**Paper de Hyndman et al. (2011)**

El artículo “Optimal combination forecasts for hierarchical time series” (Hyndman, Ahmed, Athanasopoulos & Shang, 2011) aborda el problema de pronosticar series temporales jerárquicas, es decir, conjuntos de series que pueden organizarse en distintos niveles de agregación.

Ejemplos:

* Ventas totales → por región → por ciudad.
* Empleo total → por sector → por ocupación.
* En el caso de precios: IPC total → por grupo → por clase → por subclase → por artículo.

El desafío metodológico es que, cuando se pronostican las series de manera independiente, los resultados de niveles bajos no suman exactamente a los niveles superiores. Esto genera problemas de coherencia interna.

Metodología Propuesta:

El artículo propone un método de combinación óptima de pronósticos (optimal combination forecasts), que mejora los enfoques tradicionales:

* Top-down: pronosticar el total y desagregarlo hacia abajo usando proporciones históricas.
* Bottom-up: pronosticar cada serie desagregada y sumarlas hacia arriba.
* Middle-out: mezcla de los anteriores.

La propuesta consiste en:

1. Generar pronósticos base para todas las series en todos los niveles (ej. con ARIMA, ETS u otro modelo).
2. Reconciliar los pronósticos mediante un modelo de regresión restringida, asegurando que:
   * Los artículos sumen a la subclase.
   * Las subclases sumen a la clase.
   * Las clases sumen al grupo, y así hasta el total.
3. La reconciliación se hace con un estimador de OLS,que resulta ser óptimo bajo supuestos razonables.
   * Fórmula clave:

donde son los pronósticos independientes, es la matriz de agregación, y los pronósticos reconciliados.

Resultados del paper:

* En simulaciones, la combinación óptima es más precisa que top-down y bottom-up.
* En un caso aplicado al turismo australiano, mostró mejoras importantes en niveles intermedios y bajos de la jerarquía.

**Nuestro problema:**

Los precios minoristas y el IPC del DANE tienen una estructura similar de jerarquización:

* La jerarquía de series es: total IPC → grupos → clases → subclases → artículos.
* Dispones de ponderaciones hasta el nivel de subclase, pero no hasta el nivel de artículo.
* Los pronósticos deben ser consistentes a lo largo de la jerarquía (coherencia agregada).

Incorporación de la metodología de Hyndman:

1. Coherencia:
   * Cuando se generan pronósticos independientes por artículo y por subclase, los resultados rara vez son consistentes: la suma de los artículos no coincide con la subclase, ni la suma de las subclases coincide con el total del IPC.
   * La metodología de combinación óptima de Hyndman resuelve este problema aplicando una reconciliación estadística. De esta manera, los pronósticos finales garantizan que:

Artículos    →    Subclase    →    Clase    →    Grupo    →    Total IPC

se mantengan coherentes en toda la jerarquía.

* + Las ponderaciones detalladas de artículo no están disponibles y, sin este paso, habría discrepancias frente a las ponderaciones oficiales publicadas por el DANE en niveles superiores.

1. PCA como proxy de ponderaciones de artículo
   * El PCA permite extraer factores comunes de variación entre los precios de los artículos de una misma subclase, y sus cargas factoriales indican la importancia relativa de cada artículo en la dinámica de esa subclase.
   * Al normalizar estas cargas (para que sumen 1), se obtienen ponderaciones estadísticas flexibles que actúan como los pronósticos base de cada artículo dentro de su subclase.
2. Integración con la jerarquía oficial del DANE:
   * La clave es que la metodología de Hyndman no reemplaza los pesos oficiales disponibles, sino que los integra:
     1. El DANE publica ponderaciones a nivel de subclase.
     2. Los resultados del PCA generan señales relativas entre artículos.
     3. El método de combinación óptima usa la matriz de agregación para que las proporciones generadas por PCA se escalen y respeten exactamente el peso oficial de la subclase.

**Forecasting Using Principal Components from a Large Number of Predictors Stock & Watson (2002)**

Este artículo plantea un problema central en la predicción macroeconómica: se dispone de miles de series económicas que podrían ser útiles como predictores, pero los modelos tradicionales (como VAR o ARIMA) solo permiten trabajar con unas pocas variables a la vez. La solución propuesta por Stock & Watson es el uso de un modelo de factores aproximado, donde toda la información contenida en muchas series puede resumirse en unos pocos factores comunes.

El procedimiento es el siguiente:

1. Se estima un conjunto reducido de factores no observados mediante PCA aplicados a un panel amplio de series.
2. Estos factores se utilizan luego como regresores en modelos de predicción de la variable de interés (ejemplo: inflación o crecimiento).
3. Los autores demuestran que, bajo supuestos razonables, los factores estimados por PCA son consistentes, y que los pronósticos basados en ellos se acercan a los óptimos, incluso cuando el número de series predictoras es mucho mayor que el número de observaciones temporales.

En términos simples: el PCA permite aprovechar la información de cientos de series al mismo tiempo, eliminando ruido idiosincrático y capturando solo la dinámica común relevante para los pronósticos.

**Nuestro problema:**

* En nuestro caso, el panel de predictores no son miles de series macroeconómicas, sino los precios de artículos dentro de cada subclase del IPC.
* El PCA nos permite sintetizar esa gran cantidad de precios en factores representativos, que capturan la tendencia común de los artículos de una subclase (por ejemplo, el movimiento conjunto de frutas).
* Estos factores se convierten en un proxy estadístico de las ponderaciones relativas entre artículos.
* Al usarlos como insumo de modelos de pronóstico, replicamos la lógica de Stock & Watson: en vez de seleccionar arbitrariamente unos pocos artículos, aprovechamos toda la información disponible y la condensamos en factores estadísticamente relevantes.