



Faculté

des **sciences économiques** et de **gestion**

Université de Strasbourg

**Mémoire d'Économétrie Avancée III:**  
**Techniques de prévision des rendements financiers**

FROEHLICH Marc

MIRANDA André

*Titulaire du cours* : Dr. Nathalie Picard

*Date* : 31/01/2024

---

## SOMMAIRE

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Revue de la littérature</b>	<b>3</b>
2.1	Le rôle central des prévisions dans les marchés financiers . . . . .	3
2.2	Les multiples rôles des prévisions dans la prise de décisions . . . . .	3
2.3	Stratégies de trading et élaboration de politiques . . . . .	3
2.4	Les fondements de l'analyse technique et fondamentale . . . . .	3
2.5	Remise en question par l'Hypothèse du Marché Efficient (HME) . . . . .	4
2.6	Stratégies de Prédiction dans la Littérature . . . . .	4
2.7	Modèles HMM et ARIMA . . . . .	5
2.8	Evaluation des Stratégies . . . . .	5
<b>3</b>	<b>Méthodologie</b>	<b>7</b>
3.1	Collecte des données . . . . .	7
3.2	Prétraitement des données . . . . .	8
3.3	Modèle ARIMA . . . . .	11
3.3.1	Composante Auto-Régressive (AR) . . . . .	11
3.3.2	Composante d'Intégration (I) . . . . .	11
3.3.3	Composante à Moyenne Mobile (MA) . . . . .	12
3.3.4	Application du Modèle ARIMA . . . . .	12
3.4	Analyse et Modélisation ARIMA pour les Rendements du S&P500 . . . . .	12
3.4.1	Stationnarité de la Série Temporelle . . . . .	12
3.4.2	Choix du Degré du Modèle ARIMA . . . . .	13
3.4.3	Résultats du Modèle ARIMA(4, 0, 5) . . . . .	13
3.4.4	Prévisions . . . . .	15
3.4.5	Conclusion modèle ARIMA . . . . .	16
3.5	Modèle HMM . . . . .	17
3.6	Application et Analyse du HMM pour le Cours du S&P500 . . . . .	18
3.6.1	Ajustement du modèle . . . . .	18
3.6.2	Généralité du modèle . . . . .	19
3.6.3	Etats Cachés . . . . .	22

3.6.4	Prédiction . . . . .	22
<b>4</b>	<b>Conclusion</b>	<b>23</b>
<b>5</b>	<b>Bibliographie</b>	<b>24</b>
<b>6</b>	<b>Annexe</b>	<b>25</b>

# 1 Introduction

La capacité à anticiper les mouvements futurs des marchés financiers revêt une importance cruciale dans un contexte économique en perpétuelle évolution. Dans ce cadre dynamique, où les volatilités sont monnaie courante, les prévisions financières émergent comme un outil stratégique incontournable, offrant aux investisseurs, analystes et décideurs une boussole indispensable pour naviguer dans les eaux complexes de la finance mondiale. Ces prévisions ne sont pas simplement des conjectures, mais plutôt des guides prévisionnels qui permettent de saisir les opportunités et de mitiger les risques dans un paysage financier en constante métamorphose.

Au cœur de ces préoccupations se dresse l'indice S&P 500, un véritable baromètre emblématique des marchés boursiers mondiaux. En agrégeant les performances des 500 plus grandes entreprises américaines, le S&P 500 devient le reflet non seulement des dynamiques économiques aux États-Unis mais aussi un indicateur puissant des tendances globales. C'est un miroir qui capture les nuances de l'économie mondiale, résonnant des forces et des faiblesses qui façonnent le paysage financier international.

Ainsi, plonger dans les prévisions du S&P 500 revient à s'aventurer dans un territoire où la complexité des marchés s'entremêle avec les ambitions et les incertitudes des acteurs économiques. Les investisseurs cherchent à déchiffrer les signaux, les analystes à interpréter les tendances, et les décideurs à éclairer leurs choix stratégiques. Ce mémoire s'engage dans cette exploration fascinante, en se penchant sur deux modèles prédictifs de renom, les modèles Hidden Markov (HMM) et AutoRegressive Integrated Moving Average (ARIMA), afin de jeter une lumière analytique sur les contours futurs du S&P 500. En effet, au-delà d'une simple quête de prédictions, cette étude s'inscrit dans la recherche d'une compréhension approfondie des mécanismes qui gouvernent les fluctuations des marchés et, partant, dans la quête de perspectives éclairantes pour ceux qui naviguent dans ces eaux tumultueuses.

En tant qu'indicateur avancé des mouvements économiques mondiaux, le S&P 500 joue un rôle crucial dans la projection des dynamiques à venir. Ainsi, la compréhension des mécanismes de prévision du S&P 500 acquiert une pertinence singulière dans le contexte complexe et interconnecté de l'économie mondiale. Elle devient une clé analytique pour déchiffrer les évolutions économiques à venir, tout en offrant des perspectives cruciales pour ceux qui cherchent à naviguer avec succès dans le dédale des marchés financiers à l'échelle mondiale. En explorant les méandres des prévisions du S&P 500, cette étude s'inscrit dans une quête d'éclaircissements indispensables pour ceux qui aspirent à appréhender l'avenir de l'économie mondiale à travers le prisme des marchés financiers.

Cette étude s'engage à approfondir notre compréhension des modèles HMM (Hidden Markov Model) et ARIMA (AutoRegressive Integrated Moving Average) qui se positionnent comme des piliers dans le

domaine des prévisions financières. À travers une exploration de leurs principes fondamentaux, nous aspirons à démystifier leur fonctionnement complexe, à décortiquer les mécanismes sous-jacents qui les rendent efficaces dans la modélisation des séries temporelles financières. Notre objectif est de fournir une base théorique solide permettant aux lecteurs acquérir des nouvelles capacités et étendre leurs connaissances dans le contexte spécifique des marchés financiers et les outils statistiques de prévision.

Au-delà de la compréhension théorique, notre démarche réside dans l'application concrète des modèles HMM et ARIMA à la tâche spécifique des prévisions du S&P 500. En confrontant ces modèles aux réalités mouvantes du marché, nous voulons extraire des perspectives éclairantes sur la performance anticipée de l'indice. En exploitant les données historiques du S&P 500, notre démarche se veut pragmatique et orientée vers des résultats tangibles. Ces résultats non seulement contribueront à enrichir la littérature académique sur les prévisions financières, mais, surtout ils ouvriront la voie à des prises de décision plus informées pour les acteurs du marché, renforçant ainsi leur capacité à naviguer dans un environnement financier complexe et dynamique.

## **2 Revue de la littérature**

### **2.1 Le rôle central des prévisions dans les marchés financiers**

L'importance de la prédiction précise dans le domaine financier réside dans son impact direct sur les acteurs clés des marchés boursiers, notamment les investisseurs, les chercheurs et les gouvernements. Cette affirmation souligne que la capacité à anticiper avec précision les mouvements du marché boursier est cruciale pour prendre des décisions éclairées. Les investisseurs, en particulier, utilisent ces prévisions comme un guide essentiel pour leurs décisions d'investissement, cherchant à maximiser les rendements tout en minimisant les risques inhérents aux marchés financiers volatils.

### **2.2 Les multiples rôles des prévisions dans la prise de décisions**

Les prévisions ne sont pas simplement des outils pour les investisseurs. Elles jouent également un rôle central dans l'identification des opportunités et des défis présents sur le marché (Krauss et al., 2017). La capacité à anticiper les mouvements futurs permet aux acteurs du marché de se positionner de manière proactive pour exploiter les opportunités émergentes ou, à l'inverse, pour se prémunir contre les défis potentiels. Cette capacité d'anticipation est essentielle dans un environnement financier en constante évolution.

### **2.3 Stratégies de trading et élaboration de politiques**

Outre les investisseurs individuels, les prévisions jouent un rôle crucial dans l'élaboration de stratégies de trading. Les modèles de prévision, qu'ils soient basés sur l'analyse technique ou fondamentale, fournissent des orientations quant à l'achat, la vente ou le maintien d'actifs financiers. Ces stratégies sont élaborées en tenant compte des informations passées, souvent supposant que les tendances historiques se répéteront à l'avenir. Ainsi, la construction de ces modèles implique une analyse approfondie des données historiques, comme le souligne Guerard Jr. (2013).

### **2.4 Les fondements de l'analyse technique et fondamentale**

Les deux principales approches pour construire des modèles de prévision sont l'analyse technique et l'analyse fondamentale. L'analyse technique, axée sur les mouvements de prix et les volumes, repose sur l'hypothèse que les modèles de prix passés se répéteront dans le futur (Leigh et al., 2002). D'un autre côté, l'analyse fondamentale se base sur des données spécifiques à l'entreprise, comme les bénéfices

et les perspectives, pour évaluer la valeur intrinsèque des actifs (Bodie et al., 2009). Ces deux approches complémentaires cherchent à exploiter différentes facettes des marchés financiers pour générer des prévisions robustes.

## **2.5 Remise en question par l'Hypothèse du Marché Efficient (HME)**

Cependant, cette quête de prédictions précises n'est pas sans controverse, car l'Hypothèse du Marché Efficient (HME) remet en question la validité même de ces modèles. Selon l'HME, les prix du marché suivent une marche aléatoire, rendant toute prédiction basée sur les mouvements passés non fondée (Leigh et al., 2002). Cette perspective théorique influence la construction des modèles de prévision, soulignant la nécessité de repenser les approches traditionnelles.

## **2.6 Stratégies de Prédiction dans la Littérature**

La littérature financière a connu une pléthore de stratégies de prédiction visant à anticiper avec précision les mouvements des marchés boursiers. Parmi ces stratégies, diverses approches ont été explorées, allant des machines à vecteurs de support aux réseaux neuronaux, en passant par les algorithmes génétiques. Les travaux de Guo et al. (2014) ainsi que ceux de Kim et al. (2004) illustrent la diversité des méthodologies adoptées pour prédire les évolutions du marché.

Les machines à vecteurs de support ont été largement utilisées dans la littérature financière (Guo et al., 2014; Kim, 2003). Dans le cadre de la prédiction du marché boursier de Shanghai et de l'indice Dow Jones, Guo et al. (2014) ont intégré 39 caractéristiques, dont le prix d'ouverture, le prix élevé, le prix bas, les moyennes mobiles, les termes de momentum, le RSI, les stochastiques %K et %D, le MACD, le momentum, et d'autres indicateurs techniques. Leur modèle a surpassé d'autres modèles comparatifs, soulignant l'efficacité de cette approche dans la capture des mouvements du marché.

Les réseaux neuronaux ont également démontré leur efficacité dans la prédiction des marchés financiers (Kim et al., 2004; Cao et Tay, 2001). Les travaux de Kim (2003) sur la bourse coréenne (KOSPI) ont utilisé 12 indicateurs techniques communs, montrant que les machines à vecteurs de support surpassaient le réseau neuronal de référence et le modèle CBR. De même, Cao et Tay (2001) ont utilisé une machine à vecteurs de support pour prédire l'indice quotidien du SP500, obtenant des résultats supérieurs à un réseau neuronal de rétropropagation.

Les algorithmes génétiques ont été intégrés avec succès dans la prédiction des indices boursiers (Kim et al., 2004; Kim, Min, Han, 2006). Kim et al. (2004) ont utilisé un mécanisme d'intégration hybride avec un algorithme génétique flou comprenant neuf indicateurs techniques. Leur approche a généré des résultats précis pour la prédiction de l'indice boursier coréen KOSPI. Les travaux de Kim, Min, Han

(2006) ont combiné les connaissances d'un réseau neuronal et d'experts humains pour un algorithme génétique, surpassant les méthodes de référence.

## **2.7 Modèles HMM et ARIMA**

Dans cette revue, notre attention se porte spécifiquement sur deux modèles de prédiction - les modèles HMM (Hidden Markov Models) et ARIMA (AutoRegressive Integrated Moving Average). Ces deux approches méthodologiques offrent des perspectives distinctes pour la prédiction des rendements hebdomadaires du S&P500.

Les modèles HMM, inspirés par la théorie des processus stochastiques, sont particulièrement adaptés à la modélisation de séries temporelles où la dynamique sous-jacente peut changer de manière non linéaire (Rabiner, 1989). Ces modèles permettent de capturer les transitions cachées entre différents états du marché, offrant ainsi une représentation plus sophistiquée des dynamiques complexes du marché.

D'autre part, les modèles ARIMA sont bien établis dans la littérature pour la prédiction de séries temporelles (Box et al., 1994). Basés sur des composantes auto-régressives, intégrées et à moyenne mobile, les modèles ARIMA sont adaptés à des données où les tendances passées peuvent avoir un impact significatif sur les futurs mouvements.

## **2.8 Evaluation des Stratégies**

La section finale de cette revue consacrée à l'évaluation des stratégies de prédiction, incluant celles qui reposent sur les modèles HMM et ARIMA, revêt une importance capitale pour la compréhension globale des performances de ces modèles dans le contexte dynamique du marché financier, particulièrement du S&P500.

L'évaluation comparative des stratégies se base sur une analyse approfondie de leur rendement par rapport à une stratégie de conservation, souvent représentée par l'approche "buy-and-hold". Cette approche consiste à détenir des actifs sur le long terme sans chercher activement à réajuster le portefeuille en réponse aux fluctuations du marché. La comparaison avec cette stratégie de référence fournit un point de repère essentiel pour déterminer l'efficacité relative des modèles de prédiction étudiés.

Par ailleurs, une exploration détaillée des contributions des classes prédites aux rendements permet d'apporter une nuance cruciale à l'évaluation des modèles. En analysant comment chaque classe anticipée impacte les résultats finaux, il devient possible de discerner la capacité des modèles à détecter et à tirer profit des différentes tendances du marché. Cette approche fine de l'évaluation offre une compréhension plus riche des forces et des faiblesses de chaque modèle, contribuant ainsi à une prise de décision informée.



L'évaluation ne se limite pas à la performance financière brute. Elle englobe également des aspects tels que la stabilité des prédictions dans des conditions de marché variées, la robustesse face à des événements inattendus, et la réactivité aux changements structurels. Ces dimensions fournissent une perspective holistique sur la fiabilité et l'applicabilité réelle des modèles dans des contextes financiers dynamiques et parfois volatils.

Une attention particulière doit être accordée à la validation des résultats obtenus par les modèles, en mettant en lumière la période de test, les métriques de performance utilisées, et la cohérence des résultats avec les attentes théoriques. Cette rigueur méthodologique est cruciale pour garantir que les conclusions tirées de l'évaluation reflètent de manière fiable les capacités prédictives des modèles.

Enfin, cette section d'évaluation offre également l'opportunité d'identifier des pistes d'amélioration potentielles pour les modèles HMM et ARIMA. En scrutant les limitations observées et les situations où ces modèles pourraient montrer des signes de faiblesse, il devient possible d'orienter les futures recherches vers des développements et des ajustements spécifiques, contribuant ainsi à l'évolution continue des méthodologies de prédiction financière.

En somme, l'évaluation des stratégies, au-delà d'être une étape cruciale de cette revue, représente une clé pour déverrouiller les enseignements pratiques et théoriques offerts par les modèles HMM et ARIMA dans le contexte complexe du marché boursier, éclairant ainsi la voie vers des avancées futures dans la science des prévisions financières.

## 3 Méthodologie

### 3.1 Collecte des données

Les données cruciales pour cette analyse ont été extraites de manière minutieuse à partir de la plateforme Yahoo Finance en utilisant la fonction `getSymbols` du package R, `quantmod`. Ce processus a permis d'obtenir un ensemble complet d'informations couvrant différentes facettes du comportement du S&P500 au cours des séances de négociation.

#### Variables extraites :

- **OPEN (Prix d'ouverture)** : Représente la valeur initiale à laquelle le S&P500 a été coté au début de chaque journée de négociation.
- **HIGH (Prix le plus élevé)** : Désigne le point culminant atteint par le S&P500 en termes de valeur au cours d'une journée donnée.
- **LOW (Prix le plus bas)** : Indique la valeur minimale atteinte par le S&P500 au cours d'une séance de négociation.
- **CLOSE (Prix de clôture)** : Représente la valeur finale à laquelle le S&P500 a été coté à la fin de chaque journée de négociation.
- **VOLUME (Volume des transactions)** : Représente la quantité totale d'actions du S&P500 échangées au cours d'une journée de négociation.
- **ADJUSTED CLOSE (Prix ajusté en clôture)** : Tient compte des ajustements, tels que les dividendes ou les émissions d'actions, et représente la valeur finale ajustée du S&P500.
- **RETURNS (Rendements)** : représente les rendements associés à chaque prix de clôture du S&P500.

Ces valeurs découlent de la formule

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

avec  $P_t$  le prix de clôture à la période  $t$ .

La période d'étude englobe le laps de temps débutant de 2007 à 2024. Cette sélection minutieuse vise à embrasser un spectre varié de conditions de marché, allant des phases de croissance économique aux périodes de récession. En élargissant notre fenêtre temporelle, nous nous assurons d'intégrer une diversité de contextes économiques qui peuvent influencer de manière significative le comportement du S&P500 sur les marchés financiers.

### 3.2 Prétraitement des données



Figure 1: Cours du S&P500

La période de 2008 à 2023 dans l'historique du S&P500 a été caractérisée par une séquence d'événements majeurs ayant exercé une influence significative sur les mouvements du marché. Cette analyse approfondie vise à examiner ces fluctuations de cours en identifiant les principaux chocs et en expliquant leurs impacts dynamiques sur l'indice boursier.

La crise financière mondiale de 2008 a constitué l'un des chocs les plus marquants de cette période. Déclenchée par la crise des prêts hypothécaires à risque aux États-Unis, elle a engendré une baisse significative du S&P500 reflétant la panique des investisseurs face aux risques systémiques, particulièrement ressentis dans les secteurs financier et immobilier. La crise financière mondiale déclenchée par la crise des subprimes en 2008 a entraîné une baisse significative du cours du S&P500. Le sommet pré-crise du S&P500 était atteint en octobre 2007, avec un niveau d'environ 1 565 points. À la suite de la crise des subprimes, le S&P500 a connu une baisse substantielle. Le point le plus bas de la crise a été enregistré en mars 2009, avec le S&P500 chutant à environ 676 points. Ainsi, la chute du S&P500 pendant la crise des subprimes a représenté une baisse d'environ 57% par rapport à son sommet pré-crise en octobre 2007. Cela souligne l'ampleur significative de l'impact de la crise des subprimes sur les marchés financiers mondiaux. Cependant, la période postérieure à la crise a été caractérisée par une récupération soutenue à partir de 2009. Les mesures de relance économique mises en place ont stimulé la reprise, avec le S&P500 affichant des gains significatifs. Les politiques monétaires accommodantes ont joué un rôle crucial dans le rétablissement de la confiance des investisseurs.

En 2016, le vote du Royaume-Uni en faveur du Brexit a créé une nouvelle vague d'incertitude. Bien que le S&P500 ait initialement réagi avec une volatilité accrue, il a finalement retrouvé son équilibre à mesure que les investisseurs ont assimilé les implications économiques de cet événement.

L'année 2020 a été marquée par la pandémie mondiale de COVID-19, provoquant des fermetures économiques massives et une panique sur les marchés financiers. Le S&P500 a subi une correction rapide, mais a connu une reprise impressionnante grâce à des mesures de relance massives et aux développements rapides des vaccins. La pandémie mondiale de COVID-19 a profondément impacté les marchés financiers, marquant un point bas pour le S&P500 en mars 2020, où l'indice a atteint environ 2 237 points, soit une baisse d'environ 33,9 pourcent par rapport à son sommet précédent en février 2020, qui était d'environ 3 386 points. Cette correction rapide a été attribuée à l'incertitude économique généralisée liée à la pandémie. Cependant, grâce à des mesures de relance massives et aux progrès rapides dans le développement des vaccins, le S&P500 a connu une reprise impressionnante par la suite.

Au cours des années suivantes (2021-2023), des préoccupations liées à l'inflation, aux politiques monétaires et aux tensions géopolitiques ont influencé les anticipations des investisseurs. Le S&P500 a réagi avec des signes de volatilité, les investisseurs ajustant leurs portefeuilles en réponse aux changements perçus dans l'environnement économique mondial.

<b>Statistique</b>	<b>Valeur</b>
Moyenne	2320.54
Ecart Type	1117.45
Minimum	735.09
Maximum	4927.93
Skewness (Cours à la clôture)	0.69
Kurtosis (Cours à la clôture)	2.30

Table 1: Statistiques descriptives du cours à la clôture du S&P500.

Le tableau des statistiques descriptives pour le cours à la clôture du S&P500 sur la période de 2007 à aujourd'hui fournit des indications clés sur la distribution de ces données financières. La moyenne de près de 2320.54 représente le niveau moyen du cours à la clôture sur cette période, donnant un aperçu de la tendance centrale des données. L'écart-type de 1117.45 indique une dispersion significative des cours à la clôture par rapport à la moyenne, soulignant la variabilité des valeurs observées. Le cours minimum de 735.09 indique le niveau le plus bas atteint par le S&P500 au cours de cette période, offrant un aperçu des phases de baisse importante.

Le cours maximum de 4927.93 représente le sommet atteint par le S&P500, mettant en lumière les phases de croissance significative. Une skewness de 0.69 suggère une légère asymétrie positive dans la distribution des cours à la clôture, indiquant une tendance vers des valeurs plus élevées. La skewness du S&P500 révèle une valeur de 0,69, indiquant une asymétrie positive dans la distribution des rendements. Cette asymétrie vers la droite suggère que la majorité des rendements sont concentrés du côté gauche de la moyenne, tandis que quelques valeurs extrêmes élevées tirent la distribution vers la droite. En d'autres

termes, il y a une propension à réaliser des rendements relativement plus élevés que la moyenne, avec des moments de forte croissance ponctués par des événements exceptionnels.

Pour les investisseurs, cette skewness positive pourrait être interprétée comme une plus grande probabilité de réaliser des rendements élevés, bien que cela soit associé à une certaine volatilité due aux valeurs extrêmes. Il est important de prendre en compte cette asymétrie lors de l'évaluation des risques et des rendements potentiels des placements. La kurtosis du SP500 est mesurée à 2,30, ce qui est inférieur à la kurtosis de 3,0 d'une distribution normale. Cette valeur suggère que la distribution des rendements est moins lourde que celle d'une distribution normale, mais elle indique tout de même une certaine concentration des rendements autour de la moyenne avec des queues relativement épaisses.

En interprétant une kurtosis de 2,30, on peut noter que la distribution des rendements du S&P500 présente une modération des valeurs extrêmes par rapport à une distribution normale. Cela signifie que, bien que les rendements puissent avoir des queues épaisses, la distribution est relativement moins propice à des événements extrêmes par rapport à une distribution normale. Cependant, la concentration des rendements autour de la moyenne reste importante, ce qui peut indiquer une certaine stabilité et prévisibilité dans le comportement général du marché.

Ces statistiques fournissent des informations essentielles pour comprendre la dynamique du marché financier du S&P500 au cours de la période spécifiée, en offrant des indications sur la tendance, la variabilité et la forme de la distribution des cours à la clôture. Ces éléments contribuent à une analyse approfondie de la performance du S&P500 au fil du temps.

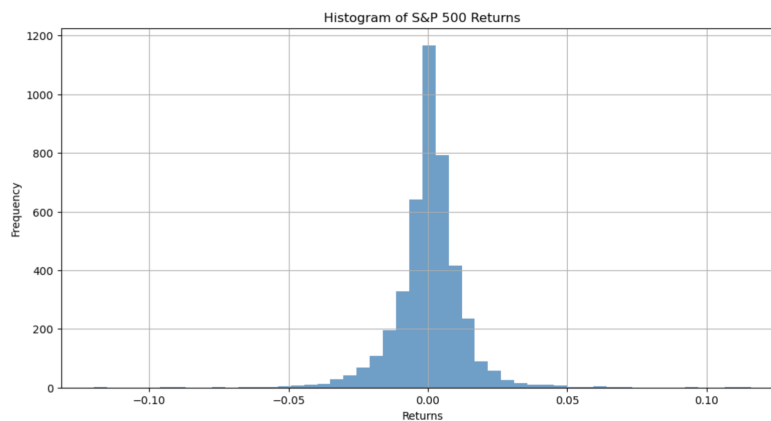


Figure 2: Histogramme des Rendements S&P500

L'histogramme du cours du S&P500 présentant une ressemblance avec une distribution normale centrée en 0 peut être attribué à divers facteurs, bien que les rendements des marchés financiers ne suivent généralement pas une distribution parfaitement normale. La distribution normale est souvent utilisée comme une approximation dans le domaine financier, et cela peut être influencé par l'effet central

limite.

Selon l'effet central limite, la somme de nombreuses variables aléatoires indépendantes et identiquement distribuées tend vers une distribution normale, même si les variables d'origine ne sont pas normalement distribuées. Les fluctuations quotidiennes du marché, résultant de divers facteurs, peuvent converger vers une distribution qui semble normale. Cependant, il est important de noter que la réalité des marchés financiers est souvent plus complexe, et les distributions peuvent présenter des caractéristiques différentes de la pure normalité.

En ce qui concerne la skewness de 0,69, cela suggère une asymétrie positive dans la distribution des rendements du SP500. Une skewness positive indique que la queue droite de la distribution est plus étirée ou plus épaisse que la queue gauche. Cela pourrait signifier une propension accrue à des rendements positifs plus élevés que la moyenne.

Quant à la kurtosis de 2,30, elle est inférieure à la kurtosis de 3,0 d'une distribution normale. Cela indique que la distribution des rendements est moins lourde que celle d'une distribution normale, mais elle montre toujours une certaine concentration des rendements autour de la moyenne avec des queues relativement épaisses.

### **3.3 Modèle ARIMA**

Le modèle ARIMA est une approche puissante pour modéliser et prévoir des séries temporelles. Il combine trois composantes principales : l'Auto-Régressif (AR), l'Intégration (I), et à Moyenne Mobile (MA).

#### **3.3.1 Composante Auto-Régressive (AR)**

L'aspect auto-régressif du modèle ARIMA capture la dépendance de chaque observation avec ses valeurs passées. L'ordre  $p$  représente le nombre de termes auto-régressifs. Formellement, l'équation AR s'exprime comme une somme pondérée des valeurs passées, où les coefficients  $\phi_1, \phi_2, \dots, \phi_p$  déterminent l'importance respective des observations antérieures :

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

#### **3.3.2 Composante d'Intégration (I)**

L'intégration est nécessaire pour rendre la série temporelle stationnaire, c'est-à-dire pour supprimer les tendances et les variations systématiques. L'ordre  $d$  représente le nombre de différenciations nécessaires. L'équation d'intégration consiste à soustraire la valeur actuelle de la valeur précédente, et cela peut être

noté comme :

$$Y'_t = Y_t - Y_{t-1}$$

La série différenciée  $Y'_t$  est utilisée dans le modèle ARIMA à la place de la série originale.

### 3.3.3 Composante à Moyenne Mobile (MA)

La composante à Moyenne Mobile capture les erreurs de prédiction passées, incorporant l'effet des résidus dans la modélisation. L'ordre  $q$  représente le nombre de termes à moyenne mobile. L'équation à Moyenne Mobile peut être formulée comme une somme pondérée des résidus passés, où les coefficients  $\theta_1, \theta_2, \dots, \theta_q$  déterminent la contribution respective des erreurs passées :

$$Y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

### 3.3.4 Application du Modèle ARIMA

L'ensemble du modèle ARIMA est défini par  $ARIMA(p, d, q)$ , où  $p$ ,  $d$ , et  $q$  sont les ordres AR, d'intégration, et MA respectivement. L'identification des ordres optimaux  $p, d, q$  est souvent réalisée en analysant les graphiques de la série temporelle, les fonctions d'autocorrélation partielle et totale, et en utilisant des critères d'information comme l'AIC (Akaike Information Criterion) ou le BIC (Bayesian Information Criterion). Le modèle ARIMA offre une approche robuste pour modéliser les séries temporelles en prenant en compte les tendances, la stationnarité, et les composantes d'erreur, fournissant ainsi des prévisions précises sur la base des données historiques observées. Dans le cadre de modélisation, on analyse les rendements (et pas le prix) du S&P500.

## 3.4 Analyse et Modélisation ARIMA pour les Rendements du S&P500

### 3.4.1 Stationnarité de la Série Temporelle

Afin d'appliquer le modèle ARIMA de manière adéquate, il est crucial de s'assurer que la série temporelle des rendements du S&P500 soit stationnaire. Cela a été confirmé par le test Augmented Dickey-Fuller (Figure 7). Ce test évalue l'hypothèse nulle selon laquelle la série temporelle possède une racine unitaire, ce qui signifie qu'elle n'est pas stationnaire. La p-value extrêmement faible ( $3.95e-29$ ) obtenue indique que nous pouvons rejeter cette hypothèse nulle, validant ainsi la stationnarité de la série temporelle.

### 3.4.2 Choix du Degré du Modèle ARIMA

Le choix des ordres dans le modèle ARIMA est une étape cruciale. Pour ce faire, une combinaison de graphiques d'autocorrélation (ACF) et d'autocorrélation partielle (PACF) a été utilisée (Figure 8). Ces graphiques offrent des indications sur la dépendance temporelle dans les données, orientant ainsi le choix des ordres AR et MA. De plus, des critères d'information tels que l'Akaike Information Criterion (AIC) ont été employés. L'AIC cherche à minimiser la perte d'information lors de la modélisation, favorisant ainsi la simplicité tout en maintenant la qualité des ajustements. Le modèle ARIMA(4, 0, 5) a été retenu comme le meilleur en termes de compromis entre ajustement aux données et simplicité du modèle. De plus comme montre le graphique en annexe (Figure 9) les résidus semblent fluctuer autour d'une moyenne de zéro, ce qui semblerait cohérent avec une. Cependant, il y a des pics indiquant des moments d'instabilité financière : crise de 2009 et crise covid-19. La densité des résidus standardisés est légèrement asymétrique, suivent pas parfaitement une loi normale. Cela pourrait indiquer la présence de mécanismes non linéaires dans les données que le modèle ARIMA ne peut pas capturer. Pour le graphique Q-Q la plupart des points se situent le long de la ligne. Or, on peut aussi voir que la distribution des résidus a des queues plus épaisses que la loi normale, un phénomène courant dans les séries temporelles financières.

Pour le corrélogramme ne montre aucune autocorrélation significative dans les résidus, suggérant que le modèle capture bien la structure temporelle.

### 3.4.3 Résultats du Modèle ARIMA(4, 0, 5)

Suite à l'ajustement du modèle ARIMA(4, 0, 5) aux données des rendements du S&P500, divers résultats sont à noter. Les coefficients obtenus à partir des termes auto-régressifs (AR) indiquent la force et la direction de la corrélation entre la valeur actuelle et les valeurs passées. Les termes à moyenne mobile (MA) capturent la dépendance temporelle dans les erreurs de prédiction du modèle.

Les diagnostics du modèle, visant à évaluer sa robustesse, sont présentés ci-dessous en détail :

Le test de Ljung-Box évalue l'autocorrélation des résidus. La statistique de test est notée Prob(Q) et dans notre cas, elle est de 0.81. Cette valeur suggère une absence d'autocorrélation significative dans les résidus, renforçant ainsi la robustesse du modèle. Les hypothèses nulles et alternatives du test sont respectivement formulées comme suit : Hypothèse nulle ( $H_0$ ) : Les résidus ne présentent pas d'autocorrélation significative jusqu'à un certain retard. Hypothèse alternative ( $H_1$ ) : Les résidus présentent une autocorrélation significative jusqu'à un certain retard.

Le test de Jarque-Bera évalue la normalité des résidus. La statistique de test est notée Prob(JB) et



dans notre cas, elle est de 0.00. Cette valeur indique que les résidus ne suivent pas une distribution normale. Les hypothèses nulles et alternatives du test sont formulées comme suit : Hypothèse nulle (H0) : Les résidus suivent une distribution normale. Hypothèse alternative (H1) : Les résidus ne suivent pas une distribution normale.

Le test d'hétéroscédasticité évalue la constance de la variance des résidus à travers les observations. La statistique de test est notée Prob(H) avec une valeur de 0.00 et H=0.63. Cette valeur indique la présence d'hétéroscédasticité. Les hypothèses nulles et alternatives du test sont formulées comme suit : Hypothèse nulle (H0) : La variance des résidus est constante à travers les observations. Hypothèse alternative (H1) : La variance des résidus n'est pas constante à travers les observations.

Les implications du modèle sont variées. En termes d'ajustement, le modèle semble bien capturer l'autocorrélation, mais présente des problèmes de normalité et d'hétéroscédasticité. Bien que des coefficients significatifs suggèrent un pouvoir prédictif, la prudence est de mise en raison des résidus non normaux et de l'hétéroscédasticité. Il serait également pertinent d'explorer des modèles alternatifs malgré la cohérence des éléments prédictifs du modèle actuel.

Term	Coef	Std Err	$P >  z $
const*	0.0004	0.000	0.05
$\phi_1$ ***	-0.6675	0.207	0.001
$\phi_2$	0.4031	0.223	0.062
$\phi_3$	-0.0094	0.221	0.972
$\phi_4$ ***	-0.5531	0.180	0.001
$\theta_1$ ***	0.5414	0.209	0.008
$\theta_2$ **	-0.4653	0.203	0.016
$\theta_3$	0.0634	0.221	0.771
$\theta_4$ ***	0.5067	0.161	0.001
$\theta_5$ ***	-0.0782	0.017	0.000
sigma2***	0.0002	$1.49 \times 10^{-6}$	0.000

Table 2: Coefficients d'un ARIMA(4, 0, 5)

Le modèle ARIMA(4, 0, 5) appliqué à la série temporelle présente des résultats significatifs qui peuvent être interprétés comme suit :

La constante est statistiquement significative, indiquant un effet significatif au décalage zéro dans le modèle.

Les valeurs  $\phi_1$  et  $\phi_4$  sont négatives et significatives, suggérant une forte influence négative du premier et du quatrième délai sur la série temporelle. En revanche, les coefficients  $\phi_2$  et  $\phi_3$  ne sont pas statistiquement significatifs.

Les valeurs  $\theta_1$  et  $\theta_4$  montrent une influence positive significative de l'erreur sur la série temporelle. Les coefficients  $\theta_2$  et  $\theta_5$  présentent une influence négative significative, tandis que le coefficient  $\theta_3$  n'est pas significatif.

Les termes d'erreur faibles ( $\sigma^2$ ) indiquent que le modèle ARIMA(4, 0, 5) prédit relativement bien les valeurs de la série temporelle, avec une bonne adéquation aux données observées.

Le modèle identifié semble être approprié pour modéliser la série temporelle, avec des coefficients significatifs pour les délais AR et MA, et une constante ayant un effet même si de faible amplitude. Les termes d'erreur faibles suggèrent une bonne capacité du modèle à capturer la variabilité des données observées.

#### 3.4.4 Prévisions

La Table 5 (en annexe), intitulée "Prévisions du rendement", offre des aperçus détaillés des projections de rendements financiers pour des dates spécifiques à venir. Chaque ligne de la table correspond à une date donnée, fournissant des informations cruciales pour les investisseurs et les analystes financiers.

La colonne "Forecast" présente les prévisions de rendement pour chaque date spécifiée. Par exemple, pour le 26 janvier 2024, la prévision est de 0.000168. Les colonnes "Lower CI" (Limite inférieure de l'intervalle de confiance) et "Upper CI" (Limite supérieure de l'intervalle de confiance) indiquent respectivement les bornes inférieure et supérieure de l'intervalle de confiance associé à chaque prévision. Pour la date du 26 janvier 2024, ces limites sont -0.024885 et 0.024890.

La colonne "Variation (%)" exprime la différence en pourcentage entre la limite supérieure et la limite inférieure de l'intervalle de confiance. Une variation plus faible suggère une plus grande confiance dans la projection. Par exemple, pour le 26 janvier 2024, la variation est de 0.016842%. De plus les prévisions fournissent une estimation du rendement futur, tandis que les intervalles de confiance délimitent la fourchette probable du rendement réel. Un intervalle de confiance plus étroit indique une confiance accrue dans la projection. Ces prédictions de rendements sont représentées par la Figure 3 ainsi que l'intervalle de confiance pour chaque jour.

En regardant les données (Table 5) ainsi que la Figure 3, on remarque rapidement que l'intervalle de confiance est beaucoup trop large comparé à nos prédictions : prédictions autour de  $10^{-4}$ , bornes de l'intervalle au niveau de  $10^{-2}$ . Ces prédictions sont trop imprécises: on ne sait pas si on aura des gains ou des pertes alors un enthousiaste de la finance ne sera pas intéressé par ce modèle ni ces résultats.

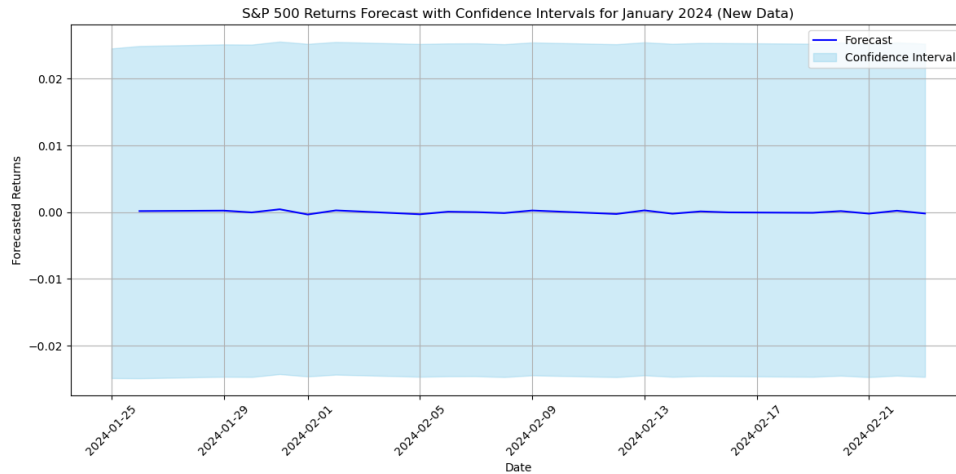


Figure 3: Rendemenets prédits

### 3.4.5 Conclusion modèle ARIMA

Le modèle ARIMA(4, 0, 5) se révèle être une bonne base pour modéliser les rendements du S&P500 et un bon modèle de référence en général. Les avantages qu'on peut tirer de ce modèle se trouvent principalement dans la facilité à obtenir des résultats numériques et des prédictions cohérentes et "tangibles", qui rendent le modèle ARIMA un bon modèle introductif dans cette étude.

Néanmoins, il faut souligner que même si les résultats sont cohérents, le modèle ARIMA est très imprécis, dû largement à son incapacité de capturer des relations avec des erreurs hétéroscédastiques et la distribution non-normale des observations. Ceci met en évidence les limites du modèle ARIMA

### 3.5 Modèle HMM

Le modèle Hidden Markov Model (HMM), ou modèle de Markov caché, est une approche probabiliste puissante utilisée pour modéliser des systèmes dynamiques où les variables observées résultent de processus complexes et non observables. Appliqué à la prédiction financière, le HMM permet d'analyser les séquences temporelles des marchés en identifiant des régimes cachés qui influent sur les comportements observables. Cette méthode offre ainsi une perspective probabiliste sur l'évolution des marchés, permettant des prédictions basées sur des transitions entre états cachés.

1. **États Cachés ( $S_t$ ):** Les états cachés forment une séquence discrète à travers le temps, notée comme  $S = \{S_1, S_2, \dots, S_T\}$ , où  $T$  est le nombre total d'étapes temporelles.
2. **Observations ( $O_t$ ):** Les observations sont générées à partir des états cachés, notées comme  $O = \{O_1, O_2, \dots, O_T\}$ . Chaque observation est conditionnée par l'état caché correspondant.
3. **Matrice de Transition ( $A$ ):** La matrice de transition  $A$  spécifie la probabilité de transition entre les différents états cachés. Si  $a_{ij}$  est la probabilité de passer de l'état caché  $i$  à l'état caché  $j$ , alors la matrice de transition est définie comme :

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{bmatrix}$$

4. **Matrice d'Émission ( $B$ ):** La matrice d'émission  $B$  spécifie la probabilité d'observer une certaine valeur à partir d'un état caché particulier. Si  $b_{ij}$  est la probabilité d'observer  $O_j$  à partir de l'état caché  $i$ , alors la matrice d'émission est définie comme :

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1M} \\ b_{21} & b_{22} & \dots & b_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ b_{N1} & b_{N2} & \dots & b_{NM} \end{bmatrix}$$

Ici,  $M$  est le nombre d'observations possibles.

5. **Distribution Initiale ( $\pi$ ):** La distribution initiale  $\pi$  spécifie la probabilité initiale d'être dans chaque état caché. Si  $\pi_i$  est la probabilité initiale d'être dans l'état caché  $i$ , alors  $\pi = \{\pi_1, \pi_2, \dots, \pi_N\}$ .

Les équations d'algorithme de Forward ( $\alpha$ ) et Backward ( $\beta$ ) sont utilisées pour calculer la probabilité d'observation totale et estimer les paramètres du modèle. La formule de Bayes est appliquée pour mettre à jour les probabilités des états cachés, et l'algorithme de Viterbi est utilisé pour trouver la séquence d'états cachés la plus probable.

### 3.6 Application et Analyse du HMM pour le Cours du S&P500

#### 3.6.1 Ajustement du modèle

Pour utiliser un modèle optimal, on va chercher à trouver le nombre d'états qui maximise la log-vraisemblance tout en minimisant l'AIC pour trouver le modèle le mieux adapté à nos données.

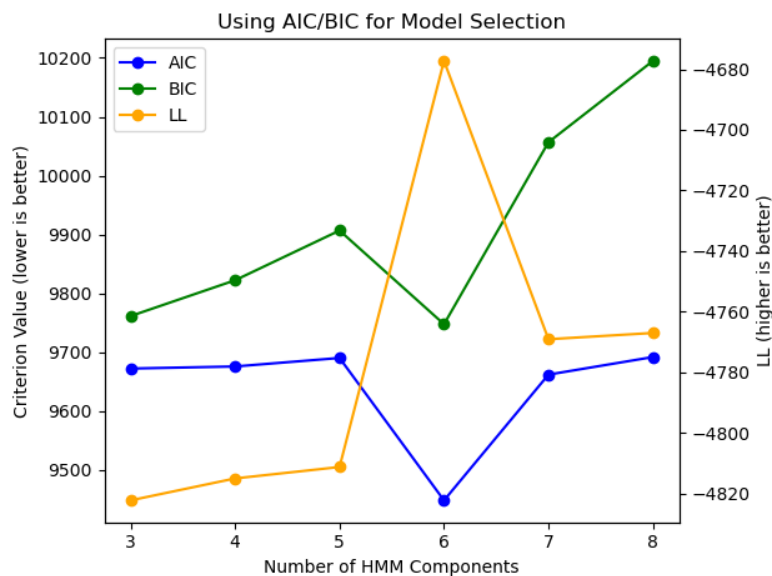


Figure 4: Critères d'Ajustement en fonction du Nombre d'états

Ici, on a trouvé que le nombre d'états optimal est 6 en optimisant les critères d'ajustement suivants:

- AIC minimal, inférieur à -4820
- BIC minimal d'environ -4760
- Log-Vraisemblance maximale d'environ 4680

Or, un bon ajustement aux données n'est pas forcément le seul critère pour choisir un modèle.

### 3.6.2 Généralité du modèle

En s'agissant une prédiction pour des données futures, une très bonne qualité d'ajustement aux données n'est pas suffisant pour construire le meilleur modèle.

On construit 3 nouvelles variables :

1. Changement Fractionnel des Prix d'Ouverture et de Clôture (fracocp) : Ce paramètre reflète le changement relatif du prix de l'action entre l'ouverture et la clôture de la journée de trading. C'est un indicateur direct de la performance de l'action ce jour-là.

$$\frac{P_{Close} - P_{Open}}{P_{Open}}$$

2. Changement Fractionnel des Prix les Plus Élevés (frachp) : Cela mesure le changement relatif dans le prix le plus élevé de l'action sur une période spécifique. Cela aide à comprendre la performance maximale de l'action en trading. Les prix les plus élevés sont souvent associés au niveau de résistance de l'action - un point où l'intérêt de vente surpasse la pression d'achat et le prix commence à baisser à nouveau.

$$\frac{P_{High} - P_{Open}}{P_{Open}}$$

3. Changement Fractionnel des Prix les Plus Bas (fraclp) : Cela suit le changement relatif dans le prix le plus bas de l'action sur une période donnée. Les prix les plus bas sont liés aux niveaux de support - un niveau de prix où on peut s'attendre à une pause de la tendance baissière en raison d'une concentration de la demande. Comprendre ces niveaux peut être crucial pour identifier les renversements potentiels dans la tendance des prix de l'action.

$$\frac{P_{Low} - P_{Open}}{P_{Open}}$$

Cette approche et méthode sont basées sur la méthode de prédiction utilisée par Ahmad A Fayad Chowdury et MD Amimul Ehsan :

<https://www.kaggle.com/code/ehsanamim/stock-market-prediction-using-hmm>

Cet algorithme se base sur deux fonctions: "augment\_features" qui crée ces trois variables en fonction d'un dataframe "transitoire". Puis la fonction "extract\_features" qui extrait ces trois variables du dataframe transitoire.

Ces valeurs extraites sont après gardées dans une matrice, nommons là "K". D'après le modèle, on attribue à ces valeurs une probabilité d'arriver, et sont rangés dans un ordre "plus probable" à "moins probable". À chaque jour, on attribue un prix de clôture qui est donné par

$$P_t^{close} = P_t^{open} * (1 + \max_i \{K_{i,j,k}\})$$

On va ainsi échantillonner nos données en deux groupes:

- Le premier 80% des données pour entrainer le modèle
- Le 20% des données plus récentes pour le tester

Ainsi, on peut évaluer la capacité du modèle à prédire les prix futures avec des données passées, respectant l'ordre chronologique des observations des séries temporelles. En minimisant le MAE, on peut choisir le modèle avec les meilleures capacités de prédiction.

$$\min \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

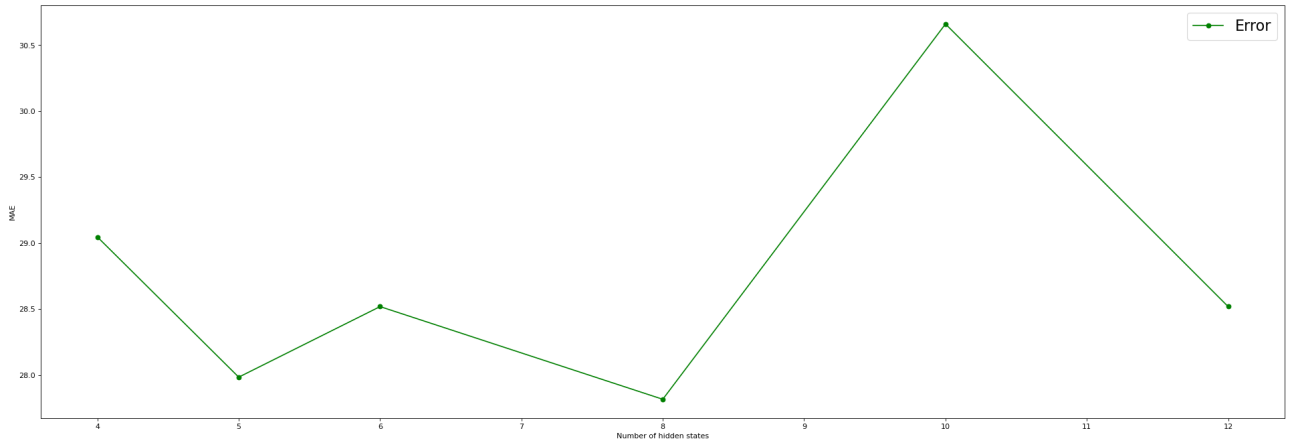


Figure 5: MAE en fonction du nombre d'Etats

On va donc garder un modèle avec 8 états cachés présentant une erreur moyenne absolue inférieure à 28. Maintenant qu'on a le modèle optimal, mettons-le à preuve.

Enfin, pour évaluer le modèle HMM avec 8 états cachés, on va comparer les prédictions du cours du S&P500 faites par le modèle avec les vraies valeurs du cours pendant ces jours. Le modèle sera évalué sur les prix du S&P500 entre Août 2020 et Novembre 2021.

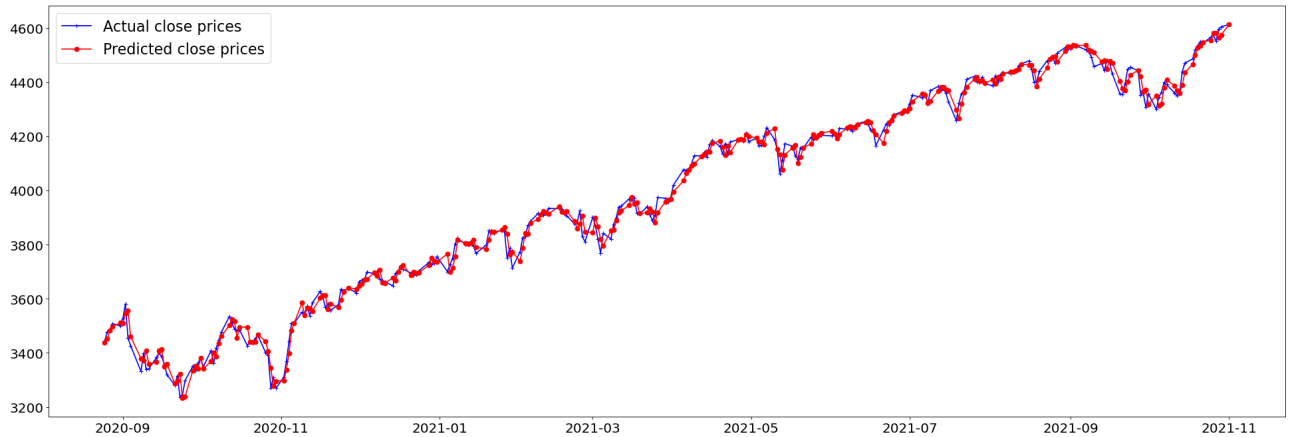


Figure 6: Comparaison entre Prix de Clôture Réels et Prédits

Comme illustré dans le Graphique, les prédictions présentées par le modèle sont assez proches des vrais prix de clôture du S&P500 pendant cette période. En s'ajoutant a ceci, on a obtenu un MAE de 21.352 ce qui est très faible, particulièrement comparé au MAE minimal précédent ( $\approx 27.5$ ).



### 3.6.3 Etats Cachés

En obtenant ce modèle optimal, on peut donc présenter les informations associées à chacun des huit états cachés, ainsi qu'à la matrice de transition des états (Annexe).

État	Moyenne			Variance		
	fracocp	frachp	fraclp	fracocp	frachp	fraclp
0	-0.0388	0.0056	0.0548	0.0016	0.0012	0.0016
1	0.0118	0.0155	0.0033	$9.58 \times 10^{-5}$	$6.83 \times 10^{-5}$	$4.03 \times 10^{-5}$
2	0.1001	0.1023	0.0034	0.0050	0.0049	0.0048
3	0.0233	0.0333	0.0138	0.0007	0.0004	0.0003
4	-0.0667	0.0039	0.0725	0.0021	0.0018	0.0021
5	-0.0260	0.0089	0.0359	0.0006	0.0003	0.0005
6	0.0009	0.0041	0.0038	$2.36 \times 10^{-5}$	$1.46 \times 10^{-5}$	$1.51 \times 10^{-5}$
7	-0.0140	0.0022	0.0182	$8.28 \times 10^{-5}$	$3.09 \times 10^{-5}$	$6.87 \times 10^{-5}$

Table 3: Résumé des moyennes et variances pour chaque état du HMM

### 3.6.4 Prédiction

En appliquant la capacité prédictive du modèle, on peut obtenir une unique prédiction: le prix de clôture du S&P 500 pour le 25 Janvier 2024 (1 jour après la dernière date observée dans la base de données).

Le modèle prédit un prix de \$4891.28, soit une hausse de 0.467%. Dans la réalité, le prix de clôture du 25 Janvier 2024 était \$4894.16, c'est à dire une hausse de 0.526%. Ainsi, notre modèle a sous estimé le cours du S&P 500 de \$2.88 ou la variation du cours de 0.095%.

## 4 Conclusion

En conclusion, l'application des modèles Hidden Markov (HMM) et AutoRegressive Integrated Moving Average (ARIMA) à la prédiction du cours du S&P500 s'est avérée être une entreprise intéressante.

Le modèle ARIMA, en prenant en compte les aspects auto-régressifs, d'intégration, et à moyenne mobile, a démontré être un outil pratique, cohérent et très illustratif pour prévoir les rendements du S&P500 malgré ses limites et imprécisions.

Le modèle HMM, en identifiant et en modélisant les régimes cachés du marché, a apporté une perspective probabiliste riche qui a contribué à anticiper les transitions entre différents états du marché. De plus, grâce aux changements fractionnaires de prix d'ouverture et clôture, maximales, et minimales on a réussi à construire un modèle de prédiction précis et robuste.

On peut aussi voir que le HMM est beaucoup plus précis en travaillant avec des séries financières que le modèle ARIMA. D'où il faut remarquer que le fait que le modèle HMM ne puisse produire qu'un unique résultat met en évidence la nature des marchés financiers. La prévision d'actifs risqués comme le S&P 500 n'a aucun intérêt dans le long-terme dû à la haute volatilité associée à ces actifs, peu importe la qualité du modèle. Les prévisions à moyen/long terme sont beaucoup plus intéressantes pour des actifs très stables comme les obligations, or pour des actifs risqués comme le S&P 500, des outils de prévisions journalières sont adaptés aux dynamiques de marchés et intéressants pour les enthousiastes de la finance.

Cette approche progressive a permis de voir les complexités du comportement du marché financier ainsi que les différentes capacités des modèles statistiques utilisés. Ces résultats servent comme une introduction dynamique et compréhensive au monde des prévisions du marché financier.

## 5 Bibliographie

1. Box, G. E., Jenkins, G. M., & Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. Prentice Hall.
2. Bodie, Z., Kane, A., & Marcus, A. J. (2009). *Investments*. McGraw-Hill.
3. Cao, L. J., & Tay, F. E. H. (2001). *Financial forecasting using support vector machines*. *Neural Computing and Applications*, 10(2), 184-192.
4. Guerard Jr., J. B. (2013). *Statistical Arbitrage: Algorithmic Trading Insights and Techniques*. John Wiley & Sons.
5. Guo, J., He, Q., & Zhang, L. (2014). *Combining AdaBoost and Support Vector Machines for financial trading prediction*. *Neurocomputing*, 139, 28-41.
6. Kim, K. J. (2003). *Financial time series forecasting using support vector machines*. *Neurocomputing*, 55(1-2), 307-319.
7. Kim, K. J., & Han, I. (2006). *Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index*. *Expert Systems with Applications*, 31(2), 241-247.
8. Krauss, C., Do, X. A., Huck, N., & Ruenzi, S. (2017). *Machine learning in finance: Status quo and prospects*. *Journal of Financial Stability*, 29, 36-48.
9. Leigh, W., Modani, N., Purvis, R., & Roberts, H. (2002). *A comparison of neural networks and linear scoring models in the prediction of bankruptcy*. *Expert Systems with Applications*, 23(3), 321-331.
10. Rabiner, L. R. (1989). *A tutorial on hidden Markov models and selected applications in speech recognition*. *Proceedings of the IEEE*, 77(2), 257-286.
11. Al Fayad Chowdury, A., Ehsan, M. D. Al, & G. C. (2022). *Stock market prediction using HMM*, kaggle.com
12. Kim, M. J., & Han, I. (2003). *The Discovery of Experts' Decision Rules from Qualitative Bankruptcy Data Using Genetic Algorithms*, *Expert Systems with Applications*, 25(4), 637-646.
13. Tay, F. E. H., & Cao, L. J. (2001). *Improved financial time series forecasting by combining Support Vector Machines with self-organizing feature map*. *Intelligent Data Analysis*, 5, 339-354.

## 6 Annexe

---

Test Statistic	-1.621520e+01
p-value	3.949040e-29
#Lags Used	1.700000e+01
Number of Observations Used	4.272000e+03
Critical Value (1%)	-3.431882e+00
Critical Value (5%)	-2.862217e+00
Critical Value (10%)	-2.567130e+00
dtype:	float64

---

Figure 7: Test ADF

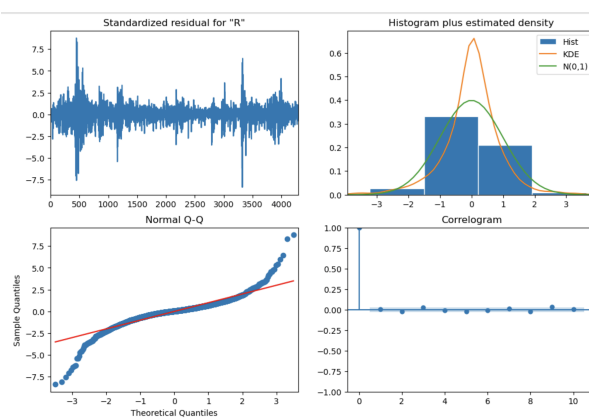


Figure 8: Diagnostics

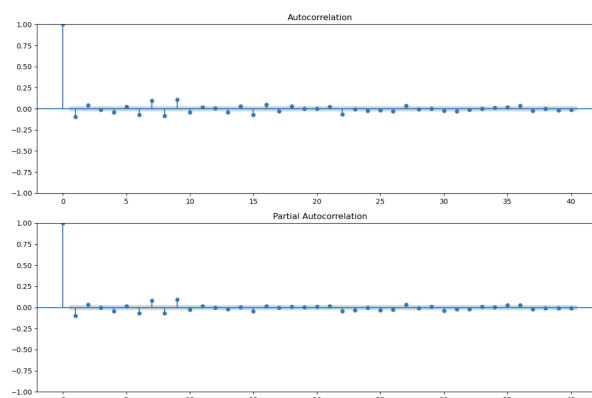


Figure 9: ACF et PACF

NB: La matrice de Transition des états est beaucoup trop grande pour ce document alors on vous présente l'output du code Python:

Matrice de Probabilités de Transition : [[1.70454978e-03 1.48476803e-15 3.85942423e-05 2.84482126e-03 4.11268706e-07 9.95411500e-01 1.23526598e-07 1.81848970e-14] [3.99723241e-03 3.11934427e-01 4.78191513e-16 2.26226227e-06 3.20012275e-13 2.50760945e-14 3.67741230e-01 3.16324848e-01] [2.78405152e-20 1.85580673e-16 3.83516216e-07 5.01750925e-07 9.21930593e-14 9.99993663e-01 2.80633395e-18 5.45168261e-06] [1.04368711e-01 1.68978921e-01 3.43289114e-06 2.42927338e-01 6.02067777e-05 4.83660076e-01 3.26660225e-08 1.28204622e-06] [8.76780419e-04 1.42176958e-14 2.04456630e-06 9.99121119e-01 5.55674483e-08 4.93703978e-11 2.07248931e-24 7.06119305e-21] [2.16148985e-04 8.12260744e-10 3.83073285e-02 8.20150169e-01 1.03331918e-01 2.19786172e-02 6.14534028e-15 1.60158170e-02] [2.81563356e-10 4.54884115e-02 4.89922319e-22 1.21564100e-12 1.30522391e-11 1.00765266e-05 8.66381851e-01 8.81196605e-02] [1.93652897e-07 4.48794583e-01 7.47013398e-08 1.01957344e-04 5.91467135e-11 2.17566392e-02 2.89704475e-01 2.39642078e-01]]

Term	Coef	Std Err	$P >  z $
const*	0.0004	0.000	0.05
$\phi_1$ ***	-0.6675	0.207	0.001
$\phi_2$	0.4031	0.223	0.062
$\phi_3$	-0.0094	0.221	0.972
$\phi_4$ ***	-0.5531	0.180	0.001
$\theta_1$ ***	0.5414	0.209	0.008
$\theta_2$ **	-0.4653	0.203	0.016
$\theta_3$	0.0634	0.221	0.771
$\theta_4$ ***	0.5067	0.161	0.001
$\theta_5$ ***	-0.0782	0.017	0.000
sigma2***	0.0002	$1.49 \times 10^{-6}$	0.000

Table 4: Coefficients d'un ARIMA(4, 0, 5)

Dep. Variable:	Returns		No. Observations:		4290	
Model:	ARIMA(4, 0, 5)		Log Likelihood		12675.654	
Date:	Mon, 22 Jan 2024		AIC		-25329.307	
Time:	15:45:52		BIC		-25259.303	
Sample:	0		HQIC		-25304.581	
			- 4290			
Covariance Type:			opg			
	coef	std err	z	P> z	[0.025	0.975]
const	0.0004	0.000	1.960	0.050	5.08e-08	0.001
ar.L1	-0.6704	0.203	-3.309	0.001	-1.067	-0.273
ar.L2	0.4004	0.214	1.867	0.062	-0.020	0.821
ar.L3	-0.0072	0.206	-0.035	0.972	-0.410	0.396
ar.L4	-0.5479	0.171	-3.198	0.001	-0.884	-0.212
ma.L1	0.5387	0.204	2.644	0.008	0.139	0.938
ma.L2	-0.4663	0.194	-2.401	0.016	-0.847	-0.086
ma.L3	0.0601	0.207	0.291	0.771	-0.345	0.465
ma.L4	0.4991	0.153	3.271	0.001	0.200	0.798
ma.L5	-0.0809	0.017	-4.672	0.000	-0.115	-0.047
sigma2	0.0002	1.49e-06	106.363	0.000	0.000	0.000
Ljung-Box (L1) (Q):			0.06	Jarque-Bera (JB):		20232.64
Prob(Q):			0.81	Prob(JB):		0.00
Heteroskedasticity (H):			0.63	Skew:		-0.51
Prob(H) (two-sided):			0.00	Kurtosis:		13.59

Figure 10: Coefficients ARIMA

Table 5: Prévisions du rendement

Date	Forecast	Lower CI	Upper CI	Variation (%)
2024-01-25	nan	-0.024858	0.024526	nan
2024-01-26	0.000168	-0.024885	0.024890	0.016842
2024-01-29	0.000239	-0.024652	0.025135	0.023929
2024-01-30	-0.000031	-0.024684	0.025106	-0.003084
2024-01-31	0.000439	-0.024255	0.025555	0.043882
2024-02-01	-0.000342	-0.024598	0.025213	-0.034200
2024-02-02	0.000273	-0.024334	0.025496	0.027309
2024-02-05	-0.000307	-0.024650	0.025198	-0.030718
2024-02-06	0.000076	-0.024577	0.025275	0.007561
2024-02-07	0.000012	-0.024568	0.025291	0.001231
2024-02-08	-0.000126	-0.024694	0.025165	-0.012591
2024-02-09	0.000258	-0.024436	0.025424	0.025819
2024-02-12	-0.000265	-0.024703	0.025161	-0.026501
2024-02-13	0.000275	-0.024432	0.025441	0.027532
2024-02-14	-0.000223	-0.024660	0.025222	-0.022337
2024-02-15	0.000120	-0.024544	0.025345	0.011975
2024-02-16	-0.000026	-0.024572	0.025321	-0.002597
2024-02-19	-0.000085	-0.024656	0.025236	-0.008458
2024-02-20	0.000168	-0.024488	0.025405	0.016841
2024-02-21	-0.000212	-0.024701	0.025193	-0.021248
2024-02-22	0.000225	-0.024479	0.025420	0.022486
2024-02-23	-0.000191	-0.024672	0.025233	-0.019052