# Techniques de prévision des rendements financiers

#### FROEHLICH Marc MIRANDA André

Université de Strasbourg

31/01/2024

Introduction

Introduction

2 Revue de la littérature

3 Analyse de données

Modèles

Conclusion

# Importance du sujet

- Comparaison d'outils prédictifs dans les marchés fianaciers
- Objectif: créer un outil de prévision d'actifs financiers
- Modèles utilisés : ARIMA, HMM

- Les multiples rôles des prévisions dans la prise de décisions (Krauss et al., 2017)
- Les fondements de l'anayse techniques et fondamentales (Leigh et al., 2002)
- Remise en question par l'Hypothèse de marché Efficent (HME).
- Stratégie de Prédiction dans la Litérrature (Guo et al. (2014)), (Guo et al., 2014; Kim, 2003), (Kim et al., 2004; Kim, Min, Han, 2006)

#### Collecte des données

- Les données sont obtenues sur le site de Yhaoo Finance en utilisant la fonction getSymbols du package quantmod.
- Les données sont étudiées sur la période 2007 à 2024.
- Variables extraites : OPEN, HIGH, LOW, CLOSE, VOLUME, ADJUSTED CLOSE, RETURNS
- Pour calculer les rendements, nous avons utilisé la fonction CalculateReturns du package PerformanceAnalytics.

# Statistiques descriptives

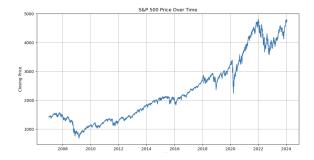


Figure 1: Cours du S&P500

# Statistiques descriptives

Statistique	Valeur
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

Figure 2: Statistiques descriptives

# Statistiques descriptives

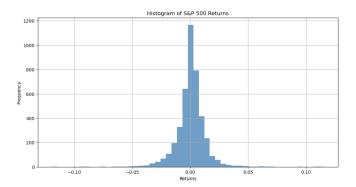


Figure 3: Histogramme

Modèles

#### ARIMA

- Composante Auto-Régressive (AR) :
- $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon$
- AR exprime comment Y est influencé par ses observations antérieures avec des poids déterminés

#### • Composante d'Intégration (I)

• 
$$Y'_t = Y_t - Y_{t-1}$$

#### Anima

- Composante à Moyenne Mobile (MA)
- $Y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$
- Modélise les variations non systématique des données

#### Stationnarité

Table 1: Tests de stationnarité

Variable	Test	Statistique	P_value
$S$ $\mathcal{E}$ $P$ $500$	ADF	-4,40774835	$3.94040 \times 10^{-29}$

# Résulats et Interprétations

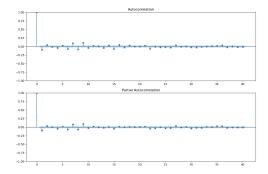


Figure 4: ACF, PACF

- Le test de Ljung-Box révèle un Prob(Q) de 0.81, suggérant une absence d'autocorrélation
- Jarque-Bera indique un Prob(JB) de 0.00, signalant que les résidus ne suivent pas une distribution normale
- $\bullet$  le test d'hétéroscédasticité affiche un Prob(H) de 0.00 avec H=0.63, présence d'hétéroscédasticité

# Diagnostics

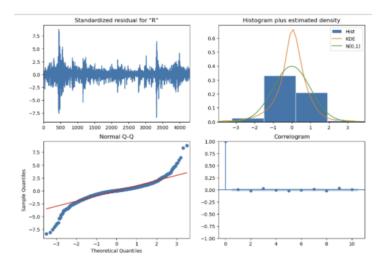


Figure 5: Diagnostic

# Résultats et interprétations

- Les résidus semblent fluctuer autour d'une movenne de zéro. 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.
- 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.
- Le corrélogramme ne montre aucune autocorrélation significative dans les résidus, suggérant que le modèle capture bien la structure temporelle.

# Coefficients associés au modèle optimal : AR(4); I(0); MA(5)

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	$1.49\times10^{-6}$	0.000

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

## Interprétation des coefficients

- Constante : faible mais significative
- Coefficients AR
  - $\phi_1$  et  $\phi_4$  indiquent une forte influence négative et significative du premier et du quatrième délai.
  - $\phi_2$  et  $\phi_3$  n'ont pas d'effets statistiquement significatifs
- Coefficients MA
  - $\theta_1$  et  $\theta_4$  montrent une influence positive significative de l'erreur.
  - $\theta_2$  et  $\theta_5$  présentent une influence négative significative
  - $\theta_3$  non significative
- sigma2: termes d'erreur faibles indiquant que le modèle prédit relativement bien les valeurs de la série

# Résulats et Interprétations

Table 3: 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

Figure 6: prévisions des rendements

#### • 2024-02-01 :

- Forecast : La prévision est de -0.000342, indiquant une légère baisse du rendement.
- Lower CI: -0.024598 et Upper CI: 0.025213 délimitent l'intervalle de confiance.
- Variation : La variation de -0.034200 suggère une baisse significative du rendement par rapport à la date précédente.

# Rendements prédits

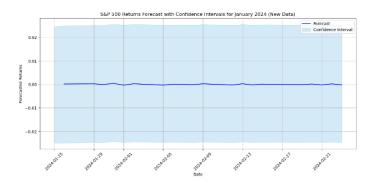


Figure 3: Rendmenets prédits

Figure 7: Rendements prédits

- États Cachés  $(S_t)$ : Les états cachés forment une séquence discrète à travers le temps, notée comme  $S = \{S_1, S_2, ..., S_T\}$ , où T est le nombre total d'étapes temporelles.
- Observations  $(O_t)$ : Les observations sont générées à partir des états cachés, notées comme  $O = \{O_1, O_2, ..., O_T\}$ . Chaque observation est conditionnée par l'état caché correspondant.

• 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:

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

### HMM

ullet 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  $O_i$  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.

#### HMM

- 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  $\pi = \{\pi_1, \pi_2, ..., \pi_N\}$ .
- Les équations de l'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.

# Ajustement du modèle

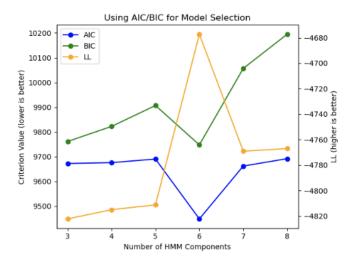


Figure 8: Critères d'Ajustement en fonction du nombre d'états

#### Construction de Nouvelles Variables

- 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.
- 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.
- Ochangement 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.

#### Construction de Nouvelles Variables

• Formule pour fracocp:

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

• Formule pour frachp:

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

• Formule pour fraclp:

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

#### MAE en fonction du nombre d'états cachés

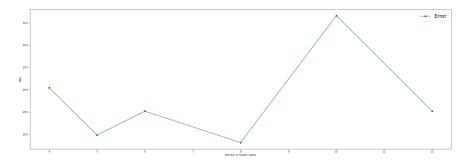


Figure 9: États cachés

### Evaluation du modèle

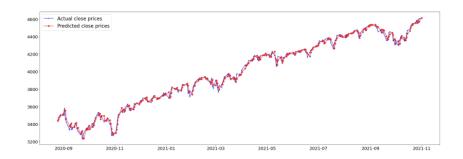


Figure 10: Comparaison entre Prix de cloture réel et prédits

MAE = 21.352

#### Résumé des Etats Cachés

État	Moyenne			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	$9.58 \times 10^{-5}$	$6.83  imes 10^{-5}$	$4.03  imes 10^{-5}$
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	$8.28 \times 10^{-5}$	$3.09\times10^{-5}$	$6.87\times10^{-5}$

Figure 11: États cachés

- Une seule et unique prédiction produite
- Prédiction pour le 25/01/2024 : prix de \$4891.28 à la clôture, soit une hausse de 0.467%
- Réalité pour le 25/01/2024 : \$4894.16, c'est à dire une hausse de 0.526%
- Le modèle a sous-estimé le prix de \$2.88 soit, ila sous-estimé la hausse de 0.095%.

Notre modèle a sous estimé le cours du S&P 500 de \$2.88 soit 0.00.

#### Conclusions

- Facilité à utiliser ARIMA (pas du tout précis)
- HMM très pertinent pour ce type de séries.
- Prédictions limitées, en concordance avec la nature des actifs risqués: aucun intérêt à prédire le cours d'un actifs risqué pour le prochain an (complètement irréaliste)
- Poursuite d'améliorations (SVM)