auto valores, reforzando la predominancia de PC1 en la representación de los datos.

Tabla 2: Componentes principales

Variable	PC1	PC2	PC3	PC4	PC5
Exportaciones	0.3593	-0.4619	-0.2023	0.2419	0.1608
Consumo	0.3711	0.1070	0.3845	-0.4282	-0.2458
Desocupación	-0.4089	-0.2820	0.2126	-0.3100	-0.1845
Empresas	0.2214	-0.4351	-0.1544	-0.3302	-0.2888
IPOM	-0.3003	-0.2382	0.0799	-0.4958	0.2565
Generación eléctrica mwh	0.2998	-0.0316	-0.4048	-0.4413	0.3075
ISUP	0.2337	-0.0471	0.2429	0.0366	0.6228
EMAT	0.2431	0.5253	-0.1731	-0.1712	-0.2727
Saldo cuentas	0.0953	-0.1884	0.5622	0.1867	-0.0881
Deuda	-0.3889	-0.0918	-0.3957	0.0850	-0.0913
Superficie autorizada	0.2509	-0.3558	-0.1022	0.1994	-0.3970

Aunque el primer componente es destacado, no podemos subestimar la importancia del segundo componente principal. La relevancia de PC2 es evidente dada su proporción significativa de la varianza total explicada, especialmente al observar su auto valor en la Tabla 1. En el análisis de componentes principales, es esencial no solo considerar el componente dominante sino también los siguientes componentes que, en su conjunto, pueden ofrecer una perspectiva más rica sobre la estructura de datos.

Al analizar las cargas factoriales en la Tabla 2, notamos que las series "exportaciones", con una carga factorial de -0.4619, y "EMAT", con una carga factorial de 0.5253, son altamente influyentes en PC2 así Saldo cuentas en el PC3 con 0.5622. Estas series proporcionan una dimensión adicional de la dinámica macroeconómica, complementando la información encapsulada en PC1.

A pesar de que la metodología de Stock y Watson tradicionalmente se enfoca en un único factor, la inclusión de tres componentes en nuestro caso, subraya la idea de que la actividad económica es multifacética y puede ser mejor representada al considerar múltiples dimensiones. La decisión de considerar tanto PC1 como PC2 busca capturar la mayor variabilidad inherente en el conjunto de datos y garantizar una representación robusta de la dinámica macroeconómica en estudio.

Para finalizar nuestra selección basada en el análisis de componentes principales, y con miras a crear un indicador que refleje de manera fiel la actividad económica de la Región de O'Higgins, hemos identificado un conjunto de series clave. Las series "exportaciones",

"consumo", "desocupación", "EMAT" y " Saldo cuentas" han sido seleccionadas dadas sus cargas factoriales significativas en los tres primeros componentes. La integración de estas series en nuestro análisis final garantiza un abordaje comprehensivo y robusto, permitiendo la construcción de un indicador sintético que abarque ampliamente las dinámicas macroeconómicas de interés.

## **Ajuste estacional**

En el ámbito macroeconómico, las series temporales a menudo presentan patrones estacionales que pueden enmascarar tendencias clave y oscurecer las dinámicas reales de la economía. Considerando que el modelo de Stock y Watson no atiende directamente a esta estacionalidad, es de suma importancia tratarla de manera adecuada. Para abordar este desafío, recurrimos al algoritmo X-13 ARIMA-SEATS. Esta herramienta no solo destaca por su habilidad de desestacionalizar las series, sino que además cuenta con una característica invaluable: la selección automática del modelo ARIMA.

La selección automática del modelo ARIMA es un proceso que busca optimizar los parámetros del modelo en función de la serie en estudio. Específicamente, el algoritmo X-13 ARIMA-SEATS decide automáticamente sobre los órdenes de diferenciación  $d$  y  $D$ , los componentes auto regresivos y de media móvil, tanto estacionales como no estacionales, representados por los órdenes  $p$ ,  $q$ ,  $P$  y  $Q$ . Esta automatización permite adaptarse de manera más precisa a las particularidades de cada serie. La fórmula general del modelo ARIMA que adoptamos, denotada como

$$\Phi\phi(B)\phi(B)(1-B)^d(1-B^S)^DY_t = \Theta(B)\theta(B)a_t,$$

representa una estructura que considera tanto componentes estacionales como no estacionales, ajustándose eficazmente a las dinámicas específicas de nuestras series económicas.

Una vez definidos los modelos para cada serie, la elección de estos se justificó en base a criterios informativos consolidados en la literatura estadística, como el AICc y el BIC. El AICc o Criterio de Información de Akaike Corregido, se define matemáticamente como  $AICc = AIC + \frac{2k(k+1)}{n-k-1}$  donde  $IC = 2k - 2\ln(\hat{L})$ ,  $k$  es el número de parámetros en el modelo,  $n$  es el número de observaciones, y  $\hat{L}$  es el valor máximo de la función de verosimilitud para el modelo.

Por otro lado, el BIC, o Criterio de Información Bayesiano, se expresa como  $BIC = \ln(n)k - 2\ln(\hat{L})$  donde nuevamente  $k$  es el número de parámetros en el modelo,  $n$  es el número de observaciones, y  $\hat{L}$  es el valor máximo de la función de verosimilitud para el modelo.

Una vez que hemos establecido y justificado el modelo ARIMA para cada serie en función de criterios como AICc y BIC, es esencial validar la idoneidad del modelo a través de pruebas que se centren en los residuos. En este sentido, la Prueba Box-Ljung es una herramienta fundamental para verificar la autocorrelación en los residuos del modelo. Matemáticamente, la Prueba Box-Ljung se define como  $Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k}$  donde  $\hat{\rho}_k$  es la autocorrelación muestral de orden  $k$ ,  $n$  es el tamaño de la muestra, y  $h$  es el número de retrasos. La ausencia de auto correlaciones significativas en los residuos es indicativa de un buen ajuste del modelo a los datos.

Además, es imperativo asegurarse de que los residuos del modelo se distribuyan normalmente para validar muchos procedimientos estadísticos. Para esto, el Test Shapiro-Wilk es ampliamente utilizado. Matemáticamente, se define como  $W = \frac{(\sum_{i=1}^n a_i x_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$  donde  $a_i$  son los coeficientes derivados de los valores esperados de los ordenados de una muestra de una distribución normal estándar. Un valor cercano a uno en esta prueba es un indicativo sólido de que los residuos siguen una distribución normal, validando así la confiabilidad de inferencias posteriores basadas en el modelo.

Estas herramientas de validación, junto con la cuidadosa selección del modelo ARIMA, aseguran que las series están adecuadamente preparadas y son representativas, formando la base para construcciones posteriores como la de un indicador al estilo Stock y Watson.

En el proceso de ajustar y modelar nuestras series temporales, aplicamos rigurosos criterios y pruebas para garantizar la idoneidad y robustez de los modelos seleccionados. Los detalles específicos se encuentran plasmados en la Tabla 3. Aquí, se detalla la especificación ARIMA de cada serie, junto con métricas de ajuste y validación.

Al observar la serie "exportaciones", por ejemplo, se nota un patrón ARIMA (1 0 0) (1 1 0). Este modelo, seleccionado basándose en los criterios AICc y BIC con valores de 745.8 y 763.9 respectivamente, indica un componente autor regresivo y una diferenciación estacional. Este enfoque se tomó buscando equilibrar la representatividad del modelo con la complejidad, evitando sobreajustes que puedan comprometer las inferencias posteriores.

Es esencial que los residuos de los modelos no muestren patrones o correlaciones no deseadas. En este sentido, recurrimos a la prueba Box-Ljung, la cual se aplicó para verificar la ausencia de autocorrelación en los residuos. Un caso destacado es la serie "EMAT", cuyo valor de 22.1 sugiere que los residuos están adecuadamente distribuidos.

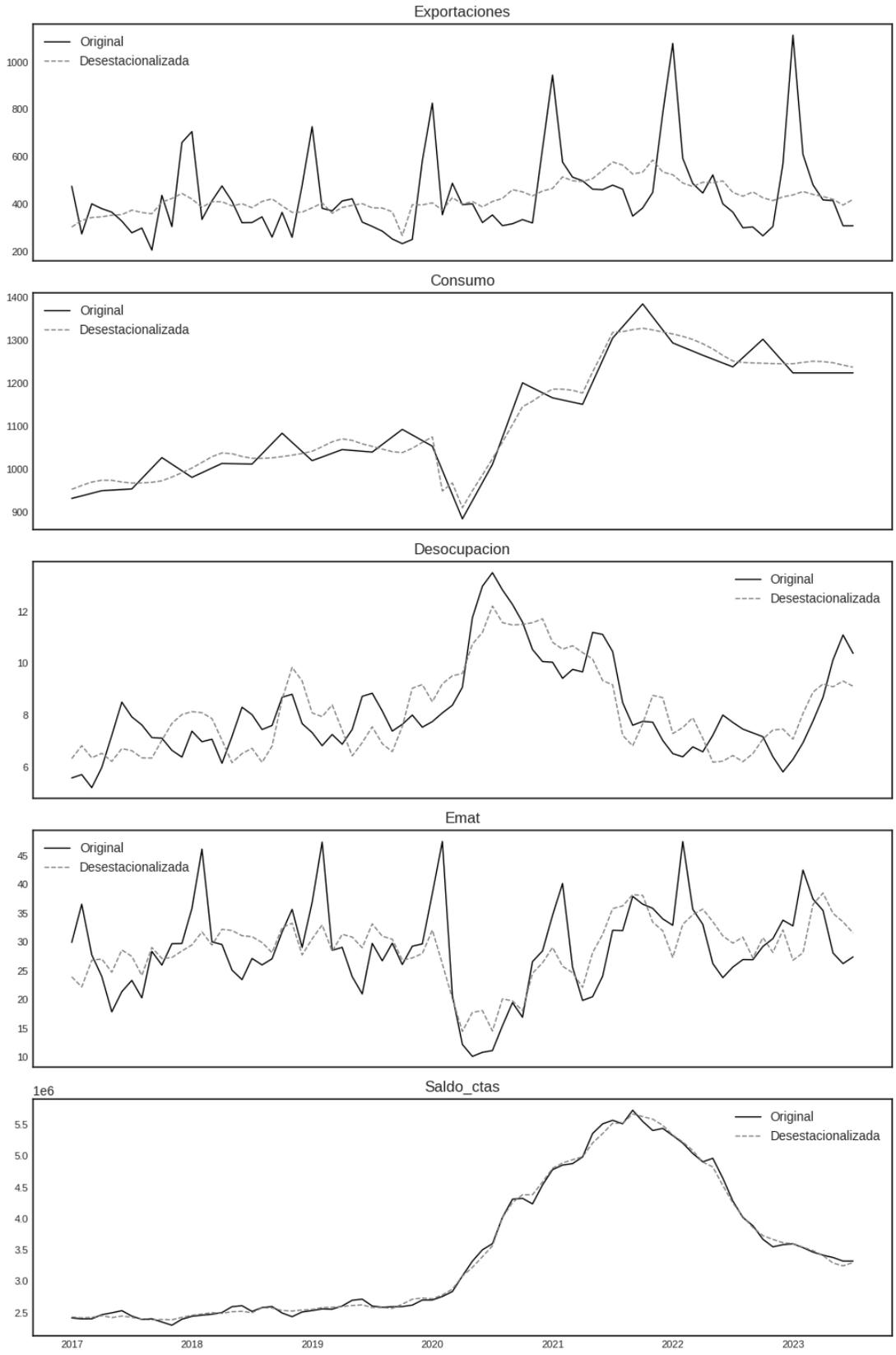
Adicionalmente, se buscó garantizar que los residuos siguieran una distribución normal. Para ello, empleamos el coeficiente Shapiro. En nuestra serie de "exportaciones", por ejemplo, un valor de 0.9901 se acerca al ideal de 1, lo que confirma la normalidad deseada.

Tabla 3: Ajuste estacional de las series

Serie	Modelo ARIMA	Transformación	AICc	BIC	Box-Ljun <sup>3</sup>	Shapiro
Exportaciones	(1 0 0)(1 1 0)	log	745.8	763.9	23.72	0.9901
Consumo	(2 1 1)(0 1 1)	none	473.1	494.5	27.31	0.9859
Desocupación	(1 1 0)(0 1 1)	none	123.8	131.9	29.1	0.992
EMAT	(1 0 0)(0 1 1)	none	386.4	396.4	22.1	0.9891
Saldo cuentas	(3 1 1)(1 0 1)	log	1990	2006	25.51	0.9953

Los criterios de selección del modelo ARIMA, la efectiva desestacionalización evidenciada por los valores QS, y la validación de los residuos a través de las pruebas Box-Ljung y Shapiro convergen hacia un conjunto de series bien ajustadas y modeladas. Estas series, al ser robustas y representativas, se convierten en candidatas ideales para la construcción de un indicador al estilo Stock y Watson, lo que garantizará una representación precisa y confiable de la actividad económica subyacente. A continuación, se presenta cada serie en su versión original y desestacionalizada.

<sup>3</sup>  $Q = n(n + 2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k}$  donde  $\hat{\rho}_k$  es la autocorrelación muestral de orden  $k$ ,  $n$  es el tamaño de la muestra, y  $h$  es el número de retrasos.



## Resultados

Con las series desestacionalizadas y normalizadas, se aplicó el Modelo Factorial Dinámico (DFM) como una herramienta analítica clave para identificar y describir las interacciones latentes entre las variables seleccionadas. La relación fundamental entre las series temporales y los factores latentes en el DFM se basa en la Ecuación (1).

El Modelo Factorial Dinámico (DFM) adoptado en este estudio ha sido estructurado con una configuración basada en un único factor latente y un orden de 1. Esta elección resalta nuestra intención de capturar las interdependencias complejas que puedan existir entre las series temporales con un factor común que pueda encapsular estas dinámicas. Este enfoque nos permite aislar y examinar el efecto de un factor dominante que podría estar influenciando todas las series simultáneamente.

Por otro lado, la dinámica temporal de los factores latentes se describe mediante la Ecuación (2) Esta ecuación, que modela los factores como un proceso autorregresivo, es vital para entender cómo los factores evolucionan a lo largo del tiempo y cómo se transmiten los shocks o perturbaciones entre ellos.

Además, es crucial mencionar que las dinámicas de los factores latentes no están aisladas; interactúan entre sí, y estas interacciones se pueden describir y analizar mediante matrices de covarianza, como se detalla en las ecuaciones asociadas al Filtro de Kalman. Estas ecuaciones, particularmente las Ecuaciones (3) a (6), iteran a través del tiempo, refinando las estimaciones de los factores a medida que se procesan las observaciones. Esta característica permite que el modelo DFM sea adaptativo y capaz de incorporar nueva información a medida que se dispone de ella.

El Modelo Factorial Dinámico (DFM) se implementó sobre las series desestacionalizadas y normalizadas con el propósito de identificar interacciones latentes entre las distintas series temporales estudiadas.

Las métricas de información, como el Criterio de Información de Akaike (AIC) y el Criterio de Información Bayesiano (BIC), ofrecen una evaluación cuantitativa del ajuste del modelo en relación a su complejidad. Con un AIC de -1207.12 y un BIC de -1181.20, el modelo demuestra un ajuste razonable a los datos.

Los factores de carga resultantes reflejan el grado en el que cada serie temporal está influenciada por el factor latente. Un factor de carga más alto indica una mayor influencia del factor latente en la respectiva serie. Los factores de carga, denotados como  $\lambda_i$  representan la relación entre el factor común  $F_t$  y cada serie temporal  $Y_{it}$  según:  $Y_{it} = \lambda_i F_t + \varepsilon_{it}$ .

Al examinar las cargas factoriales, vemos que las diferentes series tienen diferentes grados de asociación con el factor latente. Específicamente, la serie 'Saldo cuentas' tiene una carga de -0.0128, lo que indica una fuerte y negativa relación con el factor. Estos valores, en términos matemáticos, representan la proyección de cada serie en el espacio latente definido por el factor.

Tabla 4: Factores de carga

Serie	Coeficiente	Std. Err	Z-valor	P> z	[0.025	0.975]
Exportaciones	-0.0076	0.007	-1.159	0.246	-0.020	0.005
Consumo	-0.0039	0.002	-2.241	0.025	-0.007	-0.000
Desocupación	-0.00008	0.008	-0.011	0.991	-0.015	0.015
EMAT	-0.0051	0.008	-0.601	0.548	-0.022	0.012
Saldo cuentas	-0.0128	0.003	-4.162	0.000	-0.019	-0.007
<i>Factor equation f1</i>	<i>0.8701</i>	<i>0.061</i>	<i>14.192</i>	<i>0.000</i>	<i>0.750</i>	<i>0.990</i>

## Evaluación de Residuos

Para verificar la calidad de nuestro modelo, hemos realizado diversas pruebas estadísticas sobre los residuos, centrando nuestra atención en la independencia temporal, la normalidad y la homocedasticidad. En particular, se evalúan los estadísticos Ljung-Box, Jarque-Bera, Heterocedasticidad, asimetría y curtosis.

La prueba de Ljung-Box se utiliza para examinar la independencia temporal de los residuos a partir de la existencia de auto correlaciones en ellos. La hipótesis nula ( $H_0$ ) es que los datos son independientes (es decir, no auto correlacionados), mientras que la hipótesis alternativa ( $H_1$ ) es que los datos muestran Autocorrelación. Un valor  $p$  superior a 0.05 sugiere que el modelo ha capturado adecuadamente la estructura temporal. La estadística de la prueba se define como:  $Q = n(n + 2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k}$ , donde  $\hat{\rho}_k$  es la autocorrelación muestral de orden  $k$ ,  $n$  es el tamaño de la muestra, y  $h$  es el número de retrasos.

Si  $Q$  sigue aproximadamente una distribución  $\chi^2$  con  $h$  grados de libertad bajo la hipótesis nula. Si el valor  $p$  asociado a  $Q$  es pequeño (por ejemplo, menor que 0.05), se rechaza la hipótesis nula, lo que indica autocorrelación en los residuos.

Por otro lado, la prueba de Jarque-Bera es una prueba de bondad de ajuste que examina si una serie de datos tiene la asimetría y la curtosis que se esperarían de una distribución normal. Es comúnmente utilizada para probar la normalidad de los residuos y se define como  $JB = \frac{n}{6} \left( s^2 + \frac{1}{4}(K - 3)^2 \right)$  donde  $n$  es el número de observaciones,  $S$  es la asimetría de la muestra,  $K$  es la curtosis de la muestra. Bajo la hipótesis nula ( $H_0$ ) de que los datos provienen de una distribución normal,  $JB$  sigue asintóticamente una distribución  $\chi^2$  con 2 grados de libertad. Si el valor  $p$  asociado es pequeño (menor a 0.05), se rechaza la hipótesis nula, indicando que los datos no siguen una distribución normal.

En orden de profundizar en la heterocedasticidad, se evaluó la constancia de la varianza de los residuos. Valores de  $H$  cercanos a 1 sugieren que la varianza de los residuos es constante a lo largo del tiempo. Se examinó la constancia de la varianza de los errores a lo largo del tiempo. Matemáticamente,  $Var(u_t|X_t) \neq \sigma^2$  donde  $u_t$  es el error en el tiempo  $t$ ,  $X_t$  es un conjunto de variables explicativas. Si la varianza condicional no es constante ( $\sigma^2$ ) hay heteroscedasticidad. Un valor de  $H$  significativamente diferente de 1 podría indicar la presencia de heterocedasticidad.

Tabla 5: Evaluación de Residuos

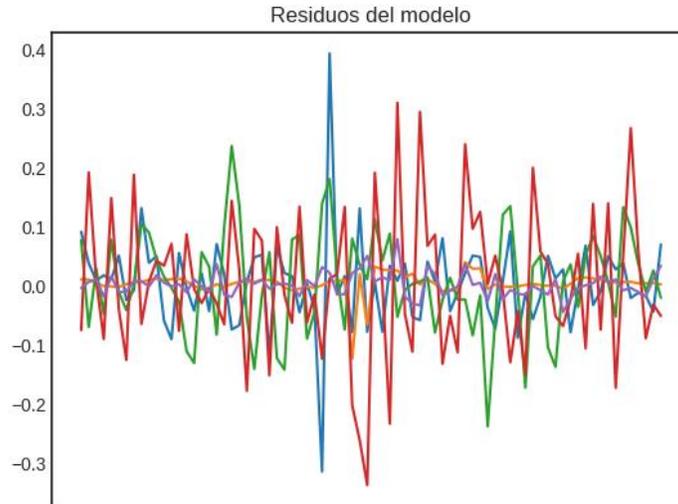
Serie	Ljung-Box (Q)	Prob(Q)	Jarque-Bera (JB)	Prob(JB)	Heteroskedasticity (H)	Prob(H)	Skew	Kurtosis
Exportaciones	9.09	0.00	308.84	0.00	0.68	0.34	0.74	12.63
Consumo	0.14	0.71	2111.20	0.00	2.16	0.06	-4.06	27.16
Desocupación	4.50	0.03	0.46	0.79	1.09	0.83	-0.14	3.24
EMAT	2.41	0.12	0.19	0.91	1.20	0.65	0.11	3.11
Saldo cuentas	0.18	0.67	29.64	0.00	2.00	0.08	1.02	5.23

En primer lugar, los resultados de la prueba de Ljung-Box indican que series como Consumo con un valor p de 0.71, Sakdo cuentas con 0.67 o EMAT con 0.12 muestran que sus residuos no presentan auto correlaciones significativas, lo cual es deseable.

Observando las puntuaciones, la serie Consumo tiene un p-valor de 0.0, lo que sugiere que sus residuos no siguen una distribución normal. Sin embargo, la serie Desocupación con un p-valor de 0.7942 sugiere que sus residuos sí se ajustan a una distribución normal. Destaca a su vez la serie Consumo con un alto valor de 2111.20 y una probabilidad cercana a 0, lo que indica una desviación significativa de la normalidad.

Enfocándonos en los valores más destacados de la tabla, es evidente que hay varias dinámicas subyacentes en las diferentes series temporales. La serie exportaciones muestra signos claros de no autocorrelación con un valor de Ljung-Box de 9.09 y un p-valor cercano a cero. Sin embargo, su asimetría positiva de 0.74 y una alta curtosis de 12.63 sugieren una distribución lejos de la normalidad. Por otro lado, la serie Consumo presenta una fuerte desviación de la normalidad, con un inmenso valor de Jarque-Bera de 2111.20, y su asimetría negativa de -4.06 respalda este hallazgo. Aunque sus residuos parecen no tener autocorrelación, la heterocedasticidad es pronunciada con un valor de 2.16. Por último, es curioso observar que la serie Desocupación tiene residuos que se aproximan a una distribución normal, pero su asimetría ligeramente negativa y su baja curtosis sugieren algunas peculiaridades.

A continuación, se presenta un gráfico visual de los residuos para cada serie, que proporciona una representación intuitiva de su distribución y posibles patrones no capturados por el modelo



La matriz de error de covarianza, representada en la Tabla 6, ofrece información sobre la varianza y covarianza de los errores de las series temporales del modelo. Es esencial examinar esta matriz para identificar posibles correlaciones no capturadas entre las series temporales.

Matemáticamente, si denotamos  $\mathbf{u}_t$  como el vector de errores en el tiempo  $t$  y  $E[\ ]$  como el operador de expectativa, la matriz de covarianza del error  $\Sigma_u$  está dada por:  $\Sigma_u = E[\mathbf{u}_t \mathbf{u}_t']$  donde  $\mathbf{u}_t'$  es el transpuesto del vector de errores. La diagonal de  $\Sigma_u$  contiene las varianzas  $\sigma_1^2, \sigma_2^2, \dots$  de los errores de cada serie, y los elementos fuera de la diagonal representan las covarianzas entre los errores de diferentes series.

La serie exportaciones tiene un coeficiente de error de covarianza de 0.0059, que es significativo. Esto indica una variabilidad no explicada por el factor común. En el caso de Consumo, su coeficiente de 0.0004, aunque pequeño, es altamente significativo, lo que sugiere que, aunque hay una parte no explicada por el factor común, esta variabilidad es relativamente baja.

Tabla 6: Matriz de covarianza de errores

	coef	std err	z	P> z	[0.025	0.975]
sigma2.Exportaciones	0.0059	0.000	12.477	0.000	0.005	0.007
sigma2.Consumo	0.0004	0.00002	17.373	0.000	0.000	0.000
sigma2.Desocupación	0.0072	0.001	6.196	0.000	0.005	0.009
sigma2.EMAT	0.0158	0.004	4.454	0.000	0.009	0.023
sigma2.Saldo cuentas	0.0001	0.00005	2.512	0.012	0.00003	0.009

La matriz de covarianza de errores ofrece detalles técnicos sobre la varianza no explicada para cada serie. Estas métricas son esenciales para entender el grado en que el factor común puede o no puede explicar la variabilidad en las series. Es notable que la serie 'EMAT' tiene la varianza residual más alta de 0.0158, indicando que, a pesar de la influencia del factor común, hay dinámicas únicas en esta serie que el modelo no captura completamente

Conectar estos hallazgos con los resultados anteriores, es evidente que, aunque el factor común captura una parte sustancial de la dinámica de las series, todavía hay variabilidades específicas en cada serie que no son completamente explicadas. Esta observación respalda la necesidad de incluir estas series en el modelo, incluso si algunas de ellas pueden presentar características no ideales desde el punto de vista de los residuos, ya que proporcionan una representación más completa del fenómeno subyacente que se está modelando.

La naturaleza compleja de las series temporales examinadas requiere un enfoque sofisticado como el Modelo Factorial Dinámico que hemos implementado. Es evidente que, aunque algunas series, como Consumo, presentan residuos con características no deseables, la representación factorial se justifica para capturar las relaciones latentes y las dinámicas compartidas entre las distintas series. El sacrificio de ciertas propiedades estadísticas es necesario cuando se intenta sintetizar la información contenida en varias series relacionadas.

Sin embargo, el factor latente claramente desempeña un papel vital en explicar las fluctuaciones conjuntas y las interdependencias. La heterocedasticidad y las desviaciones de la normalidad en los residuos, aunque no ideales, son contrapartidas de un esfuerzo

para extraer una señal común subyacente de un ruido específico de cada serie. En última instancia, es crucial recordar que la modelización es una tarea de equilibrio, y el Modelo Factorial Dinámico proporciona un marco robusto para analizar múltiples series simultáneamente, a pesar de las imperfecciones individuales en la adaptación de cada serie.

## **Comparación PIB regional**

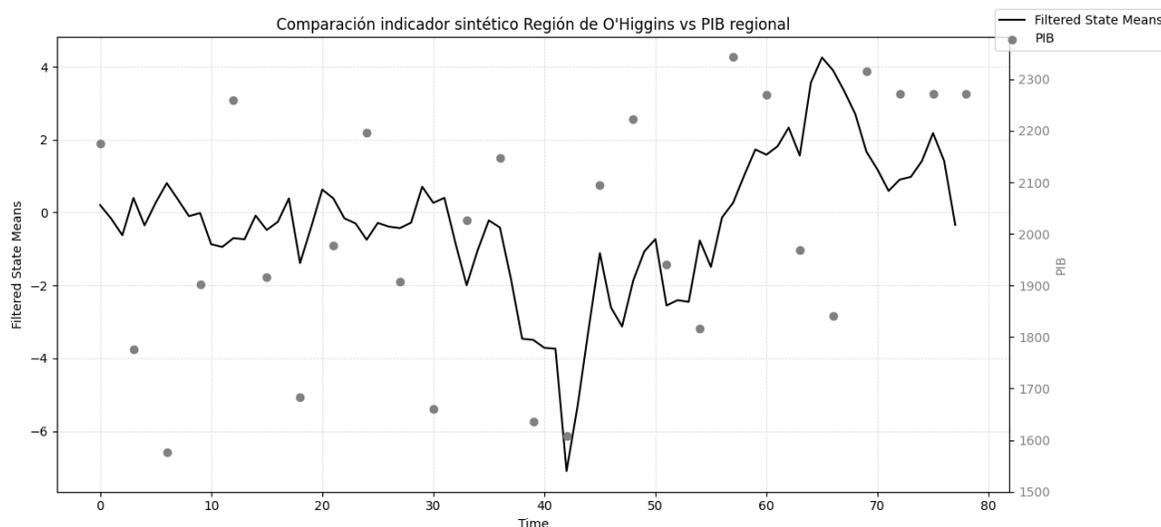
Se comparan las series temporales del indicador desarrollado con el PIB regional. Aunque no se disponga de otras métricas para una comparativa más amplia, es una práctica estándar en estudios que emplean el DFM contrastar el indicador principal con métricas macroeconómicas consolidadas, como el PIB, para evaluar su pertinencia y relevancia en el contexto económico.

Al analizar el comportamiento de nuestro indicador y el PIB regional, se observa una relación intrigante. A pesar de ser series que en ocasiones presentan movimientos disímiles, se detecta que ambos tienden a moverse en la misma dirección en aproximadamente 6 de cada 10 ocasiones. Es decir, cuando nuestro indicador muestra una tendencia al alza o a la baja, el PIB de la región tiende a seguir esa misma dirección con cierta frecuencia.

Esta coincidencia en tendencias es particularmente notable considerando que estamos comparando un indicador de desarrollo propio, único en su tipo, con un agregado macroeconómico ampliamente reconocido como es el PIB. Mientras el PIB proporciona una perspectiva holística y amplia del comportamiento económico de la región, nuestro indicador ofrece insights más detallados y específicos.

Lo que es verdaderamente relevante aquí es que, a pesar de su naturaleza detallada y específica, el indicador no sólo refleja cambios macroeconómicos regionales, sino que también tiene el potencial de anticipar tendencias. Específicamente, puede capturar fluctuaciones más rápidas o responder a eventos económicos específicos que el PIB podría tardar más tiempo en reflejar.

La capacidad de nuestro indicador para coincidir y, en ciertos casos, anticipar la dirección del PIB, subraya su valor como herramienta analítica. En un contexto donde no existen indicadores comparables, este emerge no sólo como una innovación, sino también como una pieza esencial para la toma de decisiones informadas en la región.



## Reflexiones finales

La adopción del Modelo Factorial Dinámico en el contexto de series temporales para la Región de O'Higgins ha permitido desentrañar complejas interrelaciones económicas subyacentes. Mediante el análisis de los factores de carga, hemos identificado que ciertas dimensiones económicas están intrínsecamente influenciadas por el factor latente, evidenciando una interacción macroeconómica común que permea diversas métricas de la región.

La validación diagnóstica, si bien importante, ha evidenciado que ciertas series podrían beneficiarse de optimizaciones adicionales. No obstante, la congruencia de la mayoría de las series con el modelo refuerza la robustez del enfoque adoptado. Es esencial recalcar la relevancia de este modelo en proporcionar insights sobre la economía de la Región de

O'Higgins, particularmente en el contexto de construir un indicador sintético que refleje con precisión la dinámica económica regional.

Una observación sustantiva del análisis es la relación emergente entre el indicador construido y el PIB de la región. A pesar de que se identificó una capacidad limitada del indicador para anticipar el signo del cambio del PIB, en un contexto donde carecemos de otros indicadores comparativos, esta relación adquiere relevancia, al menos en términos de tendencias generales.

Futuras investigaciones podrían expandir el alcance de este trabajo considerando una perspectiva más amplia, tal como la del IMACEC nacional. Esto permitiría una comparación más directa entre las dinámicas regionales y las nacionales. Una posible metodología sería aplicar filtros de Kalman para cada región y luego ponderarlos para ajustarse al ciclo del PIB nacional. Esta ponderación inicial podría ser abordada mediante el método de los mínimos cuadrados ordinarios (OLS), pero una alternativa más avanzada involucraría el uso de redes neuronales, específicamente aprovechando técnicas como el gradiente descendente. El beneficio de usar redes neuronales radica en su capacidad para capturar relaciones no lineales y complejas interacciones entre variables. Esto permitiría determinar una ponderación óptima de cada región, alineando con precisión cada factor regional con el ciclo económico nacional y, a su vez, calibrar este último en relación con el IMACEC

Concluyendo, el Modelo Factorial Dinámico ha demostrado ser una herramienta valiosa en el análisis económico de la Región de O'Higgins. La evolución continua de las circunstancias económicas subraya la necesidad de revisiones y refinamientos metodológicos periódicos para asegurar que el modelo y el indicador sigan siendo pertinentes y reflejen fielmente la realidad económica de la región.

## Bibliografía

- Abadir, K. M., Distaso, W., & Giraitis, L. (2014). An cointegration model with piecewise linear trends: Likelihood analysis and application. *Journal of Econometrics*, 178, 383-396.
- Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191–221.
- Banco de Inglaterra. (2003). A Kalman filter approach to estimating the UK NAIRU. Por J. V. Greenslade, R. G. Pierse y J. Saleheen. Working Paper no. 179, Publications Group, Bank of England.
- Federal Reserve Board (FRB) de Estados Unidos (2020), tal como se refleja en la publicación 'Improving the Forecast Accuracy of Provisional Data: An Application of the Kalman Filter to Retail Sales Estimates'.
- Feldkircher, M. (2015). The determinants of vulnerability to the global financial crisis 2008 to 2009: Credit growth and other sources of risk. *Journal of International Money and Finance*, 51, 49–77.
- Gobierno de Santa Fe. (2014). ISAE - Indicador Sintético de la Actividad Económica. [www.santafe.gov.ar](http://www.santafe.gov.ar).
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1), 35–45.
- Sigal, F., Camusso, J., & Navarro, A. (2022) Argentine Regions Based on Dynamic Criteria en SSRN: [enlace del documento].
- Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2), 147–162
- Stock, J. H., & Watson, M. W. (1989). A probability model of the coincident economic indicators. *NBER, Macroeconomics Annual*, 351–394.