

# Extraction de facteurs communs par filtre de Kalman : Application à des données macroéconomiques françaises

Projet d'Économétrie Financière

Morgan Jowitt, Pierre Liberge, Aya Mokthar, Gaétan Dumas

Mai 2025

## Abstract

Ce projet vise à reproduire et illustrer les résultats méthodologiques présentés dans l'article "Factor extraction using Kalman Filter and smoothing" de Poncela, Ruiz et Miranda (2021), en les appliquant à un nouveau jeu de données macroéconomiques françaises. L'article propose une méthode robuste pour extraire des facteurs communs latents à partir de séries temporelles multivariées, à travers un Modèle à Facteurs Dynamiques (DFM) associé à un filtre et un lisseur de Kalman. À l'aide d'une implémentation en Python, nous appliquons cette approche à un panel macroéconomique français couvrant la période 1975-2025, et comparons ses résultats à ceux obtenus par une Analyse en Composantes Principales (ACP). Nous démontrons que, malgré sa nature statique, l'ACP offre dans ce cas des corrélations plus fortes avec les variables observées que le modèle dynamique fondé sur le filtre de Kalman.

# 1 Introduction

Dans un contexte économique où les décideurs doivent suivre un grand nombre d'indicateurs macroéconomiques, les modèles à facteurs dynamiques (DFM) permettent de résumer efficacement l'information en extrayant quelques facteurs communs non observés. L'article "Factor Extraction using Kalman Filter and Smoothing" (Poncela et al., 2021) propose une méthodologie complète pour extraire ces facteurs en combinant une représentation en espace d'état, le filtre de Kalman et l'algorithme EM pour estimer les paramètres du modèle.

L'objectif de ce projet est de reproduire la méthodologie de l'article sur un nouveau jeu de données économiques françaises. Pour cela, nous suivons les mêmes étapes: préparation des données, estimation des paramètres par maximum de vraisemblance via l'algorithme EM, extraction des facteurs à l'aide du filtre et lisseur de Kalman, puis comparaison avec l'approche par composantes principales (PCA).

## 2 Modèle DFM et Filtrage de Kalman

Le modèle dynamique à facteur unique est représenté par :

$$Y_t = \Lambda F_t + \varepsilon_t, \quad F_t = \Phi F_{t-1} + \eta_t$$

où :

- $Y_t$  est un vecteur de variables observées,
- $F_t$  est un facteur latent,
- $\Lambda$  est la matrice de chargement,
- $\varepsilon_t \sim \mathcal{N}(0, H)$  et  $\eta_t \sim \mathcal{N}(0, Q)$  sont les bruits.

## 3 Méthodologie

Pour reproduire l'approche du papier, nous avons appliqué plusieurs méthodes d'extraction de facteurs à un panel mensuel de cinq indicateurs macroéconomiques français entre 1996 et 2023 : climat des affaires, taux de chômage (mensualisé), indice des prix à la consommation, production industrielle et taux d'intérêt à 10 ans.

Les données ont été fusionnées, normalisées, et nettoyées (traitement des valeurs manquantes par interpolation et remplissage, détection des valeurs extrêmes via la méthode IQR). Ces étapes assurent la qualité et la comparabilité des séries.

L'objectif est de résumer l'évolution conjointe de ces variables à travers un ou plusieurs facteurs latents, en comparant trois approches :

- **DFM (Dynamic Factor Model)** via `statsmodels`, utilisant une structure en espace d'état et le filtre de Kalman.
- **DFM personnalisé**, codé avec `MLEModel`, pour plus de flexibilité sur les paramètres internes et les méthodes d'optimisation.
- **ACP (Analyse en Composantes Principales)**, méthode de référence non dynamique permettant une estimation rapide des facteurs.

Les facteurs extraits sont ensuite comparés entre méthodes et analysés économiquement via leur corrélation avec les variables observées.

## 4 Résultats Empiriques

### 4.1 Performance des modèles

	Kalman (DFM)	ACP
Log-vraisemblance	-1867.64	-
AIC	3773.28	-
Variance expliquée (facteur 1)	-	52.1%
Corrélation moyenne (facteur 1)	0.548	0.676

Table 1: Comparaison des performances des méthodes sur 604 observations (1975–2025)

### 4.2 Visualisation des facteurs

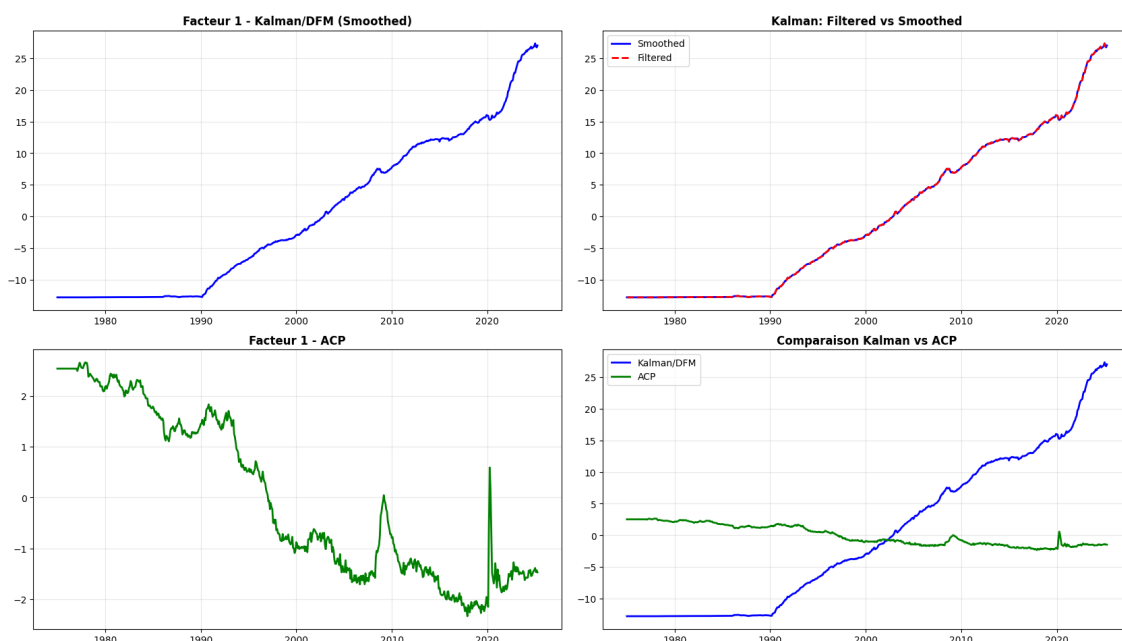


Figure 1: Comparaison des facteurs extraits (Kalman Smoothed & Filtered vs ACP)

Lecture de la figure :

- **Haut gauche** : Le facteur 1 extrait via Kalman (smoothed) montre une forte dynamique monotone croissante sur plus de trois décennies, traduisant une tendance structurelle longue.

- **Haut droite** : Le facteur lissé (bleu) et le facteur filtré (rouge pointillé) coïncident fortement, ce qui suggère un bon ajustement du modèle.
- **Bas gauche** : Le facteur 1 de l'ACP est centré, plus volatil, avec des pics nets autour de 2009 et 2020, cohérents avec les crises financières et COVID.
- **Bas droite** : La comparaison directe montre un décalage structurel : Kalman suit une trajectoire de niveau tandis que l'ACP reste stationnaire autour de zéro, mais reflète mieux les chocs de court terme.

### 4.3 Corrélations avec les variables observées

Variable	Kalman (Facteur 1)	ACP (Composante 1)
Indice des prix à la consommation	+1.000	+0.875
Taux à 10 ans	-0.915	+0.952
Production industrielle	+0.513	+0.772
Taux de chômage	+0.263	+0.515
Climat des affaires	+0.048	+0.267

Table 2: Corrélations des variables avec le facteur Kalman et la composante ACP

#### Corrélation entre les deux approches :

Le facteur Kalman (1) et la composante ACP (1) sont fortement corrélés entre eux (0,879), ce qui indique qu'ils capturent une dynamique latente similaire, bien que de forme différente.

#### Interprétation :

Le premier facteur extrait via le filtre de Kalman est fortement corrélé à l'indice des prix à la consommation (corrélation parfaite) et négativement au taux d'intérêt à 10 ans, ce qui suggère qu'il capte principalement les tensions inflationnistes et les réponses de politique monétaire. En revanche, le premier facteur issu de l'ACP présente des corrélations plus équilibrées avec l'ensemble des variables macroéconomiques, notamment la production industrielle, les taux d'intérêt, et l'inflation, traduisant davantage une dynamique de cycle économique global.

#### Recommandation finale :

L'analyse en composantes principales apparaît ici statistiquement plus performante, avec une meilleure corrélation moyenne et une part de variance expliquée plus élevée. Elle constitue donc un meilleur outil de synthèse conjoncturelle. Le modèle à facteurs dynamiques (Kalman/DFM) conserve néanmoins un intérêt analytique en tant qu'indicateur latent des pressions inflationnistes, utile dans une optique de modélisation structurelle ou de prévision ciblée.

## 5 Conclusion

Dans notre étude, le filtre de Kalman n'a pas surpassé l'ACP en termes de performance explicative, malgré ses avantages théoriques. L'ACP s'est révélée plus efficace pour résumer l'information globale des séries, avec des corrélations plus homogènes et une variance totale expliquée supérieure. Toutefois, le modèle dynamique à facteurs, estimé à l'aide du filtre de Kalman, permet une lecture structurelle plus approfondie : le facteur principal extrait reflète clairement les dynamiques d'inflation et de taux longs, et permet d'isoler plus finement les tendances de fond. Ce cadre reste donc pertinent dans une perspective de modélisation temporelle avancée, notamment pour capturer des régimes économiques latents ou intégrer des contraintes structurelles. Des extensions telles que l'utilisation de plusieurs facteurs, l'estimation bayésienne ou l'introduction de changements de régime pourraient encore enrichir l'interprétation et la portée du DFM.

## 6 Pistes d'amélioration

Ce projet a permis de répliquer une partie de la méthodologie de l'article, mais plusieurs dimensions essentielles restent à explorer. Tout d'abord, l'algorithme EM n'a pas été implémenté dans sa version complète : nous avons estimé les paramètres du modèle par maximum de vraisemblance directe, sans passer par les étapes itératives (E et M) décrites dans l'article. *L'intégration d'un EM complet constituerait une amélioration importante*, permettant une meilleure convergence et une estimation plus rigoureuse des matrices de covariance. Ensuite, nous avons limité notre analyse à un modèle à deux facteurs, en supposant des erreurs non autocorrélées et indépendantes entre séries. *Or, cette hypothèse peut s'avérer restrictive dans le cas de panels plus larges*. L'article recommande ainsi d'explorer des spécifications plus flexibles, incluant des dynamiques plus riches (par exemple, des VAR(p)), ou encore des erreurs idiosyncratiques autocorrélées et/ou corrélées entre elles.

Par ailleurs, la taille du panel utilisé reste modeste (cinq séries). *Une application plus conforme à l'approche proposée par les auteurs impliquerait l'utilisation d'un jeu de données plus étendu* ( $N \geq 20$ ), rendant l'estimation plus complexe mais aussi plus informative. Enfin, notre modèle suppose implicitement que toutes les séries sont stationnaires ou ont été stationnarisées. *Cela constitue une autre limite potentielle*, dans la mesure où de nombreuses séries macroéconomiques présentent des racines unitaires ou des tendances stochastiques. Une version non stationnaire du DFM, capable de capturer des facteurs intégrés ( $I(1)$ ), constituerait ainsi une extension naturelle à explorer.