



Master Probabilités et Finance

# Analyse de la Structure du Marché Actions avec Analyse en composantes indépendantes

Allocation optimale et arbitrage multi-asset

**Auteurs :** Renaud POUH / Philippe YAO

**Encadrant :** Jean Gabriel ATTALI

Année Universitaire : 2024-2025

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Commentaire de l'article</b>	<b>3</b>
2.1	Objectifs et contributions de l'article . . . . .	4
2.2	Aspects mathématiques de l'ICA . . . . .	4
2.3	Pondération, tri et seuillage des composantes indépendantes . . . . .	5
2.4	Résultats clés . . . . .	7
2.5	Limites et perspectives de l'étude . . . . .	7
2.6	Intérêt pour notre étude . . . . .	7
<b>3</b>	<b>Données et prétraitement</b>	<b>8</b>
<b>4</b>	<b>Résultats de l'Analyse en Composantes Indépendantes</b>	<b>9</b>
4.1	Structure des Facteurs Indépendants . . . . .	9
4.2	Reconstruction des prix à l'aide des composantes indépendantes (ICs) . . .	11
4.2.1	Reconstruction de l'indice du CAC 40 par ICA sans seuillage . . . .	11
4.2.2	Reconstruction des prix avec seuillage des composantes indépendantes	12
4.3	Comparaison des résultats entre ICA et PCA . . . . .	14
<b>5</b>	<b>Limites et perspectives</b>	<b>16</b>
<b>6</b>	<b>Conclusion</b>	<b>17</b>
<b>7</b>	<b>Référence de l'article étudié</b>	<b>18</b>

# Table des figures

1	Distribution des log-rendements du CAC 40 et de quelques entreprises com- posant l'indice . . . . .	9
2	5 premières comoosantes (ICs) sans pondération . . . . .	10
3	5 premières composantes (ICs) après pondération . . . . .	10
4	Reconstruction avec les 2 ICs dominantes . . . . .	11
5	Reconstruction avec les 5 ICs dominantes . . . . .	11
6	Reconstruction des prix (seuil = 0.05) . . . . .	12
7	Reconstruction des prix (seuil = 0.01) . . . . .	12
8	Log-rendements dominants (seuil = 0.05) . . . . .	13
9	Log-rendements dominants (seuil = 0.01) . . . . .	13
10	Reconstruction des prix (seuil = 0.001) . . . . .	14
11	Log-rendements dominants (seuil = 0.001) . . . . .	14
12	Reconstruction des prix (Top 5 PCs) . . . . .	15
13	Log-Rendements reconstruits (Top 5 PCs) . . . . .	15
14	Reconstruction des prix avec seuillage (seuil = 0.1) . . . . .	15
15	Log-Rendements reconstruits avec seuillage (seuil = 0.1) . . . . .	15

## Liste des tableaux

1	Résultats des tests de normalité de Kolmogorov-Smirnov et Shapiro-Wilk sur les log-rendements . . . . .	8
---	--	---

# 1 Introduction

L'analyse factorielle est un outil essentiel en finance pour identifier les sources de variabilité des rendements et optimiser la gestion des risques. Traditionnellement, l'*Analyse en Composantes Principales* (PCA) est utilisée pour extraire les facteurs dominants en maximisant la variance expliquée. Toutefois, la PCA ne garantit pas l'indépendance des facteurs, limitant leur interprétation économique, en particulier lorsqu'il s'agit d'associer des facteurs à des phénomènes économiques distincts.

Face à ces limites, l'*Analyse en Composantes Indépendantes* (ICA) propose une approche alternative en identifiant des facteurs véritablement indépendants, potentiellement plus représentatifs de véritables moteurs économiques ou financiers sous-jacents, tels que des chocs exogènes, des dynamiques sectorielles ou des effets macroéconomiques. Appliquée au marché japonais dans l'article "*A First Application of Independent Component Analysis to Extracting Structure from Stock Returns*", elle a révélé des structures non détectables par la PCA, mettant en avant des facteurs rares mais à fort impact, comme des crises ou événements exceptionnels.

Avant de transposer cette approche au marché français, nous commencerons par une première section consacrée à un commentaire approfondi de cet article. Cette analyse nous permettra de mieux comprendre les apports théoriques et empiriques de l'ICA dans un contexte financier, et de dégager les éléments méthodologiques clés que nous chercherons à reproduire et adapter à l'étude du **CAC 40**.

Dans ce projet, nous appliquons l'ICA aux rendements des actions du **CAC 40**. Cette décennie riche en événements (explosion de la bulle internet, crise financière de 2008) offre un terrain d'étude idéal pour tester la capacité de l'ICA à isoler des facteurs pertinents et indépendants. Nous comparons également ses performances à celles de la PCA, afin d'évaluer le compromis entre qualité de reconstruction et interprétabilité des facteurs.

Ce rapport est structuré comme suit :

- La **Section 2** présente un commentaire détaillé de l'article de Back et Weigend (1997).
- La **Section 3** décrit les données et la méthodologie employée.
- La **Section 4** analyse et compare les résultats de la PCA et de l'ICA.
- La **Section 5** discute des limites de l'approche et des perspectives.
- La **Section 6** conclut en résumant les principaux enseignements.

## 2 Commentaire de l'article

L'article "*A First Application of Independent Component Analysis to Extracting Structure from Stock Returns*" de Back et Weigend (1997) constitue l'une des premières contributions majeures appliquant l'Analyse en Composantes Indépendantes (ICA) à des séries financières. Leur objectif principal était d'évaluer la capacité de l'ICA à extraire des facteurs statistiquement indépendants à partir des rendements de 28 grandes actions japonaises, sur la période 1986-1989. Cette démarche se situe dans la perspective de dépasser les limites de l'Analyse en Composantes Principales (PCA), largement utilisée jusqu'alors en finance pour des approches de réduction de dimension ou de modélisation factorielle.

## 2.1 Objectifs et contributions de l'article

L'article s'articule autour de deux ambitions principales. D'une part explorer si l'ICA est capable de révéler des structures latentes au sein des rendements boursiers, en isolant des facteurs indépendants interprétables économiquement. D'autre part comparer les performances de l'ICA avec celles de la PCA, notamment en termes de capacité de reconstruction des prix et d'identification des chocs majeurs de marché.

## 2.2 Aspects mathématiques de l'ICA

L'un des apports importants de l'article réside dans la formalisation mathématique du problème de séparation des sources appliqué aux rendements financiers. L'ICA repose sur le modèle suivant :

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (1)$$

où :

- $\mathbf{X}$  est la matrice des rendements observés ( $n$  actifs  $\times T$  jours),
- $\mathbf{S}$  est la matrice des **composantes indépendantes** (ICs) à extraire,
- $\mathbf{A}$  est la **matrice de mélange**, reliant les facteurs indépendants aux rendements des actifs.

L'objectif de l'ICA est d'estimer la matrice inverse  $\mathbf{W}$  pour obtenir :

$$\mathbf{S} = \mathbf{W}\mathbf{X} \quad (2)$$

Dans l'étude de Back et Weigend, il est supposé que la matrice de séparation  $\mathbf{W}$  est l'inverse de la matrice de mélange  $\mathbf{A}$ , c'est-à-dire :

$$\mathbf{W} = \mathbf{A}^{-1} \quad (3)$$

Cette hypothèse repose sur le cadre classique de l'ICA linéaire, où l'on suppose que le système de mélange est inversible et que le nombre de sources indépendantes est égal au nombre de signaux observés.

La séparation optimale vise donc à estimer  $\mathbf{W}$  de manière à retrouver les composantes  $\mathbf{S}$  les plus indépendantes possible à partir des observations  $\mathbf{X}$ . Cette indépendance se mesure notamment via l'annulation des cumulants d'ordre supérieur ou la minimisation de l'information mutuelle entre les composantes.

Par ailleurs, il serait pertinent d'envisager une approche alternative à la méthode algébrique classique de l'ICA, reposant sur les réseaux de neurones. Certaines architectures neuronales permettent d'estimer les composantes indépendantes en minimisant la divergence de Kullback-Leibler (KL) entre la densité jointe des sorties du réseau, correspondant aux ICs estimées, et le produit des densités marginales, qui représente le cas idéal d'indépendance

parfaite. Dans cette approche, le réseau est entraîné pour maximiser l'entropie des sorties tout en réduisant l'information mutuelle, ce qui permet d'aboutir à une séparation efficace des sources. L'utilisation des réseaux neuronaux pour l'ICA présente plusieurs avantages notables. D'une part, ces modèles sont particulièrement adaptés aux grands ensembles de données et aux structures complexes. D'autre part, ils offrent la capacité de traiter des mélanges non linéaires, là où l'ICA classique repose sur l'hypothèse restrictive de linéarité. Enfin, les réseaux de neurones s'intègrent aisément dans des cadres d'optimisation avancés, permettant l'ajout de contraintes supplémentaires pertinentes pour des applications financières.

Enfin, il est important de rappeler que l'ICA repose sur deux hypothèses fondamentales qui conditionnent la validité de ses résultats. La première stipule qu'au plus une composante suit une loi gaussienne, tandis que la seconde impose que le nombre de sources indépendantes soit égal au nombre de signaux observés. Bien que parfois restrictives, ces hypothèses restent généralement compatibles avec les caractéristiques des rendements financiers, qui présentent rarement une distribution parfaitement gaussienne et où le nombre d'actifs correspond souvent aux dimensions du modèle analysé.

## 2.3 Pondération, tri et seuillage des composantes indépendantes

Dans l'article de Back et Weigend (1997), après extraction des composantes indépendantes (ICs) par l'ICA, il est nécessaire d'évaluer l'influence de chaque IC sur les actifs financiers étudiés. Cette influence est mesurée via la matrice de mélange  $A$ . Chaque colonne de  $A$  représente la contribution d'une IC aux rendements des différentes actions.

### Pondération des ICs et lien avec la variable d'intérêt

Dans leur étude appliquée au marché japonais, une attention particulière est portée à l'action de la **Bank of Tokyo-Mitsubishi**, considérée comme un actif clé du portefeuille étudié. L'analyse de la matrice de mélange montre que certaines ICs ont un poids dominant sur cette banque, permettant ainsi de mettre en évidence des facteurs spécifiques liés à son comportement (crises financières, variations sectorielles bancaires, etc.). Cette pondération permet de hiérarchiser les ICs selon leur importance relative vis-à-vis de cette institution.

Cependant, il est intéressant de noter que l'identification des ICs et leur interprétation économique peuvent être plus évidentes lorsque l'analyse porte sur un **indice boursier** complet (tel que le CAC 40 dans notre cas) plutôt que sur un ensemble restreint de valeurs individuelles. En effet, un indice reflète des dynamiques globales de marché, souvent liées à des facteurs macroéconomiques plus facilement identifiables (taux d'intérêt, croissance, inflation, chocs sectoriels, etc.). L'analyse factorielle appliquée à un indice permet donc d'associer plus naturellement certaines ICs à des comportements économiques généraux.

### Tri des ICs par norme

Afin d'ordonner les ICs par importance, un tri est effectué à partir de la norme des colonnes de la matrice de mélange  $A$ . Ce tri permet de quantifier la contribution globale de chaque IC aux rendements des actifs. Les principales normes utilisées sont :

- **Norme  $\ell_1$**  (somme des valeurs absolues) :

$$||A_k||_1 = \sum_{i=1}^n |a_{ik}| \quad (4)$$

- Cette norme mesure l'importance globale d'une IC à travers l'ensemble des actifs.
- **Norme**  $\ell_2$  (norme euclidienne) :

$$||A_k||_2 = \sqrt{\sum_{i=1}^n a_{ik}^2} \quad (5)$$

- Cette norme met en avant les ICs ayant des contributions fortes sur plusieurs actifs.
- **Norme**  $\ell_\infty$  (valeur absolue maximale) :

$$||A_k||_\infty = \max_{1 \leq i \leq n} |a_{ik}| \quad (6)$$

Cette norme identifie les ICs avec des contributions extrêmes sur un actif spécifique.

### Critère mathématique du seuillage

Le seuillage permet de filtrer les petites contributions des ICs jugées peu significatives. Mathématiquement, le seuillage est appliqué lors de la reconstruction des rendements selon :

$$\bar{x}_i(t-j) = \sum_{k=1}^n g(\bar{y}_{ik}(t-j)), \quad j = 0, \dots, N-1 \quad (7)$$

où la fonction de seuillage  $g(u)$  est définie par :

$$g(u) = \begin{cases} u, & \text{si } |u| \geq \xi \\ 0, & \text{si } |u| < \xi \end{cases} \quad (8)$$

avec  $\xi$  le seuil fixé pour éliminer les contributions faibles.

### Commentaire sur le seuillage

Le seuillage a pour principaux objectifs de réduire le bruit statistique en éliminant les composantes de faible intensité, d'alléger la complexité du modèle afin d'en faciliter l'interprétation, et de concentrer l'analyse sur les facteurs majeurs qui présentent une signification économique forte. Cependant, l'efficacité et la pertinence du seuillage diffèrent sensiblement selon la méthode utilisée. Dans le cas de la PCA, le seuillage s'avère souvent bénéfique, car il permet de supprimer les composantes expliquant peu de variance, contribuant ainsi à simplifier le modèle sans perte majeure d'information. En revanche, pour l'ICA, le seuillage doit être appliqué avec une prudence particulière. L'ICA extrait en effet des composantes statistiquement indépendantes, et certaines ICs, bien que de faible amplitude, peuvent néanmoins capturer des dynamiques spécifiques et significatives, telles que des effets sectoriels ou des événements ponctuels. Un seuillage trop strict risquerait alors d'appauvrir le modèle en supprimant des signaux porteurs d'informations pertinentes pour l'analyse économique et financière.

En conclusion, si le seuillage optimise parfois la lisibilité et la performance de la PCA, il doit rester modéré pour l'ICA afin de conserver l'intégralité de la structure indépendante détectée dans les rendements.

## 2.4 Résultats clés

Les résultats obtenus par Back et Weigend mettent en évidence plusieurs éléments majeurs. L'ICA permet d'identifier deux types de facteurs essentiels dans la dynamique des marchés financiers. D'une part, elle isole des chocs rares et intenses, correspondant à des événements exceptionnels tels que le krach de 1987. D'autre part, elle révèle des fluctuations fréquentes et de faible amplitude, qui peuvent être assimilées à des mouvements de bruit ou à de petites corrections de marché. Par ailleurs, la reconstruction des prix à partir d'un nombre limité de composantes indépendantes s'avère efficace, puisqu'elle parvient à capturer les principaux mouvements du marché avec un nombre restreint de facteurs. Comparée à la PCA, l'ICA offre une meilleure capacité de séparation des événements significatifs, là où la PCA tend à lisser les dynamiques et à mélanger les contributions des différents facteurs sous-jacents. Enfin, l'application d'un seuillage aux composantes indépendantes permet de concentrer l'analyse sur les ICs ayant le plus grand impact, ce qui facilite l'interprétation des principaux moteurs du marché en éliminant les composantes les plus faibles ou les moins pertinentes.

## 2.5 Limites et perspectives de l'étude

Bien que pionnière, l'étude de Back et Weigend présente plusieurs limites. Tout d'abord, l'analyse est restreinte à un seul marché, celui du Japon, et couvre une période relativement courte de trois ans, ce qui limite la portée et la généralisation des résultats obtenus. De plus, l'approche adoptée reste purement statistique, sans tentative explicite de relier les composantes indépendantes extraites à des facteurs économiques précis, tels que la politique monétaire ou des crises sectorielles. Par ailleurs, la comparaison avec la PCA se concentre principalement sur la qualité de la reconstruction des prix, sans approfondir la question de la capacité prédictive des deux méthodes ou leur utilité pour anticiper des mouvements futurs du marché.

Cependant, cette étude ouvre de nombreuses perspectives de recherche. L'application de l'ICA à d'autres marchés et périodes historiques permettrait de tester la robustesse et la généralité des résultats. De plus, il serait intéressant d'explorer les liens entre les ICs extraites et des variables macroéconomiques afin de donner un sens économique plus concret aux facteurs identifiés. Enfin, l'ICA pourrait être exploitée dans des cadres pratiques de gestion du risque ou d'allocation d'actifs, en tant qu'outil permettant d'isoler et de suivre des sources spécifiques de risque ou de performance au sein d'un portefeuille.

## 2.6 Intérêt pour notre étude

L'article de Back et Weigend constitue la base méthodologique et conceptuelle de notre projet. Nous cherchons à reproduire et adapter leur approche, cette fois sur le marché français via les actions du CAC 40, sur une période de neuf ans marquée par des crises majeures. Ce travail nous permet ainsi :

- De tester la robustesse de l'ICA dans un contexte géographique et temporel différent.
- D'évaluer la pertinence économique des ICs extraites sur un marché européen.
- De comparer finement les résultats obtenus avec ceux de la PCA, tant en termes de qualité de reconstruction que d'interprétation des facteurs.



### 3 Données et prétraitement

Nous exploitons des données financières issues du **CAC 40**, principal indice boursier français, afin d’analyser la structure du marché à l’aide de l’ICA et de la PCA. Ces données couvrent la période du **2 janvier 2001 au 9 mars 2010**, soit près de **9 ans de données journalières**.

Les données initiales contiennent les **prix de clôture quotidiens** (*Last Price*) des actions composant l’indice CAC 40. Pour mener notre analyse factorielle, nous calculons les **log-rendements journaliers**, définis comme :

$$R_t = \ln \left( \frac{P_t}{P_{t-1}} \right) \quad (9)$$

où  $P_t$  représente le prix de clôture à la date  $t$ .

Avant l’application des méthodes d’analyse, un prétraitement des données est effectué :

- **Gestion des valeurs manquantes** : suppression des jours où certaines actions ont des prix manquants.
- **Calcul des rendements** : transformation des prix en log-rendements quotidiens.

Ce prétraitement est essentiel pour garantir la robustesse des résultats et assurer une comparaison cohérente entre la PCA et l’ICA.

Il est important de préciser qu’à la différence de l’article de Back et Weigend (1997), qui accorde une attention particulière à certaines entreprises spécifiques (notamment la *Bank of Tokyo-Mitsubishi*), notre étude porte principalement sur l’indice du **CAC 40** lui-même. Notre objectif est donc d’analyser des dynamiques de marché globales, liées à des facteurs économiques généraux et macroéconomiques, plutôt que de cibler le comportement d’une entreprise isolée.

L’observation des histogrammes des log-rendements du CAC 40 et de plusieurs entreprises du panel montre des distributions fortement concentrées autour de zéro, avec des queues épaisses et des pics marqués, caractéristiques classiques des séries financières. Ces distributions présentent un *leptokurtisme* prononcé et des asymétries variables selon les titres, ce qui suggère des comportements éloignés de la normalité.

Cette impression visuelle est confirmée par les tests statistiques de normalité de *Kolmogorov-Smirnov* et de *Shapiro-Wilk*, dont les résultats sont présentés ci-dessous. Les p-values extrêmement faibles (largement inférieures à 0.01) permettent de rejeter l’hypothèse de normalité avec un niveau de confiance de 99%. Ces résultats justifient l’utilisation de méthodes adaptées aux distributions non-gaussiennes, comme l’ICA.

Actif	Kolmogorov-Smirnov p-value	Shapiro-Wilk p-value
CAC 40	$3.62 \times 10^{-13}$	$5.62 \times 10^{-32}$
Accor SA	$2.81 \times 10^{-8}$	$1.22 \times 10^{-23}$
Air Liquide	$9.29 \times 10^{-8}$	$2.01 \times 10^{-26}$
Alcatel-Lucent	$3.92 \times 10^{-14}$	$6.15 \times 10^{-35}$
Alstom	$2.26 \times 10^{-19}$	$2.93 \times 10^{-42}$
Arcelor Mittal	$4.71 \times 10^{-12}$	$1.71 \times 10^{-29}$

TABLE 1 – Résultats des tests de normalité de Kolmogorov-Smirnov et Shapiro-Wilk sur les log-rendements

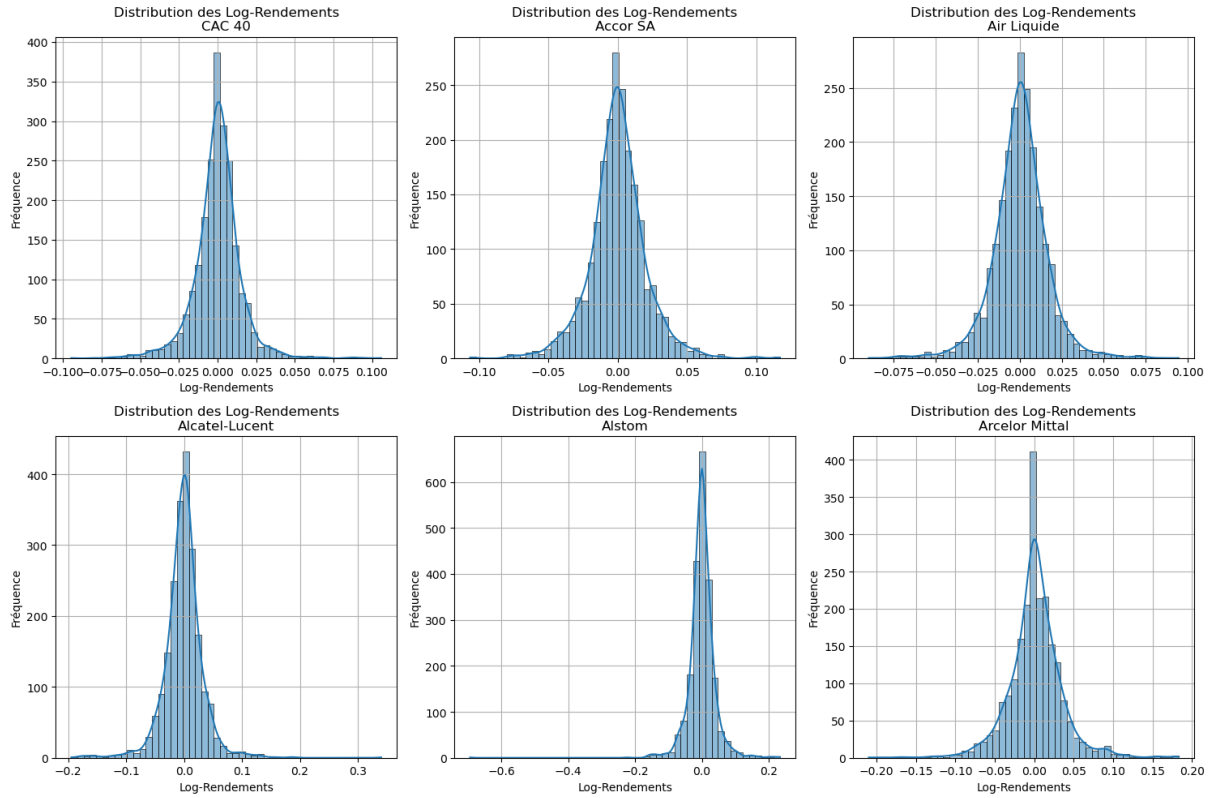


FIGURE 1 – Distribution des log-rendements du CAC 40 et de quelques entreprises composant l'indice

## Tests de stationnarité sur les log-rendements

Afin de vérifier la stationnarité des rendements, indispensable pour l'application correcte de l'ICA, nous avons réalisé le test *Augmented Dickey-Fuller* (ADF). Les résultats du test indiquent des p-values  $< 0.1$  pour l'ensemble des actions testées, ce qui nous permet de conclure à la stationnarité des séries de rendements. Cette propriété est essentielle pour garantir la stabilité statistique des données et assurer la validité des méthodes factorielles appliquées.

## 4 Résultats de l'Analyse en Composantes Indépendantes

### 4.1 Structure des Facteurs Indépendants

L'affichage des principales composantes indépendantes extraites par l'ICA, avant et après pondération, permet d'analyser leur dynamique et leur importance relative. Dans cette étude, la sélection des composantes dominantes a été réalisée à l'aide de la **norme infinie** ( $\ell_\infty$ ), qui retient, pour chaque composante, la contribution maximale parmi les poids de la matrice de mélange  $A$ . Cette approche favorise l'identification des ICs ayant un impact fort, même s'il est localisé sur un actif ou un sous-ensemble réduit.

**Avant pondération :**

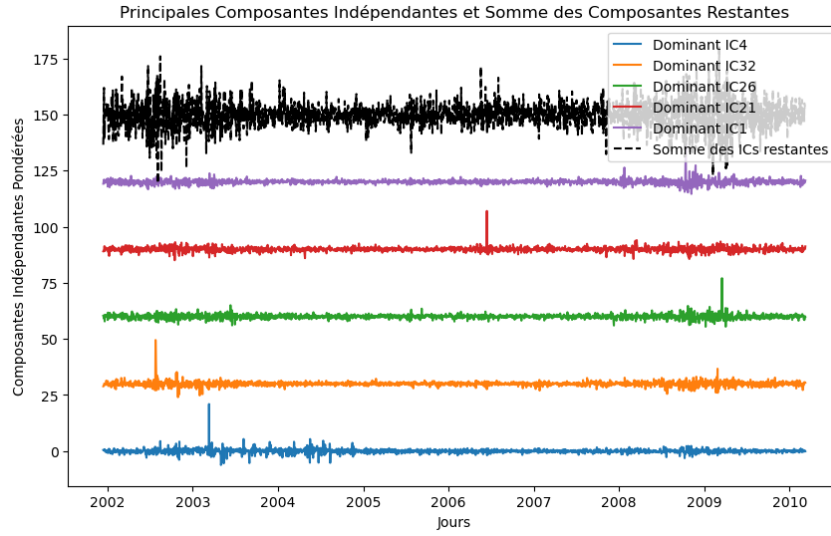


FIGURE 2 – 5 premières comoosantes (ICs) sans pondération

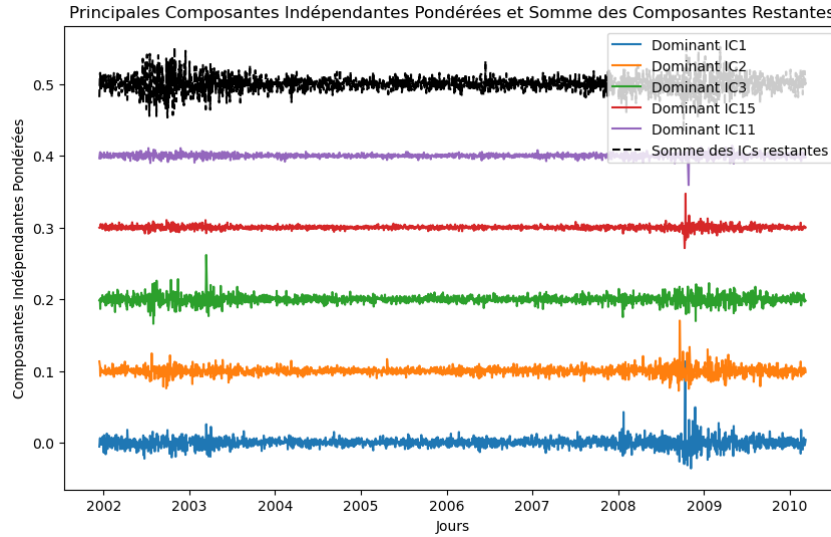


FIGURE 3 – 5 premières composantes (ICs) après pondération

- **Variabilité des composantes** : Certaines ICs présentent de fortes fluctuations, tandis que d'autres restent relativement stables, traduisant des influences hétérogènes.
- **Événements spécifiques** : Des pics marqués apparaissent, notamment en 2002-2003 et 2008-2009, en cohérence avec des périodes de crises financières.
- **Hiérarchisation limitée** : Avant pondération, les ICs sont extraites sans considération directe de leur poids relatif dans l'indice étudié, rendant leur interprétation plus délicate.

#### Après pondération :

La pondération par la matrice de mélange  $A$ , et plus précisément par les coefficients  $A_i$  associés à l'indice du **CAC 40**, entraîne une modification significative du classement et de l'interprétation des composantes indépendantes. Après pondération, les ICs dominantes identifiées diffèrent sensiblement de celles qui ressortaient avant cette étape, ce qui souligne que les composantes les plus actives en termes d'amplitude brute ne sont

pas nécessairement celles qui exercent l'influence la plus déterminante sur l'évolution globale de l'indice. Cette pondération permet également de réduire le bruit statistique en atténuant les fluctuations des composantes secondaires, ce qui améliore la lisibilité et la clarté des facteurs majeurs. Elle facilite par ailleurs l'interprétation économique des résultats, en mettant davantage en évidence des comportements de marché pertinents pouvant être associés à des dynamiques sectorielles ou à des événements économiques significatifs.

L'analyse révèle qu'un nombre restreint de facteurs clés semble gouverner la dynamique globale du marché, tandis que les autres composantes jouent un rôle plus marginal. Les pics de volatilité observés après pondération apparaissent de manière ciblée sur certaines périodes critiques, notamment lors de la crise financière de 2008, traduisant ainsi l'influence directe d'événements macroéconomiques majeurs. Enfin, les composantes résiduelles, quant à elles, se révèlent peu influentes dans l'explication des mouvements de l'indice, confirmant qu'une part importante de l'information pertinente est capturée par un sous-ensemble limité d'ICs dominantes.

**Conclusion :** L'application de l'ICA sur les log-rendements du CAC 40 met en évidence une structure de marché dominée par un nombre limité d'ICs significatives, dont l'identification et l'analyse peuvent être exploitées à des fins de gestion des risques et d'optimisation de portefeuille.

La suite logique de cette décomposition en composantes indépendantes est la reconstruction des rendements et des prix, à partir des ICs dominantes identifiées.

## 4.2 Reconstruction des prix à l'aide des composantes indépendantes (ICs)

### 4.2.1 Reconstruction de l'indice du CAC 40 par ICA sans seuillage

La reconstruction des prix de l'indice CAC 40 à partir des composantes indépendantes (ICs) permet d'évaluer la capacité de l'ICA à capturer les dynamiques majeures du marché à partir d'un nombre limité de facteurs.

Les figures ci-dessous comparent la reconstruction du prix réel du CAC 40 en utilisant :

- Les **2 ICs dominantes** (à gauche).
- Les **5 ICs dominantes** (à droite).

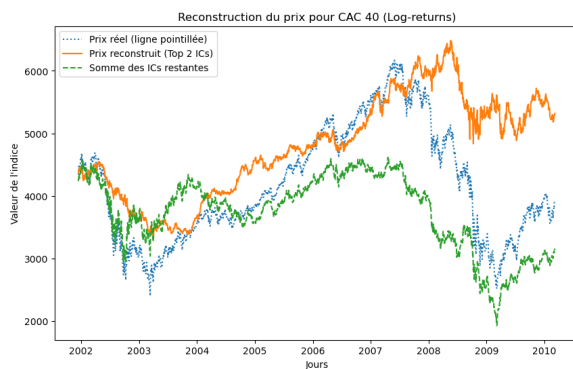


FIGURE 4 – Reconstruction avec les 2 ICs dominantes

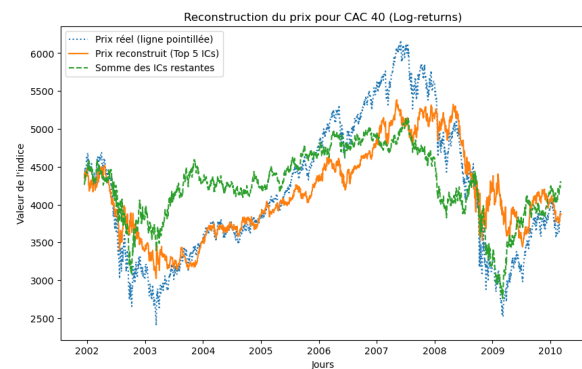


FIGURE 5 – Reconstruction avec les 5 ICs dominantes

### Analyse des résultats :

- Avec seulement **2 ICs dominantes**, la reconstruction permet de saisir les grandes tendances de long terme du CAC 40, mais manque de précision dans les phases de retournement et les crises, notamment en 2008-2009 où l'amplitude des variations est mal reproduite.
- En augmentant à **5 ICs dominantes**, la reconstruction devient beaucoup plus précise. Les dynamiques de crise, les périodes de croissance et les phases de correction sont mieux capturées, et le comportement global de l'indice s'en trouve significativement amélioré.
- On remarque également que la **somme des ICs restantes** (courbe verte) apporte des contributions non négligeables et représente un bruit de fond qui n'est pas entièrement dépourvu d'information.
- Cela illustre l'intérêt d'un compromis entre le nombre de composantes retenues et la qualité de la reconstruction : trop peu de composantes limitent la fidélité, tandis qu'un nombre trop élevé peut réintroduire du bruit.

Ces résultats mettent en évidence l'efficacité de l'ICA pour modéliser le comportement d'un indice boursier complexe comme le CAC 40. Ils confirment également que certaines ICs portent des informations majeures sur les dynamiques de marché, justifiant leur utilisation en priorité pour des objectifs de simplification et d'interprétation économique.

#### 4.2.2 Reconstruction des prix avec seuillage des composantes indépendantes

Afin de filtrer les composantes indépendantes (ICs) les moins significatives et de mieux isoler les dynamiques majeures du marché, nous appliquons un seuillage aux ICs extraites. Le seuil choisi agit directement sur les amplitudes des ICs, en annulant celles dont la contribution est inférieure à une valeur fixée ( $\xi$ ).

Nous comparons ici deux niveaux de seuillage :

- Un seuillage à **0.05**.
- Un seuillage plus faible à **0.01**.

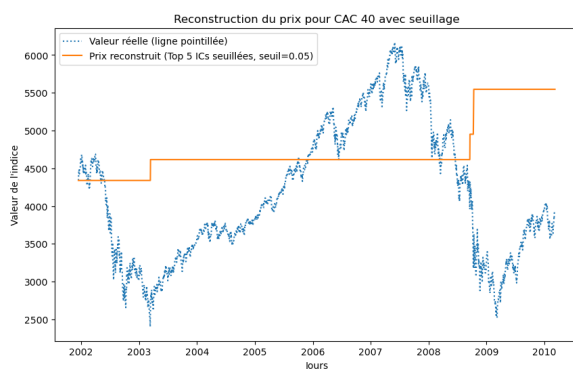


FIGURE 6 – Reconstruction des prix (seuil = 0.05)

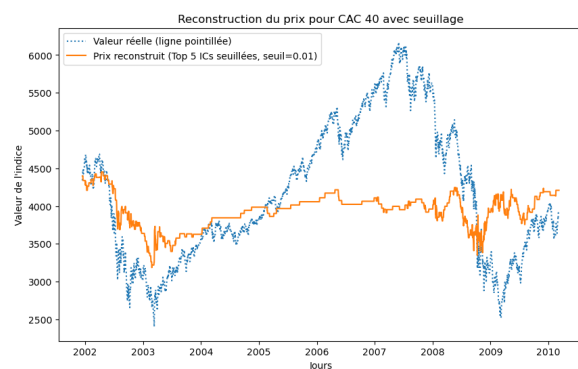


FIGURE 7 – Reconstruction des prix (seuil = 0.01)

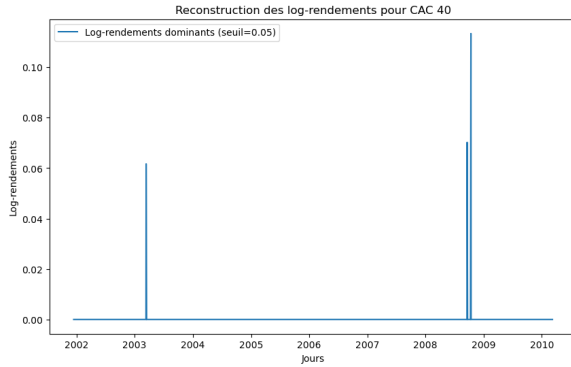


FIGURE 8 – Log-rendements dominants (seuil = 0.05)

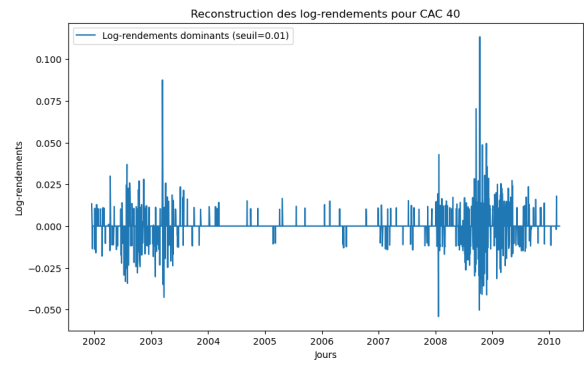


FIGURE 9 – Log-rendements dominants (seuil = 0.01)

### Commentaires et remarques :

- Avec un seuil élevé (**0.05**), la reconstruction est extrêmement dégradée : seuls quelques chocs majeurs sont conservés, ce qui produit un prix reconstruit très rigide et artificiel, en escalier, incapable de suivre les dynamiques fines du marché.
- La reconstruction des log-rendements pour ce seuil (**0.05**) montre clairement que seules quelques dates isolées présentent encore une activité significative, correspondant à des événements exceptionnels, tandis que le reste de la série est annulé.
- En abaissant le seuil à **0.01**, la qualité de la reconstruction s'améliore sensiblement. On récupère davantage de la variabilité du marché, et les log-rendements conservent de nombreuses fluctuations, permettant de mieux représenter les périodes de crise ou de volatilité accrue (comme en 2008-2009).
- Malgré l'amélioration avec un seuil plus bas, la reconstruction reste incomplète : le seuillage filtre efficacement le bruit faible, mais en contrepartie, il élimine aussi une partie des variations utiles, nuisant à la fidélité totale de la trajectoire du prix.
- Ce comportement confirme que le seuillage doit être appliqué avec prudence dans le cadre de l'ICA. Si l'objectif est d'isoler uniquement les événements majeurs ou extrêmes, un seuillage fort peut être pertinent. En revanche, pour des reconstructions précises et continues des prix, il est préférable d'opter pour un seuil faible ou de limiter l'usage du seuillage.

**Remarque générale :** Le seuillage met en évidence la capacité de l'ICA à identifier des chocs spécifiques et isolés dans le temps. Cependant, son efficacité pour une reconstruction fidèle et complète de la dynamique de l'indice est limitée, surtout lorsque le seuil est trop élevé. Il existe donc un compromis délicat à trouver entre simplification du signal et préservation de l'information.

Nous poursuivons l'analyse en abaissant davantage le seuil à **0.001**, afin de tester l'impact d'un seuillage très faible sur la qualité de la reconstruction.

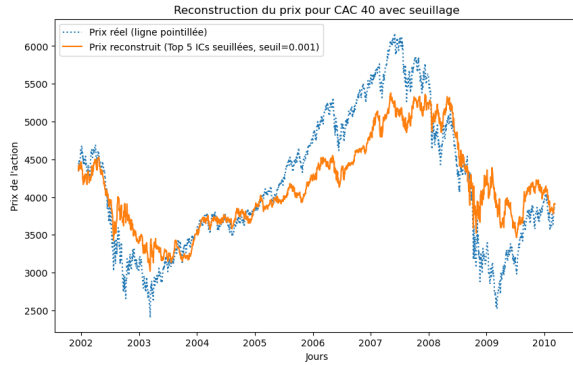


FIGURE 10 – Reconstruction des prix (seuil = 0.001)

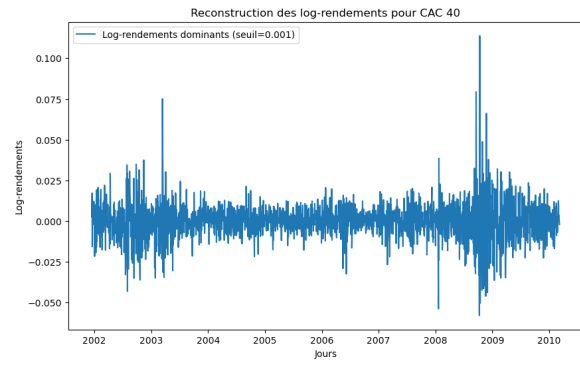


FIGURE 11 – Log-rendements dominants (seuil = 0.001)

### Commentaires et observations :

- Avec un seuil aussi faible que **0.001**, la reconstruction devient beaucoup plus fidèle aux dynamiques réelles du CAC 40. La trajectoire du prix reconstruit suit de manière plus précise les phases de hausse, de baisse et les périodes de crise.
- Les log-rendements reconstruits conservent désormais une variabilité dense et réaliste, capturant des amplitudes significatives durant les périodes de forte volatilité (notamment autour de la crise de 2008-2009).
- Ce seuil très faible permet de préserver un plus grand nombre de petites contributions des ICs, ce qui améliore la continuité du signal et la précision globale du modèle.
- Toutefois, à mesure que le seuil diminue, le risque est de réintroduire davantage de bruit dans la reconstruction. Ce compromis entre filtrage du bruit et restitution fidèle du signal doit être ajusté selon l'objectif final : détection des chocs majeurs ou modélisation précise du prix.

### Synthèse du seuillage :

- Un **seuil élevé** élimine efficacement le bruit mais sacrifie fortement l'information utile, produisant des reconstructions pauvres et artificielles.
- Un **seuil modéré** (comme 0.01) offre un équilibre intéressant, capturant les phases de volatilité tout en réduisant les fluctuations mineures.
- Un **seuil faible** (comme 0.001) conserve quasiment l'ensemble de la structure du signal et permet une reconstruction proche du prix réel, au prix d'une possible réintroduction de bruit.

En conclusion, l'application du seuillage dans l'ICA permet de simplifier l'interprétation et de mettre en évidence les facteurs dominants, mais un seuil trop strict dégrade la capacité du modèle à reproduire les dynamiques complètes du marché. Le choix optimal du seuil dépend donc de l'équilibre recherché entre simplification et fidélité du signal.

## 4.3 Comparaison des résultats entre ICA et PCA

Afin d'évaluer les performances respectives de l'ICA et de la PCA dans la reconstruction des prix et des rendements du CAC 40, nous présentons ci-dessous les résultats obtenus avec la PCA, sans et avec seuillage.



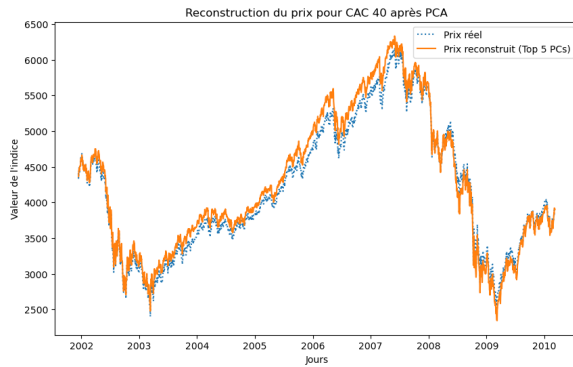


FIGURE 12 – Reconstruction des prix (Top 5 PCs)

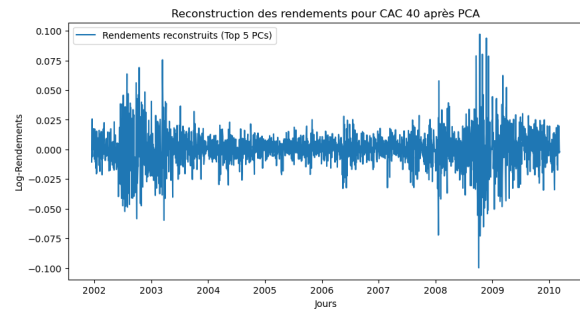


FIGURE 13 – Log-Rendements reconstruits (Top 5 PCs)

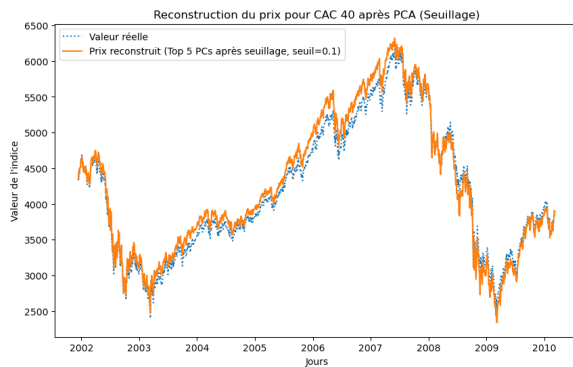


FIGURE 14 – Reconstruction des prix avec seuillage (seuil = 0.1)

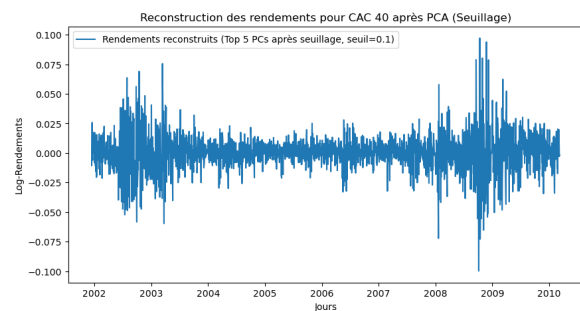


FIGURE 15 – Log-Rendements reconstruits avec seuillage (seuil = 0.1)

### Commentaires et analyse comparative :

#### — Fidélité de la reconstruction :

- La PCA, même avec un nombre limité de composantes (Top 5 PCs), offre une excellente approximation du prix réel du CAC 40, capturant avec précision les tendances générales ainsi que les phases de crise, notamment celle de 2008-2009.
- L'ICA nécessite davantage d'ajustements (pondération, seuillage fin) pour atteindre un niveau de précision comparable, bien que ses ICs soient plus interprétables économiquement.

#### — Comportement des rendements :

- Les rendements reconstruits via la PCA sont plus denses et homogènes sur toute la période, traduisant une meilleure capacité à restituer la structure globale de la volatilité.
- En comparaison, l'ICA tend à isoler des pics spécifiques correspondant à des événements majeurs, surtout après seuillage, mais perd de l'information continue.

#### — Effet du seuillage :

- Avec la PCA, le seuillage (ici fixé à 0.1) n'a qu'un impact limité sur la qualité de la reconstruction, conservant une bonne fidélité du prix et des rendements.



Cela s'explique par le fait que la PCA ordonne les composantes par variance décroissante, assurant que les premières PCs concentrent l'essentiel de l'information utile.

- Pour l'ICA, au contraire, le seuillage affecte beaucoup plus fortement la qualité de la reconstruction. Un seuillage trop fort dénature la dynamique des prix et filtre excessivement les contributions faibles mais pertinentes.
- **Robustesse face aux crises :**
  - La PCA permet une restitution fidèle et régulière, y compris lors de fortes crises.
  - L'ICA identifie très bien les chocs spécifiques (comme en 2008), mais au prix d'une perte de détail sur les périodes plus calmes ou intermédiaires.

## Synthèse

La PCA apparaît comme la méthode la plus performante lorsqu'il s'agit d'obtenir une reconstruction fidèle et continue des prix et des rendements du CAC 40. Cette efficacité tient à sa capacité à concentrer l'essentiel de l'information dans ses premières composantes, ce qui permet de restituer avec précision la dynamique globale du marché, même en utilisant un nombre limité de facteurs. À l'inverse, l'ICA se distingue par sa faculté à isoler des facteurs indépendants et spécifiques, qui s'avèrent particulièrement pertinents lors de périodes de stress ou de crises. Toutefois, cette spécialisation dans la détection d'événements singuliers s'accompagne parfois d'une perte de cohérence dans la reconstruction continue du signal, rendant l'ICA moins performante que la PCA pour reproduire fidèlement l'ensemble du comportement du marché. Par ailleurs, bien que le seuillage soit un outil utile pour simplifier l'interprétation des composantes et concentrer l'analyse sur les facteurs dominants, il doit être appliqué avec une prudence accrue dans le cas de l'ICA. En effet, contrairement à la PCA, où les premières composantes concentrent naturellement l'essentiel de la variance, certaines ICs de faible amplitude peuvent porter des dynamiques économiques importantes qu'un seuillage trop strict risquerait d'éliminer.

## 5 Limites et perspectives

Bien que l'application de l'ICA et de la PCA aux rendements du CAC 40 ait permis de mettre en lumière des dynamiques intéressantes, certaines limites doivent être soulignées.

### Limites de l'étude

- **Choix du seuil de seuillage :** La fixation du seuil de seuillage pour l'ICA s'est faite de manière empirique. Un seuil trop élevé appauvrit fortement la reconstruction, tandis qu'un seuil trop faible réintroduit du bruit. Une approche plus rigoureuse, fondée sur des critères statistiques ou économiques, pourrait optimiser ce paramètre.
- **Nombre de composantes retenues :** Tant pour l'ICA que pour la PCA, le choix du nombre de composantes (ICs ou PCs) est déterminant. Ici, nous avons utilisé les 5 principales, mais ce choix pourrait être affiné par des tests de robustesse ou des indicateurs comme par exemple l'information criterion (AIC, BIC, ...).
- **Interprétation économique des composantes :** Si la PCA maximise la variance expliquée, elle produit des facteurs peu interprétables économiquement. L'ICA propose des facteurs indépendants, mais leur signification économique reste parfois

difficile à établir sans analyse complémentaire.

- **Données limitées** : L'étude porte uniquement sur le CAC 40 sur la période 2001-2010. Une extension à d'autres indices ou à des périodes plus récentes permettrait de tester la robustesse des résultats face à différents contextes de marché.

## Perspectives

- **Optimisation du seuillage** : Implémenter des méthodes automatiques de calibration du seuil de seuillage, par validation croisée ou par minimisation de l'erreur de reconstruction.
- **Extension à d'autres indices** : Appliquer la même méthodologie à d'autres indices boursiers (S&P 500, DAX, Nikkei) pour comparer les structures de marché et vérifier la généralisation des résultats.
- **Identification économique des ICs** : Associer les ICs extraites à des variables macroéconomiques (taux d'intérêt, inflation, indicateurs sectoriels) afin d'enrichir l'interprétation économique des facteurs.
- **Utilisation de réseaux de neurones** : Explorer l'extraction des ICs via des architectures neuronales (par exemple l'ICA non linéaire ou des autoencodeurs) pour détecter des structures complexes et non linéaires.

## 6 Conclusion

Dans ce projet, nous avons appliqué l'Analyse en Composantes Indépendantes (ICA) et l'Analyse en Composantes Principales (PCA) aux log-rendements des actions du CAC 40 entre 2001 et 2010, afin d'extraire les facteurs latents structurants le marché et d'évaluer leur capacité à reconstruire fidèlement l'évolution de l'indice. L'ensemble des résultats met en évidence des différences fondamentales entre ces deux approches, tant du point de vue technique qu'économique.

La PCA s'est montrée particulièrement performante pour restituer les dynamiques globales du CAC 40, grâce à sa capacité à condenser l'essentiel de la variance dans les premières composantes. Cette efficacité lui permet d'offrir une reconstruction précise et stable du prix et des rendements, même avec un nombre limité de facteurs. Toutefois, la PCA souffre d'un manque d'interprétabilité économique, ses composantes principales étant construites uniquement sur des critères de maximisation de la variance, sans garantie d'indépendance ni de lien direct avec des phénomènes économiques identifiables.

L'ICA, quant à elle, propose une lecture complémentaire du marché. En extrayant des facteurs statistiquement indépendants, elle met en lumière des dynamiques spécifiques et souvent plus interprétables. D'un point de vue économique, les composantes indépendantes obtenues par l'ICA pourraient représenter différents moteurs du marché. Certaines ICs semblent refléter des tendances globales affectant l'ensemble des titres de l'indice, tandis que d'autres capturent des comportements plus localisés, qu'il s'agisse de dynamiques sectorielles ou de facteurs idiosyncratiques propres à certaines entreprises. Il est également probable que certaines ICs traduisent l'impact d'événements exogènes, tels que des crises financières ou des annonces macroéconomiques majeures. Néanmoins, cette interprétation reste à approfondir : il serait pertinent, dans la continuité de ce travail, de chercher à associer formellement chaque composante indépendante à des variables économiques ou financières spécifiques, par exemple à travers l'étude de leurs corrélations avec des indicateurs macro-financiers.

Par ailleurs, l'analyse a montré que l'application d'un seuillage aux ICs permet de filtrer les signaux faibles pour isoler les événements les plus marquants, au prix toutefois d'une dégradation de la qualité de la reconstruction lorsque le seuil devient trop strict. Ce constat souligne l'importance de calibrer précisément ce paramètre en fonction des objectifs recherchés, qu'il s'agisse de simplification du signal ou de fidélité de la restitution.

En définitive, la PCA et l'ICA apparaissent comme deux approches complémentaires. La PCA excelle dans la restitution fidèle des prix et des rendements globaux, tandis que l'ICA offre un potentiel d'analyse plus riche pour comprendre les facteurs profonds et indépendants qui structurent le comportement du marché. Ce travail ouvre ainsi des perspectives intéressantes pour des recherches futures, notamment sur l'interprétation économique des ICs, l'optimisation des paramètres du modèle, ou encore l'application de ces méthodes à d'autres marchés financiers et périodes historiques.

## 7 Référence de l'article étudié

Ce projet s'appuie principalement sur l'étude pionnière suivante, qui constitue la base méthodologique et théorique de notre travail :

Back, A., & Weigend, A. S. (1997). *A First Application of Independent Component Analysis to Extracting Structure from Stock Returns*.