## Introducción

Los indicadores sintéticos son herramientas que combinan múltiples variables o indicadores individuales en una sola medida para proporcionar una visión más amplia y simplificada de una situación específica. Estos indicadores suelen utilizarse en diversos campos, como economía, finanzas, medio ambiente, salud pública y más. Así, por ejemplo, el Producto Interno Bruto (PIB) tal y como es calculado para las series trimestrales no deja de ser un indicador sintético que resume la actividad económica de un país, considerando la producción de bienes y servicios en un solo número.

La creación de estos indicadores puede requerir métodos estadísticos o matemáticos para ponderar y combinar las diferentes variables de manera que reflejen con precisión la situación que se desea medir. Sin embargo, la selección de las variables y la forma en que se combinan pueden generar debates, ya que la interpretación de estos indicadores sintéticos, a veces, puede simplificar en exceso realidades complejas.

Al iniciar un proyecto de indicadores sintéticos, es importante primero definir claramente el objetivo que se busca y el problema que se intenta resolver. En este caso, el objetivo es elaborar un indicador sintético de actividad para la economía andaluza (en adelante ISAEA) que sea capaz de informar en frecuencia mensual la evolución de la actividad andaluza. No este un caso en el que se quiera sustituir al PIB como referencia de la evolución de la actividad (ni podría), sino de informar en una frecuencia mayor sobre la evolución de dicha actividad que, a su vez, puede ser también medido con mayor precisión por el PIB. Una vez definido el objetivo, el siguiente paso clave es identificar las principales variables o indicadores individuales que capturan los aspectos más relevantes del fenómeno que se busca medir, y que se desean incorporar en el indicador sintético final.

Cuando se tienen identificadas las variables individuales, es necesario recolectar los datos correspondientes a esos indicadores para los casos (países, regiones, etc.) y años que se incluirán en el análisis. Es muy importante revisar la integridad y calidad de los datos, realizando una limpieza en caso de ser necesario, para asegurarse que la información esté completa, al que le seguirá un tratamiento y estandarización de dicha información. Así, mediante dicha estandarización de indicadores, se eliminará el ruido posible que pueda transmitir las distintas métricas y escalas que no sean comparables y medidos en una escala común, así como la eliminación de la información que no resulte relevante para el objetivo, como son los factores estacionales o el ruido generado por eventos repetitivos más allá de la estacionalidad como es el calendario.

Sin embargo, la tarea más compleja de la realización de un indicador sintético es la de definir las reglas de agregación de la información. Esta agregación tradicionalmente se lleva a cabo mediante la asignación de ponderaciones o pesos a cada una de las variables, pesos que se suelen asignar de acuerdo con su relevancia en la conformación del indicador sintético final. Con estos pesos se da más o menos importancia a ciertos indicadores en función del fenómeno que se estudia. Las ponderaciones suelen obtenerse mediante diversos métodos, pero los más habituales son los derivados de juicios de expertos, a partir de análisis estadísticos (como son la regresión, el uso de simples correlaciones o el análisis factorial) o simplemente, no discriminar por variables dando a todas ellas la misma preponderancia.

Sin embargo, es muy necesario justificar cómo se calibran dichos pesos y qué se pretende con ello. Una opción es recurrir a fuentes estadísticas externas, como son los pesos de los sectores o de las ramas de actividad en el PIB de la economía, en el caso de un indicador como el que se quiere realizar para Andalucía. Otra opción es el de recurrir a estadísticos endógenos a la información ofrecida por los indicadores a usar. En este segundo caso, la idea sería usar correlaciones, coeficientes de regresiones u otras métricas que permitan definir la relevancia de cada uno de los indicadores a la hora de incorporarse al indicador sintético.

En nuestra consideración, la segunda opción representa la alternativa más favorable. En este escenario, los datos mismos, sin intervención directa del analista, determinan su propio peso en la creación del indicador. Sin embargo, este peso estará, en última instancia, sujeto a las decisiones sobre qué métrica se empleará para definirlos. Tal como se detalla en este informe, siempre surgirán decisiones sobre la metodología a utilizar para identificar qué indicadores se incluyen y con qué nivel de relevancia en el indicador sintético. Por tanto, la justificación detrás de estas decisiones debe estar enfocada en maximizar la objetividad y la calidad del indicador construido, alineado con los objetivos planteados.

En el resto del informe se describe una primera aproximación para la construcción del indicador sintético. Para ello se van a utilizar los indicadores suministrados por el IECA y se va a plantear una metodología que, dados los condicionantes existentes y que serán detallados a lo largo del texto, podría llevarnos a obtener un indicador objetivo y que maximice el uso de la información suministrada por todos y cada uno de los indicadores individuales. No obstante, consideramos que este ejercicio que se presenta en este informe es solo una primera aproximación y que existen metodologías de reducción de variables que podrían ofrecer indicadores que mejoren el cumplimiento de los objetivos propuestos.

## Un indicador de actividad para la economía Andaluza

En primer lugar, antes de construir un indicador es necesario realizar una serie de pasos que nos permitan eliminar de los indicadores individuales la información que no es relevante para el seguimiento de la actividad cíclica andaluza. Así, para construir un indicador sintético de actividad económica y que minimice las diferencias con el perfil cíclico mostrado por el PIB de un país o, en este caso, de una región, se deben realizar los siguientes pasos.

#### Selección y tratamiento de series

En primer lugar, se debe seleccionar una batería de indicadores potenciales y, entre los cuáles, poder escoger aquellos que de forma conjunta puedan aproximar de la mejor forma posible la actividad de la economía andaluza. Para ello deben seleccionarse previamente aquellos que muestren una estrecha relación contemporánea con el PIB y que cubran los principales sectores de la economía. Entre ellos, y a propuesta del IECA, se han seleccionado una batería de 14 indicadores, mensuales y trimestrales que mejor podrían capturar la información cíclica del PIB andaluz.

Estos indicadores son los siguientes:

* Consumo aparente de cemento
* Consumo de gasolina y gasóleo
* Exportaciones de bienes
* Importaciones de bienes
* IPIAN Div 23
* Índice general de producción industrial
* ICMA General sin estaciones de servicios
* Matriculación de turismos
* Índice ventas grandes superficies
* Pernoctaciones
* Cifra negocios del sector servicios
* Liquidación de presupuestos de la Junta de Andalucía. Capítulo 1
* Total de afiliados SS Total
* Índice de producción agrícola.

Los requisitos para que estas series puedan constituirse como candidatas a incluirse en el ISAEA son, al menos, tres. En primer lugar, como se ha avanzado, deben tener o bien una vinculación directa con la evolución de la actividad productiva andaluza general, o bien deben reflejar la evolución en sectores parciales de la actividad productiva. En segundo lugar, se debe exigir de estas series que se extiendan para un periodo de tiempo suficientemente largo. En tercer lugar, deben aportar una garantía mínima de calidad.

Una vez hecha la selección y aplicado un filtro de calidad y suficiencia, se procede a eliminar aquellos componentes de las series que no son necesarios, y que incluso ocultan los componentes más relevantes. Junto con la rigurosa revisión y limpieza de los datos faltantes o atípicos es necesario proceder a la tradicional desestacionalización y eliminación de los efectos del calendario en la evolución de los datos de frecuencia superior a la anual[[1]](#footnote-1).

#### Estacionariedad de los indicadores

Cualquier análisis que implique la búsqueda del mejor modelo posible para relacionar estadísticamente series temporales debe incorporar, en su inicio, además de un tratamiento como los indicados en el apartado anterior, el de transformar dichas series en series estacionarias. Es necesario que, para establecer relaciones econométricas entre series temporales, estas sean ergódicas. Generalmente dicha ergodicidad suele alcanzarse mediante la diferenciación de las series que la necesiten.

Para conocer el grado de diferenciación necesaria para transformar en estacionarios los indicadores individuales se aplica el test de Dickey-Fuller al logaritmo neperiano de cada serie. La implementación, para agilizar el procedimiento, se realizó automáticamente mediante la aplicación de un algoritmo que determinaría el grado de integración de cada serie en función de si cumple o no la hipótesis nula que marca el test aplicado.

Así, siendo el indicador, el algoritmo en el primer paso aplica el test. Si se rechaza la hipótesis nula de estacionariedad, se procede a calcular su diferenciación, obteniendo una nueva serie . La diferenciación en tres períodos responde a la idea de que la aproximación de la actividad se va a establecer respecto a una serie trimestral (el PIB) y a su crecimiento intertrimestral. A se le vuelve a aplicar el test de estacionariedad y si se acepta, el proceso finaliza. Si se rechaza, se vuelve a calcular la diferenciación hasta lograr que la hipótesis nula sea aceptada. El mismo procedimiento se replica para los datos trimestrales, aunque en este caso con diferenciaciones de un trimestre. Una vez finalizado se registra la cantidad de transformaciones necesarias para cada columna, lo que permite identificar y adecuar las series que requieren ajustes más significativos.

Sin embargo, la diferenciación de algunas series no es adecuadamente asignada por el proceso automático. Es el caso de 'Pernoctaciones', y cuya particular estructura derivada de los meses de COVID dificultan la aplicación de los tests. Es por ello por lo que a esta serie se le asigna manualmente una integración de orden 1 tras una tradicional comprobación visual de su comportamiento.

El resultado después del tratamiento y dela diferenciación hasta lograr la estacionariedad de las series puede observarse en las figuras 1 y 2, para las series mensuales y trimestrales respectivamente.

**Figura 1. Series mensuales CVEC y estacionarias.**Imagen que contiene Calendario

Descripción generada automáticamente

**Figura 2. Series trimestrales CVEC y estacionarias.**

Gráfico

Descripción generada automáticamente

## Optimización de la selección de series para el ISAEA.

#### El uso de la correlación como identificación

El objetivo de esta sección es el de identificar cuáles son los indicadores preferidos a la hora de construir el ISAEA mediante el uso de una metodología objetiva. En este caso concreto, s propone el uso de regresiones por mínimos cuadrados (OLS) para justificar el cálculo del indicador sintético mediante el uso de estadísticos endógenos (generados por los propios datos). No obstante, y como se adelantó, este ejercicio solo corresponde a una aproximación previa que se concretará, finalmente, por el uso de métodos de reducción de la dimensionalidad para elaborar un indicador sintético de actividad para la economía andaluza (ISAEA).

Así pues, y a pesar de no corresponder a un ejercicio definitivo, este nos permite valorar qué grado de aproximación podemos logar, así como fijar algunos elementos de la metodología definitiva que se aplicará para construir el ISAEA definitivo.

En primer lugar, y dado que el indicador objeto de aproximación tiene frecuencia trimestral, lo primero que debemos hacer es transformar los indicadores mensuales a esta frecuencia, aunque el indicador final tendrá frecuencia mensual. Una vez agregado a una frecuencia menor, mediante el uso de la media, se puede realizar una primera aproximación a la validez de los indicadores como potenciales fuentes de información para el seguimiento de la actividad económica andaluza. Para ello, el primer análisis natural corresponde a la visualización de la correlación contemporánea entre indicador y PIB trimestral de Andalucía. La figura 3 muestra este ejercicio, donde puede observarse que la correlación entre los diversos indicadores en tasa de crecimiento intertrimestral es elevada respecto al PIB, con excepción de los indicadores de producción agrícola, la liquidación de presupuestos y las exportaciones de bienes y servicios.

**Figura 3. Correlación entre el crecimiento de los indicadores y PIB trimestral**

Gráfico, Gráfico de barras

Descripción generada automáticamente

Una vez conocida la correlación, la primera idea por la que estaríamos tentados a proceder sería utilizar aquellos indicadores con un mayor valor de esta. Sin embargo, esto podría suponer un error por varias razones. En primer lugar, la falta de correlación simplemente indica la incapacidad de un indicador para aproximar por sí solo la evolución del PIB a lo largo del tiempo a partir de una especificación lineal. No obstante, esto no indica su capacidad para transmitir información relevante que no sea aportada por el resto de los indicadores. Es posible que su utilidad no sea nula a pesar de una baja correlación.

En segundo lugar, es importante considerar el período comprendido desde el primer trimestre de 2020 hasta mediados de 2021, el cual representa una fase que introduce variabilidad adicional en las series temporales. Esto dificulta la evaluación de las relaciones entre estas mediante medidas simples como el coeficiente de correlación u otras similares. Durante este corto período de tiempo, la variabilidad experimentada fue considerablemente alta en comparación con la tendencia histórica. Esta situación implica que, aunque las series temporales sean extensas, la correlación en estos meses recientes tendría un peso significativo en la correlación total entre las variables, potencialmente sesgando los resultados a favor de indicadores cuyo comportamiento durante la pandemia se asemejó más al del PIB. Las figuras 1 y 2 ilustran cómo cada serie temporal absorbió el impacto de la pandemia. A pesar de mostrar perfiles similares, algunas de estas series lo hicieron con una variabilidad más o menos intensa en relación con la del PIB. Por lo tanto, el uso exclusivo de este estadístico para seleccionar los indicadores que mejor reflejen la actividad económica de la región definida por el PIB puede no ser la opción más adecuada.

La idea, por ende, es emplear otras aproximaciones que nos permitan, en primer lugar, seleccionar indicadores por su relevancia para replicar la evolución de la actividad económica definida por el PIB y, en segundo lugar, extraer de ellos la información necesaria con la garantía de hacerlo correctamente.

Por todas estas razones, la estrategia previa para identificar los indicadores preferidos en la elaboración del indicador es la de encontrar la combinación de estos que mejor replique la evolución del PIB, condicionado a la particularidad que supone la existencia de fuertes "outliers" en torno a los meses de la pandemia. La variación de las series en dichos meses y la fuerte correlación “artificial” con la serie objetivo que es el PIB implica que debamos considerar estrategias de identificación que sean sensibles a este hecho y que, por lo tanto, descarten estadísticos que, o bien, incorporen dichos meses en su cálculo o bien nos permita elegir entre aquellos que mejor aproximen la evolución “a pesar” de dichos meses. Si esto no se hace, se podrá estar cometiendo un error de asignar a una serie una fuerte capacidad de información cuando, en realidad, lo que sucede es que habrá una clara dominancia de la variación de la actividad recogida por dicho indicador y el PIB entre marzo de 2020 y septiembre de 2020.

Por ejemplo, si realizamos una regresión entre el PIB y todos los indicadores, como se muestra en la siguiente tabla, el coeficiente de determinación será excepcionalmente alto, lo que implica una alta precisión de ese modelo. Sin embargo, este resultado se ve afectado, además de una evidente multicolinealidad, por los datos extremadamente atípicos de la pandemia, y que dominarán al resto por el mero hecho de su valor extremo.

Imagen que contiene Tabla

Descripción generada automáticamente

Como se puede apreciar en la figura 4, la capacidad predictiva del modelo es alta (salvo, paradójicamente, su incapacidad para aproximar los datos extremos de la pandemia), aunque no hay garantía de que el mejor modelo sea aquél que utiliza todos los indicadores. Además de lo mencionado anteriormente, debemos considerar el riesgo tradicional de sobreajuste en un modelo de regresión, así como el problema de multicolinealidad, que se manifiesta en signos inesperados en indicadores con correlaciones positivas respecto a la variable dependiente.

**Figura 4. Serie de PIB observada y predicha por la regresión (tasa de crecimiento intertrimestral)**

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Una prueba de lo que se está comentando, del posible sesgo en la estimación del modelo que supone considerar los meses de la pandemia, se pueden observar en la regresión que se obtendría si esta se realizara con los datos hasta diciembre de 2019. El coeficiente de determinación del modelo decae hasta el 85,9 %, lo que muestra que tratar de aproximar las variables que mejor aproximan al PIB usando las fechas que corresponden a la pandemia no es la mejor estrategia.

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media

La figura 5 muestra la evolución de la serie de PIB predicho y observada cuando se usa el modelo con datos hasta diciembre de 2019. Es obvio que nos muestra una aproximación bastante aceptable, pe

**Figura 5. Serie observada y predicha del PIB hasta diciembre de 2019**

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Para finalizar con esta reflexión, podemos mirar la figura 6 como una muestra más del sesgo que puede suponer tomar en consideración los meses de la pandemia. Lo que nos dice esta figura, como la anterior figura 3, es que los indicadores, en general, mantienen una elevada correlación respecto al PIB. Sin embargo, esta es diferente para algunos de ellos. Nada mejor que para poder valorar esta diferencia que mostrar en una misma figura no tanto las correlaciones sino los coeficientes de ambas regresiones lineales mostradas en un solo gráfico. Esto es lo que nos enseña la figura 7.

**Figura 6. Coeficiente de correlación hasta diciembre de 2019**

Gráfico, Gráfico de barras

Descripción generada automáticamente

Así, y por ejemplo, para las ventas en grandes superficies, el índice de producción agrícola, las exportaciones, la liquidación de presupuestos y alguno otro indicador, los coeficientes en la regresión son significativamente diferentes dependiendo de si se incluyen o no los meses posteriores a la pandemia. Esto, obviamente, como se ha adelantado, obliga a tener en consideración este período antes de seleccionar o elegir cuáles son los mejores indicadores para aproximar la evolución de la actividad andaluza.

**Figura 7. Coeficientes de las regresiones dependiendo si se incluyen meses posteriores a enero de 2020**

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media

Una prueba de que debe tenerse en cuenta estos meses como distorsionadores está en la proyección del PIB predicho a partir del modelo estimado hasta 2019. Al analizar únicamente la información disponible hasta diciembre de 2019, se observa en la figura 8 que, además de ser una buena aproximación para períodos pasados, el nuevo modelo proporciona estimaciones muy precisas durante los meses críticos de la pandemia, así como los posteriores.

**Figura 8. Comparación del PIB observado y el PIB predicho por el modelo teniendo en cuenta solo los meses hasta diciembre de 2019.**

Gráfico, Gráfico de líneas

Descripción generada automáticamente

#### Búsqueda iterativa del mejor modelo

Para solucionar los problemas planteados mientras se escogen los indicadores que optimizan la capacidad de reproducir la evolución del PIB se propone una estrategia doble. En primer lugar, y como se ha adelantado, la estimación del modelo va a realizarse hasta diciembre de 2019 (grupo de entrenamiento), dejando los datos de 2020 y posteriores fuera del cálculo del mismo. En segundo lugar, para seleccionar los mejores indicadores no se utilizarán estadísticos tradicionales de bondad de ajuste para el conjunto de los datos, sino métricas aplicadas para las predicciones posteriores a diciembre de 2019 (grupo de prueba). Este proceso de selección sigue, así, los pasos iterativos que se describen a continuación:

1. En primer lugar, el proceso se caracteriza por buscar entre los modelos predictivos aquél que optimiza una métrica definida. Concretamente, de los indicadores disponibles se estiman regresiones con diferentes combinaciones posibles entre ellos, con un mínimo de 3 indicadores y un máximo de 11. Este enfoque de búsqueda exhaustiva de combinaciones de regresores trata de encontrar la mejor combinación posible de variables para predecir el PIB. Esto puede mejorar la precisión de los modelos predictivos al identificar las variables más influyentes en la economía de Andalucía.

2. Se define una evaluación exhaustiva de las diferentes combinaciones. Al considerar una amplia gama de combinaciones de regresores (un total de 8086 combinaciones evaluadas), el proceso permite analizar sistemáticamente cómo diferentes conjuntos de variables afectan la capacidad del modelo para predecir el PIB. Esto proporciona una visión integral y detallada de la importancia relativa de cada variable en la predicción económica.

3. Este proceso permitiría, aunque parcialmente, la identificación de relaciones ocultas. Explorar múltiples combinaciones de variables puede revelar relaciones no evidentes entre diversos aspectos económicos y el PIB. Esto puede desvelar conexiones subyacentes que, de otro modo, podrían pasarse por alto al analizar individualmente cada variable. Por ejemplo, variables con escasa correlación con el PIB pero que podrían explicar una parte no evidente de su evolución y que quedaría oculta por la capacidad predictiva de otros indicadores con mayor correlación con la variable objetivo.

El proceso de evaluación propuesto en este trabajo tiene similitudes con un enfoque de aprendizaje automático, específicamente con técnicas de selección de características y ajuste de modelos. Esto es así porque, en primer lugar, explora de forma sistemática las características que mejor predicen la evolución del PIB. Esto es similar a la selección de las características que se desarrollan en técnicas de aprendizaje automático. Así, este proceso busca determinar las mejores variables (regresores) que influencian la variable objetivo (PIB). Explora múltiples combinaciones para identificar las más relevantes en la predicción.

En segundo lugar, evalúa los modelos de forma similar a como se llevan a cabo en los algoritmos de aprendizaje automático. En este caso, con modelos de regresión lineal, se utilizan métricas de desempeño como el Error Cuadrático Medio (MSE). El objetivo es encontrar el modelo con el mejor desempeño predictivo en función de las características seleccionadas.

En tercer lugar, la optimización del modelo se diseña mediante la búsqueda de la combinación óptima de características que minimiza el error en la predicción del PIB. Esto se asemeja a la optimización de hiperparámetros en el aprendizaje automático, donde se busca maximizar el rendimiento del modelo. Finalmente, el enfoque iterativo prueba múltiples configuraciones y se realiza un seguimiento de los resultados para mejorar la precisión del modelo.

Así pues, aunque no emplea algoritmos de aprendizaje automático avanzados, este proceso comparte aspectos fundamentales con las metodologías utilizadas en el aprendizaje automático para seleccionar y evaluar variables predictoras en función de su capacidad para explicar o predecir un resultado específico.

Para ello se ha desarrollado un código que lleva a cabo los siguientes pasos:

1. Primero se definen las variables y conjuntos de fechas: Se inicia definiendo una lista de 13 regresores que podrían influir en el PIB y que se corresponde con la batería de indicadores seleccionada por el IECA. Luego, se establecen fechas para dividir los datos en conjuntos de entrenamiento y prueba. El entrenamiento es hasta diciembre de 2019. La prueba, todos los meses posteriores.

2. En segundo lugar, se procede a buscar la mejor combinación de regresores. Así, el código utiliza bucles anidados para iterar a través de todas las combinaciones posibles de regresores, desde combinaciones de 3 hasta 11 regresores.

3. Se entrena y evalúan los modelos. Para cada combinación de regresores, se entrena (estima) un modelo de regresión lineal con los datos de entrenamiento y se evalúa su desempeño utilizando los datos de prueba. Se calcula el Error Cuadrático Medio (MSE) para medir la precisión de las predicciones del PIB.

4. En cuarto lugar, se selecciona el mejor modelo. Durante el proceso de iteración, se realiza un seguimiento del modelo con el MSE más bajo, actualizando el "mejor modelo" y la combinación de regresores que proporcionan la mejor predicción.

5. Pero, además del MSE, se calcula una métrica adicional que evalúa la concordancia en el cambio de dirección de las predicciones del PIB en comparación con los datos reales. No solo nos interesa si se aproxima bien, sino también si la proyección genera predicciones que captan adecuadamente los cambios en los signos de las tasas de crecimiento del PIB a lo largo de los trimestres que se evalúan como período de prueba.

El proceso busca encontrar la combinación óptima de variables que minimiza el MSE y maximiza la precisión en la predicción de la dirección del PIB, contribuyendo así a identificar los factores más relevantes que impactan en la economía de Andalucía.

Siguiendo este proceso, la combinación de indicadores que minimiza el error cuadrático medio para los meses posteriores a diciembre de 2019 Importaciones de bienes, Consumo de gasolina y gasóleo, ICMA General sin estaciones de servicios y Cifra negocios del sector servicios. Con solo cuatro indicadores es posible maximizar la capacidad de información para describir la evolución de la actividad económica andaluza similar al PIB.

Si realizamos la regresión entre el POIB y estos cuatro indicadores se obtiene el siguiente resultado:

Tabla

Descripción generada automáticamente

En la regresión se puede observar dos cuestiones. En primer lugar, los signos de los coeficientes son los espetados, aunque el de importaciones no podríamos considerarlo como significativo. En segundo lugar, la capacidad explicativa del modelo es elevada, pudiendo explicar el 90% de la variabilidad del PIB a lo largo de su serie completa.

Finalmente, la serie predicha del PIB por este modelo se muestra bastante coincidente con la observada del PIB. Incluso su capacidad de explicar ya no solo los meses posteriores al inicio del PIB sino los siguientes, es elevada.

**Figura 9 Comparación del PIB observado y el PIB predicho por el modelo teniendo en cuenta solo los trimestres hasta IV de 2019 y utilizando como prueba los posteriores. Mejor modelo**

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Pero podríamos pensar que en este ejercicio los meses de 2020 influyen en la definición del modelo. Para ello se repite el ejercicio estimando el modelo hasta 2019 y minimizando el error entre el primer trimestre de 2021 y el tercero de 2023.

Para este ejercicio, los indicadores seleccionados serían solo tres: Índice de producción agrícola – Andalucía, Consumo de gasolina y gasóleo y Cifra negocios del sector servicios – Andalucía. La capacidad de predicción del modelo solo con estos tres indicadores es muy elevada, del 88,6% para el conjunto del período.

Tabla

Descripción generada automáticamente

Finalmente, la representación gráfica de las series predichas y observadas de PIB muestran que solo estos tres indicadores reproducen el perfil cíclico marcado por el último, con un ajuste muy certero ya no solo para los trimestres de la pandemia sino para los posteriores.

**Figura 10 Comparación del PIB observado y el PIB predicho por el modelo teniendo en cuenta solo los trimestres hasta IV de 2019 y utilizando como prueba los posteriores a IV de 2020. Mejor modelo II.**

Gráfico, Gráfico de líneas

Descripción generada automáticamente

No obstante, esta aproximación que por sí sola nos podría ofrecer pesos que permitieran construir un indicador muy aproximado a la evolución cíclica de la economía andaluza se antoja insuficiente.

En primer lugar, y a pesar de su capacidad de reproducir el ciclo andaluz, el uso de tan escaso número de indicadores nos ofrece la intuición de que hay información relevante que no se está incluyendo. En segundo lugar, cuanto menor sea el número de indicadores usados, más sensible será el indicador sintético a comportamientos erráticos o outliers de los indicadores usados, por muy bien que estos aproximen la evolución histórica o para los trimestres usados como prueba de la serie PIB.

Por estas razones, lo que se propone, como segundo paso en este trabajo, es el de aplicar metodologías de reducción de la dimensión del número de variables o indicadores que pueden ser usados para construir el indicador sintético. La idea, muy simple, es poder extraer de todos los indicadores presentes en la batería aportada por el IECA, toda la información relevante que permita aproximar la evolución del PIB, y con ello de la actividad económica de Andalucía.

Sin embargo, el problema del uso de esta metodología dejaría de lado indicadores con distinta frecuencia. Por esta razón, se propone como segundo ejercicio la aplicación de un modelo dinámico factorial desarrollado por Bańbura y Modugno (2014) y que permite, no solo rellenar en el proceso los datos “missings” sino también, emplear en un mismo proceso homogéneos, indicadores tanto mensuales como trimestrales. Esto es lo que se propone para una segunda fase del trabajo.

## Bibliografía

Bańbura, Marta, and Michele Modugno. “Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data.” Journal of Applied Econometrics 29, no. 1 (2014): 133-160.

1. En este sentido, de momento, se han procedido a usar directamente las series aportadas por el IECA con los procesos de desestacionalización y ajuste de atípicos. [↑](#footnote-ref-1)