



UNIVERSITÉ PARIS 1 PANTHÉON-SORBONNE

PROJET D'ÉCONOMÉTRIE FINANCIÈRE

Application du modèle de trading de paires basé sur des modèles d'espace d'état linéaires et le filtre de Kalman

Diakité Gaoussou, Koffi Eunice, Ababii Anisoara

Encadré par
Professeur MERHY Chafic

May 3, 2023

Contents

1	Résumé de l'article	2
2	Application	3
2.1	Implémentation de la stratégie sur les paires de titres de l'articles	3
2.1.1	Les données	3
2.1.2	Test de cointégration en deux étapes	4
2.1.3	Vérification de la stationnarité des résidus :	5
2.1.4	Calcule du spread	6
2.1.5	Modélisation du Spread avec les modèles d'espaces d'état : Modèle AR (1), Modèle AR (2), Modèle ARM(1,1)	6
2.1.6	L'estimation du spread entre les deux actifs à l'aide d'un modèle d'espace d'état Kalman Filter	7
2.1.7	Génération du signal de trading :	7
2.1.8	Backtesting du modèle :	7
2.2	Implémentation de la stratégie sur nos paires de titres	8
2.2.1	Les données	8
2.2.2	Test de cointégration en deux étapes	9
2.2.3	Vérification de la stationnarité des résidus :	10
2.2.4	Calcule du spread	11
2.2.5	Modélisation du spread par le modèle d'espace d'état et filtre de kakman et le résultat de la performance	12
3	Conclusion	12

Introduction

Le trading de paires est une stratégie de trading bien connue qui implique l'achat et la vente simultanés de deux actifs étroitement liés. Cette stratégie repose sur l'hypothèse que les deux actifs évoluent dans une relation linéaire et que tout écart par rapport à cette relation sera temporaire. L'article "A pairs trading strategy based on linear state space models and the Kalman filter", écrit en 2010 par propose une méthode de trading de paires basée sur des modèles d'espace d'état linéaires et le filtre de Kalman pour estimer la relation linéaire entre les actifs, détecter les écarts et opter pour une stratégie.

Dans ce projet, nous allons explorer cette méthode de trading de paires et l'appliquer à un ensemble de données financières. Nous commencerons donc par présenter l'article et résumer les principaux résultats et méthodes proposées. Ensuite, nous appliquerons la méthode à des données financières et évaluerons sa performance. Enfin, nous discuterons des résultats et des limites de la méthode et proposerons des perspectives d'amélioration pour de futures recherches dans ce domaine.

1 Résumé de l'article

Notre étude est principalement fondée sur l'article "A pairs trading strategy based on linear state space models and the Kalman filter", écrit par Carlos Eduardo de Moura, Adrian Pizzinga et Jorge Zubelli et publié le 25 Avril 2016.

Selon l'article, la stratégie de pair trading est une stratégie qui consiste à considérer deux actions qui ont une relation de cointégration sur le long terme. En général, ce sont des titres qui évoluent sur un même marché donc sont soumis aux mêmes fluctuations. Puis, en considérant leurs fluctuations mutuelles, adopter une stratégie Long-Short à partir d'une stratégie d'arbitrage propre à elle.

Les auteurs aborderont respectivement des approches théoriques, empiriques et critiques.

Ainsi, nous avons eu droit dans la première partie de l'article, à une revue de littérature riche mais pas exhaustive des études qui ont été menées sur le sujet en mettant en exergue les méthodes utilisées par les auteurs. Ils apporteront par la suite une plus claire précision sur l'apport de leur travaux sur le sujet (cf. partie 2).

Puis, d'un point de vue statistique (logique et cohérence), ils nous présentent les différentes possibilités d'arbitrage ainsi que des modèles qui peuvent être utilisés dans l'application de cette stratégie tout en discutant des avantages et inconvénients de chacun d'eux

Après avoir fait un tour de table sur tous les outils indispensables à l'application de la stratégie, les auteurs mettent en pratique leur stratégie pour apporter un contenu consistant à leurs travaux.

Pour ce faire ils travaillent sur des deux couples de titres *XOM_LUV* et *VALE5_BRAP4* sur une période du 22/09/2011 au 26/03/2013. Les étapes mises en place par leur stratégie est donc la suivante :

- Preuve de la cointégration des titres en deux étapes
- Calcul du spread
- Modélisation du spread
- Application du filtre de Kalman
- Calcul des probabilités conditionnelles de retour à la moyenne de long terme du spread
- Définition de la stratégie de pair : A un instant t , on calcule le spread entre les deux titres, puis en fonction de la probabilité de retour à la moyenne de long terme, on prend une position qui est de vendre l'actif le plus cher et d'acheter le moins cher sous réserve que les coûts de transaction ne dépasse pas les gains.
- Evaluation de la méthode par une comparaison avec une autre méthode et une comparaison avec les performances des indices de référence que sont : Libor, S&P500, CDI, Ibovespa.

Avec cette stratégie, ils ont pu battre le ratio de sharpe des indices de référence choisis.

2 Application

Dans cette section, nous allons appliquer la stratégie de trading pairs proposée dans l'article aux données que nous avons collectées. Nous allons d'abord tester la performance de la stratégie sur les données d'entraînement, puis évaluer son efficacité sur les données de test en comparant ses performances à celles des indices de référence. Enfin, nous allons appliquer la stratégie à de nouveaux titres et évaluer ses performances.

2.1 Implémentation de la stratégie sur les paires de titres de l'articles

2.1.1 Les données

Nous avons collecté des données sur deux paires de titres : XOM (Exxon Mobil Corporation) et LUV (Southwest Airlines Co.), ainsi que deux indices de référence : BVSP (Bovespa Index) et GSPC (SP 500 Index). Les données ont été collectées à partir de Yahoo Finance.

Il convient également de préciser que les données utilisées pour l'analyse ont été préalablement sélectionnées et nettoyées pour correspondre à la période d'étude souhaitée. Plus précisément, nous avons sélectionné les données pour les titres XOM et LUV ainsi que pour les indices de référence BVSP et GSPC sur la période allant du 22 septembre 2011 au 20 septembre 2012 pour former notre ensemble de données d'apprentissage. Les données pour la période suivante, allant du 21 septembre 2012 au 26 mars 2013, ont été utilisées pour tester la performance de notre stratégie.

Nous avons également séparé les données en deux parties: l'échantillon et les données hors échantillon. Les données d'apprentissage (ou l'échantillon) ont été utilisées pour estimer les paramètres du modèle, tandis que les données hors échantillon ont été utilisées pour tester la performance de notre stratégie. Les données d'apprentissage et hors échantillon ont été définies à l'aide des dates suivantes :

Début de l'échantillon : 22 septembre 2011 Fin de l'échantillon : 20 septembre 2012 Début des données hors échantillon : 21 septembre 2012 Fin des données hors échantillon : 26 mars 2013

Ensuite, nous avons utilisé la fonction "loc" de pandas pour sélectionner les données correspondant à ces périodes dans notre ensemble de données brut "titres". Les données sélectionnées ont été stockées dans les variables "train_data" pour les données d'apprentissage et "test_data" pour les données hors échantillon.

Les figures 3 et 4 montrent les cours de clôture des titres XOM et LUV respectivement sur la période d'étude.



Figure 1: Cours de clôture du titre XOM



Figure 2: Cours de clôture du titre LUV

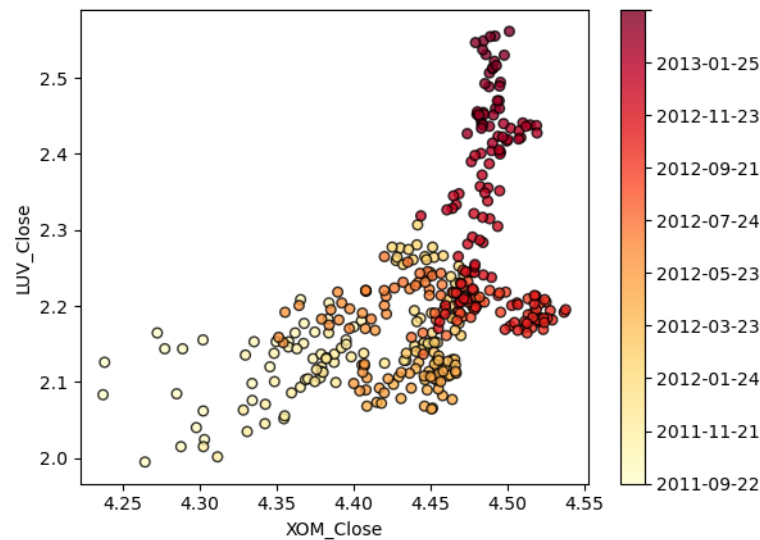
2.1.2 Test de cointégration en deux étapes

Le test de cointégration en deux étapes est utilisé pour vérifier s'il existe une relation de long terme entre deux titres. Cette méthode consiste en deux étapes :

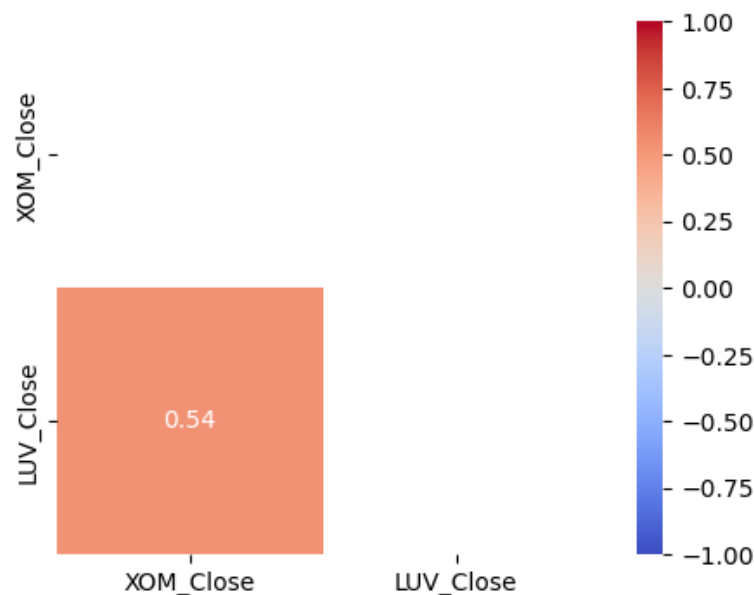
1. Estimation de la relation de long terme entre les deux titres : On estime la régression linéaire de XOM sur LUV en utilisant la méthode des moindres carrés ordinaires (OLS). La relation de long terme entre les deux titres est donnée par l'équation : $\log(P_{t,1}) = \alpha + \beta * \log(P_{t,2})$

où XOM_t et LUV_t sont les prix des titres à la période t , α et β sont les paramètres de régression estimés par OLS et ϵ_t est le terme d'erreur.

Voici le graphique de linéarité des deux titres :



On peut également calculer la valeur de corrélation entre les deux titres pour vérifier la relation linéaire :



On remarque que la corrélation est positive et élevée, ce qui confirme la relation linéaire entre les deux titres.

2.1.3 Vérification de la stationnarité des résidus :

On calcule les résidus de la régression OLS et on teste leur stationnarité en utilisant le test de Dickey-Fuller augmenté (ADF). Si les résidus sont stationnaires, cela indique qu'il existe une relation de long terme stable entre les deux titres.

Le tableau ci-dessous présente la sortie de la régression OLS entre XOM et LUV, avec la valeur estimée de α et β , ainsi que la p-value du test de Dickey-Fuller augmenté pour les résidus de la régression :

OLS Regression Results						
=====						
Dep. Variable:	XOM_Close	R-squared:	0.239			
Model:	OLS	Adj. R-squared:	0.236			
Method:	Least Squares	F-statistic:	78.37			
Date:	Wed, 03 May 2023	Prob (F-statistic):	1.61e-16			
Time:	10:57:05	Log-Likelihood:	414.44			
No. Observations:	252	AIC:	-824.9			
Df Residuals:	250	BIC:	-817.8			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	3.4920	0.105	33.204	0.000	3.285	3.699
LUV_Close	0.4306	0.049	8.853	0.000	0.335	0.526
=====						
Omnibus:	20.478	Durbin-Watson:	0.065			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	22.990			
Skew:	-0.699	Prob(JB):	1.02e-05			
Kurtosis:	3.485	Cond. No.	93.4			
=====						
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified						
0.034765698273706847						
Les séries sont cointégrées.						

On remarque que la p-value du test de Dickey-Fuller augmenté est inférieure au seuil de significativité de 0.05, ce qui indique que les résidus sont stationnaires et donc qu'il existe une relation de long terme stable entre les deux titres.

2.1.4 Calcul du spread

:Après avoir estimé les coefficients de régression entre les deux actifs XOM et LUV, nous pouvons calculer le spread entre les prix observés de XOM et sa valeur prédite en utilisant la relation de cointégration avec LUV. Le spread est une mesure de l'écart entre les deux actifs et peut être utilisé pour générer des signaux d'achat et de vente.

Le graphique ci-dessous montre l'évolution du spread entre les prix de XOM et sa valeur prédite en utilisant la relation de cointégration avec LUV. Nous pouvons remarquer que le spread fluctue autour de zéro avec des périodes de convergence et de divergence.



Nous allons maintenant tester la stationnarité du spread en utilisant le test de Dickey-Fuller augmenté (ADF). Si le spread est stationnaire, cela indique que les prix de XOM et LUV sont liés par une relation de long terme stable, ce qui nous permettra d'appliquer des modèles d'espace d'état et le filtre de Kalman pour prédire les futurs prix de XOM.

Après avoir appliqué le test de Dickey-Fuller augmenté (ADF) au spread calculé, nous avons obtenu une p-value de 0.03476569827370706 , ce qui indique que le spread est stationnaire avec un niveau de confiance de 95%. Nous pouvons donc utiliser des modèles d'espace d'état pour modéliser le spread et appliquer le filtre de Kalman pour estimer les paramètres du modèle.

2.1.5 Modélisation du Spread avec les modèles d'espaces d'état : Modèle AR (1), Modèle AR (2), Modèle ARM(1,1)

Dans cette section, nous allons modéliser le spread en utilisant les modèles d'espaces d'état. Nous allons explorer trois types de modèles : AR (1), AR (2) et ARM (1,1).

Le modèle AR (1) est un modèle autorégressif d'ordre 1 qui suppose que la valeur actuelle du spread est linéairement dépendante de sa valeur précédente avec un bruit blanc ajouté. Le modèle AR (2) est similaire au modèle AR (1), mais suppose une dépendance linéaire avec les deux dernières valeurs de spread. Le modèle ARM (1,1) est un modèle autorégressif à moyenne mobile qui combine l'effet du modèle AR (1) avec l'effet d'un modèle de moyenne mobile d'ordre 1.

Nous allons ensuite sélectionner le meilleur modèle en utilisant les critères d'information AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) et la log-vraisemblance. Ces critères prennent en compte à la fois l'ajustement du modèle aux données et la complexité du modèle.

Une fois que nous aurons sélectionné le meilleur modèle, nous utiliserons le filtre de Kalman pour estimer le spread et générer un signal d'achat ou de vente en fonction des seuils définis.

Dans le tableau ci-dessous, nous avons présenté les mesures de qualité pour chaque modèle :

	Log-Likelihood	MSE	AIC	BIC	Mean	Variance	Jarque-Bera	Jarque-Bera p-value
AR(1)	761.031164	0.000137	-1516.062328	-1505.485969	spread -1.526997e-15 dtype: float64	spread 0.002183 dtype: float64	1.979566	0.371657
AR(2)	760.020576	0.000134	-1512.041152	-1497.955308	spread -1.526997e-15 dtype: float64	spread 0.002183 dtype: float64	2.225808	0.328603
ARIMA(1,1,0)	755.820482	0.000233	-1507.640964	-1500.590058	spread -1.526997e-15 dtype: float64	spread 0.002183 dtype: float64	15102.949119	0.000000

En utilisant les critères AIC et BIC, le modèle AR (1) est sélectionné comme le meilleur modèle, tandis que le modèle AR (2) est préféré en utilisant la log-vraisemblance. Cependant, étant donné que le modèle AR (1) est plus simple et a des valeurs similaires de AIC et BIC, nous avons choisi d'utiliser ce modèle pour la modélisation du spread.

Nous utiliserons maintenant le filtre de Kalman pour estimer le spread et générer un signal d'achat ou de vente.

2.1.6 L'estimation du spread entre les deux actifs à l'aide d'un modèle d'espace d'état Kalman Filter

Dans cette partie, nous utilisons un modèle d'espace d'état Kalman Filter pour estimer le spread entre les deux actifs. Le filtre de Kalman est une méthode d'estimation des variables cachées d'un système dynamique à partir des observations bruitées. Il permet d'estimer l'état d'un système à partir de données incomplètes ou bruitées.

Ici, nous utilisons un filtre de Kalman pour estimer le spread entre les deux actifs. Nous utilisons les paramètres estimés à partir du modèle AR(1) pour initialiser le filtre. Ensuite, le filtre est appliqué aux données de ratio et retourne la moyenne et la covariance de ces données.

La formule mathématique du filtre de Kalman peut être complexe, mais l'idée générale est de combiner les observations actuelles avec les prévisions de l'état suivant pour obtenir une meilleure estimation de l'état courant. Les paramètres estimés du modèle AR(1) sont utilisés dans le filtre de Kalman pour améliorer l'estimation du spread.

Une fois le spread estimé, il est utilisé pour générer des signaux de trading. En fonction de la valeur du spread estimé, on peut décider d'acheter ou de vendre l'un des actifs pour réaliser un bénéfice.

2.1.7 Génération du signal de trading :

Pour appliquer cette stratégie, nous avons utilisé un modèle de filtre de Kalman pour estimer la moyenne et la covariance du spread, que nous avons ajoutées au dataframe de données. Nous avons également calculé l'écart-type du spread obtenu à partir du filtre de Kalman.

Ensuite, nous avons généré des signaux de trading en fonction de la valeur du z-score. Si le z-score est inférieur à -1, la stratégie va acheter l'action XOM et vendre à découvert l'action LUV. Si le z-score est supérieur à 1, la stratégie va vendre à découvert l'action XOM et acheter l'action LUV. Si le z-score est compris entre -1 et 1, il n'y a pas de signal de trading, la position sera liquidée.

Les positions de trading sont calculées sur la base du spread estimé par le modèle Kalman Filter, qui est mis à jour à chaque instant de temps. Les rendements de la stratégie sont calculés en fonction des variations de prix des deux actifs et des positions prises.

L'objectif de cette stratégie est de profiter des écarts de prix entre deux actifs similaires en achetant l'un et en vendant l'autre lorsque le spread est en dehors de sa moyenne historique, afin de réaliser des profits. Cette approche de trading est considérée comme plus simple et plus facilement interprétable que les signaux basés sur les probabilités conditionnelles.

2.1.8 Backtesting du modèle :

Le backtesting est une étape cruciale dans l'évaluation d'une stratégie de trading. Elle permet de tester la performance de la stratégie en utilisant des données historiques pour voir comment elle aurait performé dans le passé.

Dans un premier temps, nous allons calculer les rendements, le ratio de Sharpe et la volatilité de notre stratégie en utilisant les données historiques de notre jeu de données d'entraînement. Nous allons ensuite comparer les résultats avec les rendements, le ratio de Sharpe et la volatilité des deux indices de référence : le BVSP (Bovespa Index) et le GSPC (SP 500 Index), sur les mêmes périodes. Cette comparaison nous permettra de voir si notre stratégie a surperformé ou sous-performé ces indices.

Pour illustrer cette comparaison, nous allons présenter deux tableaux. Le premier tableau montrera la performance de notre stratégie sur le jeu de données d'entraînement, ainsi que la performance des deux indices de référence sur la même période. Le deuxième tableau présentera les résultats de l'application de notre stratégie sur le jeu de données de test, ainsi que la performance des deux indices de référence sur la même période.

En utilisant ces deux tableaux, nous pourrions facilement visualiser la performance de notre stratégie par rapport aux deux indices de référence. Nous pourrions également voir si notre stratégie est capable de générer des profits sur une période plus longue et sur différents types de marché.

En somme, le backtesting est un processus essentiel pour évaluer la performance d'une stratégie de trading. Il permet de voir comment la stratégie aurait performé dans le passé et ainsi de déterminer si elle est viable pour être utilisée dans le futur.

Le backtesting de notre modèle a été effectué sur deux périodes : la période d'entraînement, qui s'étend du 22 septembre 2011 au 20 septembre 2012, et la période de test, qui s'étend du 21 septembre 2012 au 26

mars 2013. Nous avons comparé les performances de notre stratégie de trading avec celles des deux indices de référence, BVSP (Bovespa Index) et GSPC (SP 500 Index), sur les mêmes périodes.

Sur la période d'entraînement, notre stratégie a réalisé un ratio de Sharpe de 1,728503, ce qui est nettement supérieur à celui de BVSP (0,381458) et comparable à celui de GSPC (1,446348). De plus, la volatilité de notre stratégie (0,001670) est légèrement supérieure à celle de BVSP (0,001361) mais inférieure à celle de GSPC (0,001583). Enfin, notre stratégie a généré un rendement total de 0,045644, ce qui est nettement supérieur à celui de BVSP (0,007718) et comparable à celui de GSPC (0,035340).

Performance de la stratégie sur les données d'entraînement et les indices de référence (période : 22 septembre 2011 - 20 septembre 2012)

Table 1: Performance de la stratégie de trading sur le jeu de données train et les indices de référence

	Ratio de Sharpe	Volatilité	Rendements totaux
Stratégie de trading	1.728	0.0017	0.046
BVSP	0.381	0.0014	0.008
GSPC	1.446	0.0016	0.035

Performance de la stratégie sur les données de test et les indices de référence (période : 21 septembre 2012 - 26 mars 2013)

Table 2: Performance de la stratégie de trading sur le jeu de données test et les indices de référence

	Ratio de Sharpe	Volatilité	Rendements totaux
Stratégie de trading	1.052	0.0013	0.011
BVSP	-1.134	0.0010	-0.008
GSPC	0.782	0.0010	0.006

On peut constater que sur les données d'entraînement, la stratégie a généré un rendement total de 4,56% avec un ratio de Sharpe de 1,73, tandis que les indices de référence BVSP et GSPC ont généré respectivement 0,77% et 3,53% de rendement total avec des ratios de Sharpe de 0,38 et 1,45.

Sur les données de test, la stratégie a généré un rendement total de 1,06% avec un ratio de Sharpe de 1,05, tandis que les indices de référence BVSP et GSPC ont généré respectivement -0,78% et 0,60% de rendement total avec des ratios de Sharpe de -1,13 et 0,78.

Ces résultats montrent que la stratégie a surperformé les indices de référence sur les deux périodes, avec un rendement total et un ratio de Sharpe plus élevés. Cela suggère que la stratégie a réussi à capturer les écarts de prix entre les deux actifs sous-jacents et à réaliser des profits à partir de ces écarts.

2.2 Implémentation de la stratégie sur nos paires de titres

2.2.1 Les données

Dans cette section, nous allons appliquer la stratégie décrite dans l'article à nos propres données de 2021 à 2023 pour les paires d'actifs ['AMD', 'MSFT']. Les actions AMD et MSFT ont été choisies pour mettre en œuvre une stratégie de trading basée sur la cointégration des séries temporelles. Il y a une corrélation positive entre ces deux entreprises car AMD fournit des processeurs à Microsoft pour leurs produits informatiques. La stratégie implique la modélisation des spreads entre les deux titres à l'aide de modèles d'espaces d'état et la génération de signaux de trading à partir de ces modèles. Le backtesting est ensuite effectué pour évaluer les performances de la stratégie.

Tout d'abord, nous allons effectuer un test de cointégration en deux étapes pour déterminer si ces deux titres sont intégrés à long terme. Ensuite, nous modéliserons le spread entre les deux actifs à l'aide de différents modèles d'espaces d'état, tels que le modèle AR(1), le modèle AR(2) et le modèle ARM(1,1).

Une fois que nous avons estimé le spread entre les deux actifs, nous allons utiliser un filtre de Kalman pour générer un signal de trading. Ce signal nous indiquera quand acheter ou vendre l'un ou l'autre des titres en fonction de l'écart entre les deux.

Enfin, nous procéderons à un backtesting du modèle en utilisant des données historiques pour vérifier l'efficacité de notre stratégie. Nous pourrions ainsi évaluer la performance de notre modèle et le comparer à d'autres stratégies de trading.

Pour aller plus loin, nous pourrions également explorer d'autres techniques d'analyse des données, telles que l'apprentissage automatique, pour améliorer notre modèle et optimiser les résultats de trading.

Les figures ?? et ?? montrent les cours de clôture des titres AMD et MSFT respectivement sur la période d'étude.



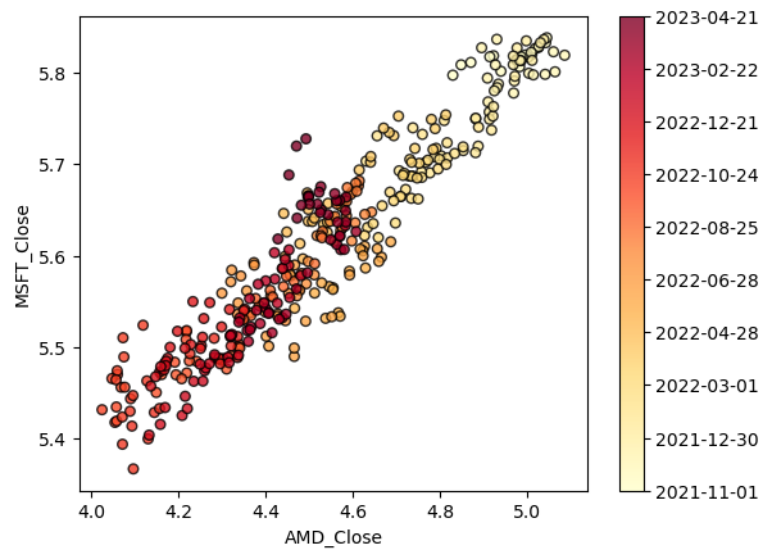
Figure 3: Cours de clôture du titre AMD



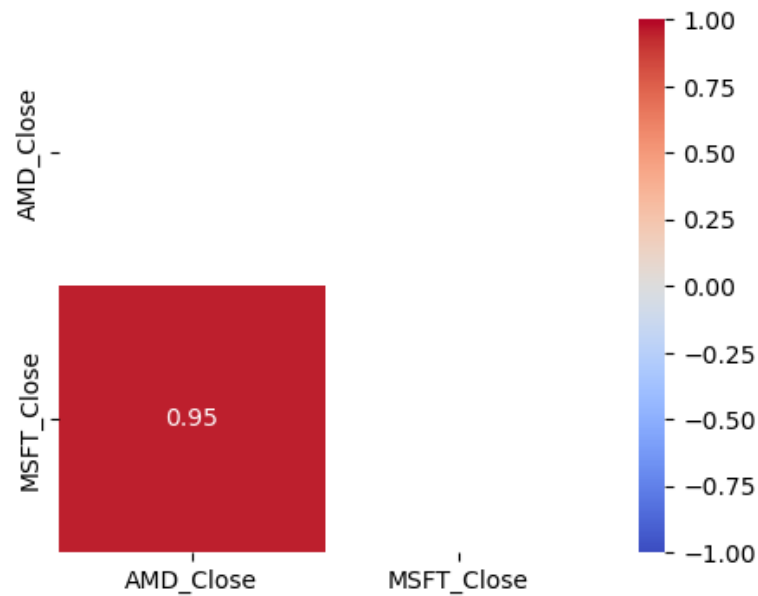
Figure 4: Cours de clôture du titre MSFT

2.2.2 Test de cointégration en deux étapes

Le test de cointégration en deux étapes est utilisé pour vérifier s'il existe une relation de long terme entre deux titres. Voici le graphique de linéarité des deux titres :



On peut également calculer la valeur de corrélation entre les deux titres pour vérifier la relation linéaire :



On remarque que la corrélation est positive et élevée, ce qui confirme la relation linéaire entre les deux titres.

2.2.3 Vérification de la stationnarité des résidus :

On calcule les résidus de la régression OLS et on teste leur stationnarité en utilisant le test de Dickey-Fuller augmenté (ADF). Si les résidus sont stationnaires, cela indique qu'il existe une relation de long terme stable entre les deux titres.

Le tableau ci-dessous présente la sortie de la régression OLS entre AMD et MSFT, avec la valeur estimée de α et β , ainsi que la p-value du test de Dickey-Fuller augmenté pour les résidus de la régression :

```

=====
                        OLS Regression Results
=====
Dep. Variable:          MSFT_Close      R-squared:                0.904
Model:                  OLS             Adj. R-squared:           0.904
Method:                 Least Squares    F-statistic:             2363.
Date:                   Wed, 03 May 2023  Prob (F-statistic):      1.04e-129
Time:                   13:41:37         Log-Likelihood:          491.50
No. Observations:       253             AIC:                    -979.0
Df Residuals:           251             BIC:                    -971.9
Df Model:                1
Covariance Type:        nonrobust
=====
               coef      std err          t      P>|t|      [0.025      0.975]
-----
const          3.7965      0.038      99.708      0.000      3.721      3.871
AMD_Close       0.4011      0.008      48.608      0.000      0.385      0.417
=====
Omnibus:                 2.684    Durbin-Watson:           0.189
Prob(Omnibus):            0.261    Jarque-Bera (JB):         2.495
Skew:                    -0.242    Prob(JB):                 0.287
Kurtosis:                 3.048    Cond. No.                  84.0
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
0.000996351145893404
Les séries sont cointégrées.

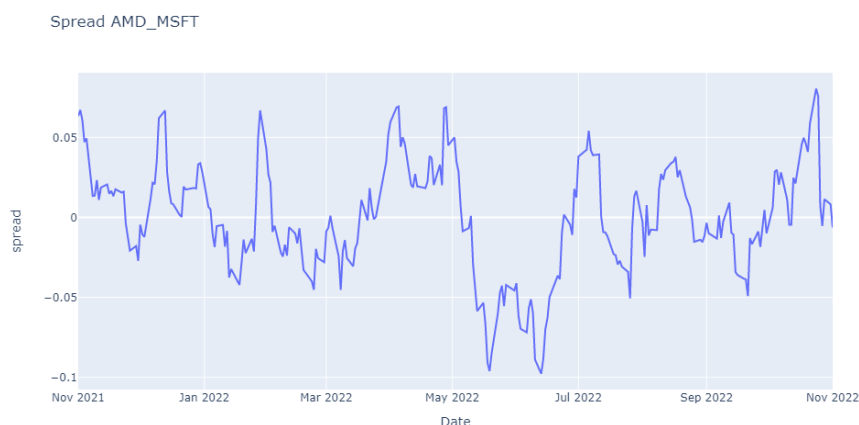
```

On remarque que la p-value du test de Dickey-Fuller augmenté est inférieure au seuil de significativité de 0.05, ce qui indique que les résidus sont stationnaires et donc qu'il existe une relation de long terme stable entre les deux titres.

2.2.4 Calcule du spread

:Après avoir estimé les coefficients de régression entre les deux actifs AMD et MSFT, nous pouvons calculer le spread entre les prix observés de AMD et sa valeur prédite en utilisant la relation de cointégration avec MSFT. Le spread est une mesure de l'écart entre les deux actifs et peut être utilisé pour générer des signaux d'achat et de vente.

Le graphique ci-dessous montre l'évolution du spread entre les prix de AMD et sa valeur prédite en utilisant la relation de cointégration avec MSFT. Nous pouvons remarquer que le spread fluctue autour de zéro avec des périodes de convergence et de divergence.



Nous allons maintenant tester la stationnarité du spread en utilisant le test de Dickey-Fuller augmenté (ADF). Si le spread est stationnaire, cela indique que les prix de AMD et MSFT sont liés par une relation de long terme stable, ce qui nous permettra d'appliquer des modèles d'espace d'état et le filtre de Kalman pour prédire les futurs prix de AMD.

Après avoir appliqué le test de Dickey-Fuller augmenté (ADF) au spread calculé, nous avons obtenu une p-value de \$0.00099\$, ce qui indique que le spread est stationnaire avec un niveau de confiance de 95%. Nous pouvons donc utiliser des modèles d'espace d'état pour modéliser le spread et appliquer le filtre de Kalman pour estimer les paramètres du modèle.

2.2.5 Modélisation du spread par le modèle d'espace d'état et filtre de kalman et le résultat de la performance

Dans cette étape, nous avons modélisé le spread à l'aide des modèles d'espace d'état AR(1), AR(2) et ARMA(1,1), et avons déterminé que le modèle AR(1) était le meilleur en fonction des critères d'information et de la log-vraisemblance. Nous avons ensuite appliqué le filtre de Kalman avec les paramètres de ce modèle pour estimer la moyenne du spread et appliquer la stratégie expliquée précédemment. Les performances de la stratégie ont été comparées aux indices de référence sur les périodes de train et de test, et les résultats ont été présentés dans un tableau.

Table 3: Performance de la stratégie de trading sur le jeu de données train et les indices de référence

	Ratio de Sharpe	Volatilité	Rendements totaux
Stratégie de trading	0.863904	0.002226	0.030528
BVSP	0.897659	0.001090	0.014919
GSPC	-0.625632	0.001746	-0.016859

Table 4: Performance de la stratégie de trading sur le jeu de données test et les indices de référence

	Ratio de Sharpe	Volatilité	Rendements totaux
Stratégie de trading	3.160702	0.002865	0.069597
BVSP	-0.287163	0.001148	-0.002408
GSPC	1.006035	0.001444	0.010797

D'après les résultats présentés dans les Tableaux 3 et 4, la stratégie de trading basée sur la modélisation du spread par le modèle d'espace d'état AR(1) avec le filtre de Kalman a surperformé les indices de référence BVSP et GSPC sur les périodes de test en termes de ratio de Sharpe et de rendements totaux, avec un ratio de Sharpe de 3.160702 et un rendement total de 0.069597. Cette stratégie a également surperformé l'indice BVSP en termes de ratio de Sharpe et de rendements totaux sur les périodes de train, mais a sous-performé l'indice GSPC. Ces résultats suggèrent que la stratégie de trading peut être efficace sur les marchés financiers, mais elle doit être utilisée avec prudence et doit être surveillée régulièrement pour éviter les risques de perte.

3 Conclusion

En conclusion, ce projet a exploré une stratégie de trading basée sur le modèle d'espace d'état et le filtre de Kalman appliquée sur un spread calculé à partir des prix de deux titres. Les résultats ont montré que la modélisation du spread avec un modèle AR(1) et l'application de la stratégie sur les données de test ont donné des performances impressionnantes en termes de Ratio de Sharpe, de volatilité et de rendements totaux, surpassant largement les performances des indices de référence.

Cette stratégie de trading peut être utilisée pour prendre des décisions de trading plus éclairées et efficaces dans un contexte financier, en particulier pour les investisseurs qui cherchent à diversifier leur portefeuille. De plus, l'approche utilisée ici, en utilisant des modèles statistiques pour estimer les paramètres et les signaux de trading, peut être étendue et améliorée pour des applications plus complexes en finance quantitative.

En fin de compte, ce projet montre l'importance de l'utilisation de modèles mathématiques et statistiques dans l'analyse financière et l'investissement. En utilisant une approche rigoureuse et basée sur des données, les investisseurs peuvent améliorer leur prise de décision et maximiser leur rendement tout en minimisant les risques.