# Manual para el código y aplicación de búsqueda eficiente del mejor modelo DFMQ

# Search\_app.py

## 1. Introducción y Objetivo del Código

El presente código ha sido desarrollado con el objetivo de implementar un modelo de factores dinámicos mixto-frecuencias para el análisis y predicción del PIB de Andalucía. Este tipo de modelos resulta especialmente útil cuando se trabaja con variables económicas que se publican con diferentes frecuencias temporales (mensuales y trimestrales), permitiendo aprovechar toda la información disponible para generar predicciones más precisas.

### Descripción General

El código implementa una metodología que combina indicadores económicos mensuales y trimestrales para estimar y predecir la evolución del PIB. El proceso incluye la transformación de las series temporales para garantizar su estacionariedad, la selección óptima de variables explicativas, y la estimación de un modelo que maximiza la capacidad predictiva mientras mantiene la estabilidad de las predicciones.

### Propósito del Análisis

El análisis persigue varios objetivos fundamentales:

* Identificar las variables económicas más relevantes para la predicción del PIB
* Generar un modelo que combine de manera eficiente información de diferente frecuencia temporal
* Obtener predicciones robustas del PIB que aprovechen la información más reciente disponible
* Evaluar la capacidad predictiva del modelo tanto en términos globales como en el período más reciente

### ## Bibliotecas y Dependencias

El código utiliza diversas bibliotecas de Python especializadas en análisis de datos y modelización estadística:

Texto

Descripción generada automáticamente

Cada una de estas bibliotecas cumple un papel específico:

* pandas y numpy proporcionan las estructuras de datos fundamentales y las operaciones matemáticas necesarias
* statsmodels ofrece las herramientas econométricas para el análisis de series temporales y la estimación del modelo
* matplotlib y seaborn se utilizan para la visualización de resultados
* tqdm permite monitorizar el progreso de los procesos de optimización
* Las bibliotecas restantes facilitan la gestión de archivos, fechas y advertencias del sistema

La combinación de estas herramientas permite crear un flujo de trabajo robusto y eficiente para el análisis econométrico propuesto, aprovechando las mejores prácticas en análisis de datos y modelización estadística con Python.

## 2. Preparación y Transformación de Datos

### Transformación de Series Temporales a Series Estacionarias

En el análisis de series temporales económicas, es fundamental trabajar con series estacionarias para garantizar la validez de los modelos econométricos. El código implementa dos funciones especializadas que transforman series temporales en series estacionarias: una para datos mensuales (make\_all\_stationary\_m) y otra para datos trimestrales (make\_all\_stationary\_q).

El proceso de transformación comienza con una conversión logarítmica de todas las series temporales. Esta transformación inicial es crucial ya que ayuda a estabilizar la varianza de las series y reduce el impacto de valores extremos que podrían distorsionar el análisis. Además, la transformación logarítmica tiene la ventaja adicional de linearizar relaciones que originalmente podrían ser exponenciales, facilitando así su posterior análisis.

Para verificar la estacionariedad de las series, ambas funciones emplean el test Augmented Dickey-Fuller (ADF). Este test estadístico es una herramienta robusta que evalúa la presencia de raíces unitarias en la serie temporal. El proceso se realiza de manera iterativa, aplicando el test repetidamente hasta confirmar que la serie se ha convertido en estacionaria, utilizando un nivel de significancia del 5% como criterio de decisión.

La principal distinción entre las dos funciones radica en su método de diferenciación. Para las series mensuales, se aplica una diferenciación de orden 3, lo que significa que se calcula la diferencia entre el valor actual y el valor de hace tres períodos. Esta elección no es arbitraria: está diseñada específicamente para eliminar patrones estacionales trimestrales que pueden estar presentes en los datos mensuales. En contraste, para las series trimestrales, se utiliza una diferenciación de orden 1, calculando simplemente la diferencia entre valores consecutivos, ya que los datos trimestrales generalmente requieren un tratamiento menos intensivo para alcanzar la estacionariedad.

### Lectura y Procesamiento de Datos

El proceso de lectura y tratamiento inicial de los datos es fundamental para el correcto funcionamiento del modelo. El código implementa una serie de pasos sistemáticos para importar y preparar la información contenida en archivos Excel.

**Lectura de Datos desde Excel**

El proceso comienza con la lectura de un archivo Excel que contiene las series temporales económicas. La estructura del código está diseñada para trabajar con un archivo específico que contiene dos hojas de interés:

Texto

Descripción generada automáticamente

La lectura de datos se realiza utilizando pandas:

Texto

Descripción generada automáticamente

**Configuración de Fechas y Formatos Temporales**

El manejo adecuado de las fechas es crucial para el análisis de series temporales. El código implementa un proceso detallado para extraer y formatear la información temporal:

Texto

Descripción generada automáticamente

Para las series temporales, se realiza un procesamiento específico de las fechas:

Texto

Descripción generada automáticamente

**Tratamiento Específico del Período COVID-19**

El código incorpora un tratamiento especial para el período de la pandemia COVID-19 mediante la creación de variables dummy trimestrales. Se crean cuatro variables (D2020Q1, D2020Q2, D2020Q3, D2020Q4) que toman el valor 1 en su trimestre correspondiente del año 2020 y 0 en el resto de períodos. Estas variables se incorporan automáticamente al conjunto de regresores trimestrales durante la estimación del modelo.

**Tratamiento Inicial de los Datos**

Una vez leídos los datos, se aplica un procesamiento inicial para preparar las series:

Texto

Descripción generada automáticamente

Este proceso incluye:

* La aplicación de una media móvil de 3 períodos para suavizar las series mensuales
* La eliminación de las primeras filas que contienen valores NaN resultantes de la media móvil
* La configuración de los índices temporales adecuados para cada tipo de serie

El resultado de este proceso son dos DataFrames (df\_mens y df\_trims) correctamente formateados y preparados para su posterior análisis, con:

* Índices temporales adecuados para cada frecuencia
* Series suavizadas para reducir la volatilidad
* Formatos consistentes para su uso en el modelo

Este procesamiento inicial es crucial ya que establece la base para todos los análisis posteriores y asegura que los datos estén en el formato adecuado para la aplicación del modelo de factores dinámicos.

## 3. Modelado y Análisis

El proceso de modelado y análisis constituye el núcleo central del código, donde se implementan las funciones necesarias para evaluar y seleccionar el mejor modelo predictivo del PIB. Esta sección se divide en tres componentes principales que trabajan de manera coordinada.

### Función de Evaluación de Modelos

La función evaluar\_modelo es responsable de estimar y evaluar cada especificación del modelo de factores dinámicos. La función calcula tres métricas principales:

* El R² global del modelo
* La variabilidad de las predicciones
* El R² específico para el período más reciente (2021Q3 en adelante)

La función incorpora automáticamente las variables dummy de COVID-19 al conjunto de regresores trimestrales antes de la estimación del modelo, aunque se puede seleccionar su eliminación en los procesos de estimación.

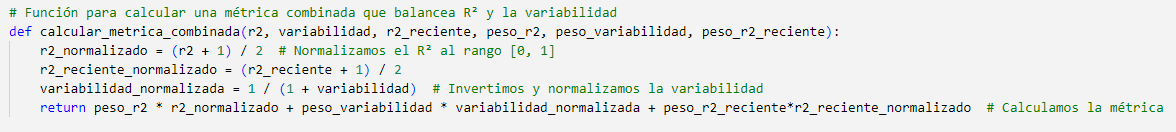
Texto

Descripción generada automáticamente

### Función de Cálculo de Métricas Combinadas

La función calcular\_metrica\_combinada integra diferentes criterios de evaluación para seleccionar el mejor modelo. La función combina tres aspectos clave con sus respectivos pesos:

* peso\_r2 (0.8): para el R² global del modelo
* peso\_variabilidad (0.1): para la medida de variabilidad de las predicciones
* peso\_r2\_reciente (0.1): para el R² específico del período reciente



### Algoritmo de Búsqueda Eficiente del Mejor Modelo

La función busqueda\_eficiente implementa un algoritmo iterativo para encontrar la mejor combinación de variables y parámetros. Los parámetros clave incluyen:

* max\_vars: número máximo de variables a incluir
* paciencia: número de iteraciones sin mejora antes de detener la búsqueda
* peso\_r2, peso\_variabilidad, peso\_r2\_reciente: pesos para diferentes componentes de la métrica combinada
* factores: lista de números de factores a considerar [1, 2, 3]

El algoritmo optimiza considerando tres criterios:

* La capacidad predictiva global (R²)
* La estabilidad de las predicciones (variabilidad)
* La precisión en el período reciente (R² reciente)

Texto

Descripción generada automáticamente

**Parámetros y Criterios de Selección**

Los parámetros clave incluyen:

- `max\_vars`: número máximo de variables a incluir

- `paciencia`: número de iteraciones sin mejora antes de detener la búsqueda

- Pesos para diferentes componentes de la métrica combinada

- Rango de factores a considerar

**Proceso Iterativo de Optimización**

El algoritmo:

1. Comienza con un conjunto mínimo de variables

2. En cada iteración:

- Evalúa la inclusión de nuevas variables

- Prueba diferentes números de factores

- Actualiza el mejor modelo si encuentra una mejora

**Control de Convergencia**

El proceso incluye varios mecanismos de control:

- Contador de iteraciones sin mejora

- Límite máximo de variables

- Umbral mínimo de mejora requerida

La salida del algoritmo proporciona:

- El mejor modelo encontrado

- Las variables seleccionadas (mensuales y trimestrales)

- El número óptimo de factores

- Las métricas de rendimiento finales

Este enfoque integral asegura una selección robusta del modelo, equilibrando:

- Capacidad predictiva global

- Estabilidad de las predicciones

- Precisión en el período reciente

- Parsimonia del modelo

## 4. Visualización y Presentación de Resultados

La fase final del análisis se centra en la presentación clara y efectiva de los resultados obtenidos, proporcionando tanto elementos visuales como métricas cuantitativas que permiten evaluar el rendimiento del modelo seleccionado.

### Generación de Gráficos Comparativos

El código implementa una visualización que permite comparar las predicciones del modelo con los valores reales del PIB:

Texto

Descripción generada automáticamente

Este gráfico permite:

- Visualizar la precisión de las predicciones

- Identificar patrones o desviaciones sistemáticas

- Evaluar el comportamiento del modelo en diferentes períodos

### Presentación de Métricas de Rendimiento

El código proporciona un resumen completo de las métricas de rendimiento del modelo seleccionado:

Texto

Descripción generada automáticamente

Estas métricas incluyen:

- R² global del modelo

- Medida de variabilidad de las predicciones

- R² específico para el período más reciente

Resumen del Modelo Óptimo

El código presenta un resumen detallado de la especificación final del modelo:

Texto

Descripción generada automáticamente

Este resumen proporciona:

- Lista de variables mensuales seleccionadas

- Lista de variables trimestrales incluidas

- Número óptimo de factores dinámicos

- Estadísticas detalladas del modelo

La presentación de resultados está diseñada para proporcionar una visión completa del rendimiento del modelo, combinando:

- Elementos visuales para una comprensión intuitiva

- Métricas cuantitativas para una evaluación objetiva

- Detalles técnicos para usuarios avanzados

Esta información permite:

- Evaluar la calidad general del ajuste

- Identificar las variables más relevantes

- Comprender la estructura del modelo óptimo

- Valorar la fiabilidad de las predicciones

La combinación de estos elementos proporciona una base sólida para la toma de decisiones basada en las predicciones del modelo, así como para la evaluación crítica de sus resultados.

## 5. Consideraciones Técnicas y Notas

El código implementado para el análisis y predicción del PIB mediante un modelo de factores dinámicos mixto-frecuencias requiere tener en cuenta diversos aspectos técnicos y limitaciones que son fundamentales para su correcta utilización e interpretación.

En lo que respecta al manejo de datos faltantes, el código implementa varias estrategias de tratamiento. Durante la transformación de las series temporales, se eliminan automáticamente los valores faltantes que surgen tras aplicar las diferenciaciones necesarias para alcanzar la estacionariedad. Además, para las series mensuales, se aplica una media móvil de tres períodos que ayuda a suavizar la serie y reduce la volatilidad, aunque esto implica la pérdida de las primeras observaciones de la muestra. En el proceso de modelado, las series se filtran automáticamente para el período de estimación especificado, verificando la completitud de los datos antes de proceder a la estimación del modelo.

Los criterios de convergencia implementados en el algoritmo de búsqueda son fundamentales para garantizar resultados robustos. El proceso iterativo se controla mediante varios parámetros: un número máximo de variables a incluir, un criterio de paciencia que determina cuántas iteraciones sin mejora se permite antes de detener la búsqueda, y un umbral mínimo de mejora en la métrica combinada. La evaluación de la estacionariedad se realiza mediante el test ADF con un nivel de significancia del 5%, aplicando diferenciaciones de forma iterativa hasta alcanzar la estacionariedad requerida.

En cuanto a las limitaciones del modelo, es importante señalar que está diseñado específicamente para combinar datos mensuales y trimestrales, no siendo directamente aplicable a otras frecuencias temporales. Las transformaciones aplicadas a los datos, incluyendo logaritmos y diferencias, modifican la interpretación directa de los resultados y son irreversibles, lo que afecta a la interpretabilidad del modelo. Además, la pérdida de observaciones debido a las diferenciaciones reduce la muestra disponible para la estimación.

Desde el punto de vista computacional, el proceso de búsqueda puede ser intensivo y el tiempo de ejecución aumenta significativamente con el número de variables consideradas. La convergencia del algoritmo no está garantizada en todos los casos, lo que requiere un seguimiento cuidadoso del proceso de estimación.

Para la aplicación práctica del modelo, se recomienda realizar reestimaciones periódicas, ya que tanto las variables seleccionadas como la estructura óptima pueden variar con la incorporación de nuevos datos. Las predicciones generadas deben considerarse como estimaciones puntuales, teniendo en cuenta la incertidumbre asociada a las mismas. La calidad de los resultados depende fundamentalmente de la calidad de los datos de entrada.

El modelo presenta diversas áreas de mejora potencial, como la incorporación de intervalos de confianza para las predicciones, la implementación de tests de robustez adicionales y el desarrollo de diagnósticos más detallados. Desde el punto de vista técnico, el código podría optimizarse para mejorar su eficiencia y se podrían implementar métodos de validación cruzada más sofisticados.

Estas consideraciones técnicas son esenciales para garantizar un uso apropiado del modelo y una correcta interpretación de sus resultados, proporcionando una base sólida para su aplicación en el análisis económico y la toma de decisiones.

# APLICACIÓN STREAMLIT PARA LA ESTIMACIÓN DE MODELOS

## Introducción

Esta herramienta ha sido diseñada para permitirte realizar análisis predictivos del PIB combinando datos mensuales y trimestrales. Si es la primera vez que utilizas la aplicación, este manual te guiará paso a paso en el proceso.

## Primeros Pasos

Lo primero que verás al abrir la aplicación es una interfaz limpia y organizada con varias secciones. La aplicación está estructurada de manera secuencial, guiándote desde la carga de datos hasta la ejecución del análisis.

**1. Cargando tus Datos**

El primer paso es cargar el archivo Excel que contiene tus datos. La aplicación espera un archivo con un formato específico:

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Para cargar tu archivo:

* 1. Localiza la sección superior con el texto "Drag and drop file here"
  2. Puedes arrastrar tu archivo directamente a esta zona o hacer clic en "Browse files"
  3. El archivo debe ser formato .XLSX y no superar los 200MB
  4. Una vez cargado, verás el nombre del archivo y su tamaño

**2. Definiendo el Período de Análisis**

Después de cargar los datos, encontrarás dos campos para definir el período de análisis:

Imagen que contiene Icono

Descripción generada automáticamente

* La fecha de inicio predeterminada es 2002/01/01
* La fecha de fin predeterminada es 2024/09/30
* Puedes ajustar estas fechas según tus necesidades, pero recuerda mantener un período suficientemente largo para que el análisis sea robusto

**3. Selección de Variables**

La siguiente sección te permite elegir qué variables incluir en tu análisis. Está dividida en dos partes:

Variables Mensuales

* Verás un contador que muestra cuántas variables mensuales están seleccionadas
* Utiliza el menú desplegable para descartar las variables que no desees incluir
* Cada vez que selecciones una variable para descartar, el contador se actualizará

Variables Trimestrales

* Similar a las mensuales, verás un contador de variables trimestrales
* Usa el menú desplegable para descartar variables
* El PIB siempre estará incluido como variable trimestral principal

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

**4. Configuración del Modelo**

Esta es la parte más técnica del proceso, pero la hemos simplificado para que sea intuitiva:

Imagen que contiene Gráfico de líneas

Descripción generada automáticamente

Variables dummy COVID-19

* Marca esta casilla si quieres que el modelo tenga en cuenta el impacto específico de la pandemia
* Recomendamos mantenerla activada para obtener resultados más precisos

Número de Factores

* Usa el control deslizante para seleccionar entre 1 y 5 factores
* El valor predeterminado es 3, que suele funcionar bien en la mayoría de los casos
* Un mayor número de factores puede capturar más patrones pero también puede hacer el modelo más complejo

Ajuste de Pesos

Esta es una parte crucial del modelo. Necesitas configurar tres pesos diferentes:

* Peso del R²: Determina la importancia del ajuste global del modelo
* Peso de la variabilidad: Controla cuánto valoras la estabilidad de las predicciones
* Peso del R² reciente: Define la importancia del ajuste en el período más reciente

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

\*\*IMPORTANTE\*\*: La suma de los tres pesos debe ser exactamente 1.00

**5. Ejecutando el Análisis**

Una vez configurado todo:

1. Revisa que todos los parámetros estén como deseas
2. Asegúrate de que la suma de los pesos sea 1.00
3. Haz clic en el botón "Ejecutar análisis"
4. Espera mientras la aplicación procesa los datos (esto puede tomar unos minutos)

### Consejos y Solución de Problemas

**Consejos para Mejores Resultados**

- Comienza con los valores predeterminados y ajusta gradualmente

- Mantén un período de análisis suficientemente largo

- Si incluyes las variables dummy de COVID-19, considera su impacto en tus resultados

### Problemas Comunes y Soluciones

Si encuentras alguno de estos problemas:

1. Error al cargar el archivo

- Verifica que sea formato .XLSX

- Comprueba que no supere 200MB

- Asegúrate de que tiene las hojas correctas

2. Error en los pesos

- Revisa que sumen exactamente 1.00

- Ajusta los controles deslizantes con cuidado

3. Tiempo de procesamiento muy largo

- Reduce el número de variables seleccionadas

- Considera un período de análisis más corto

- Verifica que tu archivo no sea demasiado grande

4. Resultados inesperados

- Revisa la selección de variables

- Comprueba los pesos asignados

- Verifica el período de análisis

Es posible que algunos modelos no logren estimarse cuando se busca el óptimo. No te preocupes, es uno de entre cientos que se estiman.