# Aplicación y código de selección de mejor modelo

Este código está diseñado para abordar una problemática común en el análisis económico: el modelado de series temporales multivariantes para predecir el comportamiento de una variable económica clave, en este caso el Producto Interno Bruto (PIB). En el contexto económico, las predicciones precisas de variables como el PIB son esenciales para la planificación de políticas públicas, el análisis macroeconómico y las decisiones empresariales. Las variables macroeconómicas suelen estar correlacionadas entre sí, lo que significa que el análisis univariante (es decir, analizar una sola variable a la vez) puede no ser suficiente para capturar toda la información disponible.

Para ello, el código utiliza un enfoque basado en modelos de factores dinámicos, que son útiles para analizar grandes conjuntos de datos donde múltiples variables económicas pueden influir en el comportamiento de una variable objetivo (como el PIB). Este enfoque permite capturar las relaciones subyacentes entre estas variables y sintetizar la información de manera que sea posible obtener predicciones más precisas.

El código incluye diversas etapas de manipulación de datos, transformación para garantizar la estacionariedad de las series temporales, y la búsqueda de combinaciones de variables explicativas (regresores) que maximicen la calidad del modelo predictivo en términos de ajuste (R²) y la estabilidad de las predicciones a lo largo del tiempo (variabilidad).

## Descripción general del proceso

El código se estructura en varias fases fundamentales, que pueden resumirse de la siguiente manera:

1. Carga y preparación de datos: Se leen los datos económicos que se encuentran en una hoja de cálculo Excel. Estos datos incluyen series temporales de indicadores económicos que son tanto mensuales como trimestrales.

2. Transformación de las series para garantizar estacionariedad: La mayoría de los modelos de series temporales requieren que los datos sean estacionarios, es decir, que sus propiedades estadísticas (como la media y la varianza) sean constantes en el tiempo. El código transforma estas series mediante logaritmos y diferenciaciones hasta que todas las series cumplan con esta propiedad.

3. Modelado mediante factores dinámicos: El modelo principal utilizado es el Dynamic Factor Model o modelo de factores dinámicos, que es capaz de capturar relaciones entre muchas variables observadas y un número reducido de factores no observables que las impulsan. El código ajusta este modelo para predecir el PIB a partir de diferentes combinaciones de series temporales económicas.

4. Búsqueda de las mejores combinaciones de variables predictoras: El proceso clave del código es la búsqueda iterativa de las mejores combinaciones de variables mensuales y trimestrales que maximicen una métrica compuesta de la bondad de ajuste (R²) y la estabilidad de las predicciones a lo largo del tiempo. Este proceso de búsqueda tiene en cuenta diferentes factores y configuraciones para encontrar la mejor solución posible.

5. Evaluación y selección del modelo: Finalmente, el código selecciona el modelo que mejor combina precisión en la predicción (medida por el R²) y estabilidad en las proyecciones a través de la minimización de la variabilidad mes a mes.

## Preparación y transformación de los datos

El primer paso importante del código es la carga y preparación de los datos. Aquí, el código lee series temporales mensuales y trimestrales desde un archivo Excel. Este archivo contiene varias variables económicas que se usarán como posibles regresores en el modelo.

Uno de los desafíos más comunes en el análisis de series temporales es que muchas de estas series no son estacionarias. La estacionariedad es una propiedad que se refiere a que las características estadísticas de una serie (como la media y la varianza) se mantienen constantes a lo largo del tiempo. Las series económicas suelen mostrar tendencias (por ejemplo, un aumento general en el número de afiliados a la Seguridad Social) o patrones estacionales (como la variación del turismo a lo largo del año), lo que implica que no son estacionarias.

El código resuelve este problema mediante dos transformaciones clave:

1. Transformación logarítmica: Esto ayuda a estabilizar la varianza y reducir la influencia de valores extremos. Las variables económicas a menudo tienen distribuciones asimétricas, y tomar el logaritmo puede ayudar a normalizar los datos.

2. Diferenciación: Para eliminar tendencias o patrones estacionales, se aplican diferencias sucesivas a las series hasta que se vuelven estacionarias. El orden de las diferenciaciones se realizan mediante el uso de los test de Dickey-Fuller de modo automatizado. Esta diferenciación convierte las series en tasas de cambio en lugar de valores absolutos, lo cual es común en el análisis macroeconómico.

## Modelos de factores dinámicos

El corazón de este análisis es el uso de un modelo de factores dinámicos (Dynamic Factor Model). Estos modelos son herramientas potentes para reducir la dimensionalidad de los datos cuando se tienen muchas variables correlacionadas, como en este caso, donde se analiza una gran cantidad de series temporales económicas.

En esencia, un modelo de factores dinámicos busca identificar un número reducido de factores latentes o no observables que explican la mayoría de la varianza de las variables observadas. Por ejemplo, en lugar de modelar cada una de las series económicas directamente, el modelo busca identificar un conjunto más pequeño de factores subyacentes que afectan a todas las series de forma conjunta. En el contexto del PIB, estos factores podrían representar dinámicas económicas generales como la demanda agregada, la inversión o las exportaciones.

El código utiliza la clase `DynamicFactorMQ` de la biblioteca `statsmodels`, que está diseñada específicamente para manejar datos de frecuencia mixta (por ejemplo, datos mensuales y trimestrales juntos), lo que es ideal para este análisis dado que las variables económicas provienen de diferentes frecuencias temporales.

## Búsqueda de las mejores combinaciones de variables explicativas

Una parte crucial de este código es la búsqueda eficiente de las mejores combinaciones de regresores. No todas las variables disponibles son necesariamente útiles para predecir el PIB, por lo que el objetivo es seleccionar un subconjunto óptimo de variables que maximicen la precisión del modelo.

La búsqueda se realiza de la siguiente manera:

1. Combinaciones de regresores: El código prueba múltiples combinaciones de variables mensuales y trimestrales. Utiliza la función `combinations` de Python para generar todas las combinaciones posibles de un subconjunto de variables, aunque se establece un límite para evitar la explosión combinatoria.

2. Evaluación del modelo: Para cada combinación de variables, el modelo se ajusta y se calcula el R² para el PIB, que mide qué tan bien la combinación de variables elegida explica las variaciones en el PIB.

3. Minimización de la variabilidad: Además de buscar maximizar el R², el código también evalúa la estabilidad de las predicciones del PIB mes a mes, penalizando aquellas combinaciones de variables que producen una alta variabilidad en las predicciones. Esto es crucial en un contexto económico, ya que las predicciones altamente volátiles son menos útiles para la toma de decisiones.

4. Métrica combinada: Para guiar la búsqueda, se utiliza una métrica que combina el R² y la estabilidad de las predicciones. Esta métrica da más peso al R² (es decir, se prioriza la precisión del modelo), pero también tiene en cuenta la variabilidad para evitar modelos demasiado volátiles.

## Optimización del número de factores

El número de factores latentes que el modelo de factores dinámicos utiliza es una decisión importante. Si se eligen muy pocos factores, es posible que el modelo no capture toda la información disponible en las variables observadas. Si se eligen demasiados, el modelo puede sobreajustarse, capturando el ruido en lugar de las verdaderas relaciones subyacentes.

El código prueba diferentes números de factores (limitados entre 1 y 3) y evalúa qué configuración proporciona el mejor balance entre precisión y estabilidad.

## Optimización y criterio de parada

Para evitar que la búsqueda de combinaciones de variables continúe indefinidamente o que el proceso se demore de forma extensa en el tiempo, el código implementa un criterio de paciencia. Este criterio establece un límite en la cantidad de iteraciones que pueden pasar sin que se encuentre una mejora significativa en la métrica combinada. Si después de un número definido de iteraciones no se logra una mejora, el algoritmo se detiene, asumiendo que no se puede mejorar más el modelo actual.

Este enfoque permite que la búsqueda sea eficiente y evita sobrecargar el proceso con combinaciones que probablemente no resultarán en mejoras sustanciales.

## Resultados y selección del mejor modelo

El resultado final del código es la selección del mejor modelo que maximiza el ajuste (R²) y minimiza la variabilidad en las predicciones del PIB. El mejor modelo seleccionado se imprime, junto con las combinaciones de variables mensuales y trimestrales que resultaron ser más efectivas.

Además, el código también imprime el resumen del modelo ajustado, que incluye estadísticas clave como los coeficientes de las variables seleccionadas, los errores estándar y la significancia estadística de los coeficientes. Este resumen es esencial para interpretar los resultados del modelo y entender qué variables económicas tienen mayor impacto en las predicciones del PIB.

## Resumen

Este código implementa un análisis sofisticado de series temporales multivariantes con el objetivo de predecir el PIB utilizando un modelo de factores dinámicos. A lo largo del proceso, se lleva a cabo una cuidadosa selección de variables explicativas, se asegura la estacionariedad de las series temporales, y se busca un equilibrio entre precisión y estabilidad en las predicciones.

La metodología empleada es de gran utilidad en el análisis macroeconómico, donde se dispone de múltiples series económicas que pueden influir en la variable objetivo. La búsqueda de las mejores combinaciones de variables y factores, junto con la penalización de la volatilidad en las predicciones, proporciona un enfoque robusto y eficiente para la modelización de datos económicos.

Este tipo de análisis tiene aplicaciones prácticas no solo para la predicción del PIB, sino también para cualquier otro tipo de variable económica clave, como la inflación, el empleo o el comercio exterior.