

Optimisation de portefeuille avec réseaux de neurones et copules

Réplication et analyse de l'article

"Neural network copula portfolio optimization for exchange traded funds"

Rhellab Lin Montariol Montagard

Université Paris-Dauphine

19 mai 2025

Gestion Quantitative

Table des matières

Introduction générale	3
1 Présentation et analyse de l'article	4
1.1 Contexte et motivation	4
1.2 Méthodologie proposée	4
1.3 Contributions principales	4
1.4 Résultats clés	5
2 Réplication de l'étude	5
2.1 Données utilisées	5
2.2 Statistiques Générales	6
2.2.1 Modèle ARMA-bench	6
2.2.2 Multilayer perceptron	10
2.2.3 Recurrent Neural Network	10
2.2.4 Psi Sigma Network	11
2.2.5 Out-of-sample statistical performance	12
2.3 Modélisation des copules	13
2.3.1 Définition	13
2.3.2 t-copula classique	13
2.3.3 La Skewed t-Copula	14
2.3.4 Résultats obtenus	14
2.4 Optimisation de portefeuille	15
2.4.1 Approche moyenne-variance traditionnelle	15
2.4.2 Approche moyenne-CVaR avec copules	15
2.5 Résultat de la Réplication	16
3 Extension	17
3.1 Critique 1 : Benchmark	17
3.2 Critique 2 : Période d'étude	18

Introduction générale

Ce document présente une réplique et une analyse de l'article "Neural network copula portfolio optimization for exchange traded funds" par Yang Zhao, Charalampos Stasinakis, Georgios Sermpinis et Yukun Shi.

1 Présentation et analyse de l'article

1.1 Contexte et motivation

L'article de Zhao et al. (2018) s'inscrit dans le cadre de la théorie moderne du portefeuille tout en cherchant à dépasser certaines limitations du modèle de Markowitz (1952). La théorie classique repose sur l'hypothèse que les rendements financiers suivent une distribution normale conjointe et que leur dépendance peut être entièrement décrite par le coefficient de corrélation linéaire. Cependant, de nombreuses observations empiriques remettent en question ces hypothèses.

Les rendements financiers présentent généralement une asymétrie négative et un excès de kurtosis. Plus important encore, les rendements des actifs sont davantage corrélés pendant les baisses de marché que pendant les hausses, phénomène connu sous le nom de dépendance asymétrique. De plus, la variance comme mesure de risque traite de manière identique les écarts positifs et négatifs, ce qui ne correspond pas à la perception du risque par les investisseurs.

Pour leur étude, les auteurs choisissent d'analyser trois ETFs représentatifs du marché américain : SPDR S&P 500 (SPY), SPDR Dow Jones (DIA) et PowerShares QQQ (QQQ). Ce choix est motivé par la popularité croissante des ETFs, qui offrent aux investisseurs la possibilité de négocier des indices boursiers à des coûts de transaction très faibles tout en bénéficiant d'un niveau élevé de diversification.

1.2 Méthodologie proposée

L'approche méthodologique de l'article repose sur l'intégration de plusieurs techniques avancées. Premièrement, les auteurs appliquent trois structures de réseaux de neurones pour prévoir les rendements quotidiens des ETFs : le Multi-Layer Perceptron (MLP), un réseau standard à propagation avant ; le Recurrent Neural Network (RNN), qui incorpore une mémoire à court terme ; et le Psi Sigma Network (PSN), un réseau d'ordre supérieur combinant apprentissage rapide et capacité de cartographie puissante. Ces modèles sont comparés à un modèle ARMA traditionnel servant de référence.

Deuxièmement, pour capturer la structure de dépendance entre les rendements des ETFs, les auteurs utilisent une copule t asymétrique dans un cadre GAS (Generalized Autoregressive Score). Cette approche permet de modéliser l'asymétrie dans la dépendance entre les actifs.

Troisièmement, au lieu d'utiliser la variance comme mesure de risque, les auteurs emploient la Conditional Value-at-Risk (CVaR), qui se concentre sur le risque de queue et constitue une mesure de risque cohérente. L'optimisation du portefeuille vise ainsi à minimiser la CVaR pour un niveau donné de rendement attendu.

L'innovation principale de l'article est l'intégration de ces trois composantes dans un cadre unifié appelé processus Neural Network Copula (NNC), qui combine les prévisions précises des réseaux de neurones avec la modélisation de la dépendance asymétrique pour optimiser un portefeuille basé sur la CVaR.

1.3 Contributions principales

La contribution majeure de cet article réside dans la démonstration empirique que les réseaux de neurones, en particulier le PSN, surpassent significativement les modèles ARMA traditionnels pour la prévision des rendements des ETFs. De plus, les auteurs montrent que la copule t asymétrique modélise mieux la dépendance entre les rendements que les copules symétriques traditionnelles.

Sur le plan méthodologique, l'article propose une approche innovante qui intègre l'asymétrie de dépendance dans le processus d'optimisation de portefeuille, caractéristique généralement

négligée par les méthodes d’optimisation traditionnelles. L’utilisation de la CVaR comme mesure de risque représente également une avancée par rapport à la variance, car elle se concentre spécifiquement sur le risque de queue, plus pertinent pour les investisseurs préoccupés par les pertes extrêmes.

Enfin, le développement du cadre NNC constitue une contribution méthodologique importante, offrant une approche unifiée qui peut être appliquée à divers types de portefeuilles d’actifs financiers.

1.4 Résultats clés

L’étude révèle que le PSN surpasse systématiquement les autres modèles en termes de précision statistique et de performance de trading. En moyenne, le PSN atteint 17,35% de rendements annualisés et un ratio de Sharpe de 2,08 après les coûts de transaction, des résultats nettement supérieurs à ceux des modèles ARMA et des autres réseaux de neurones.

Les tests empiriques confirment également la supériorité de la copule t asymétrique pour modéliser la dépendance entre les rendements des ETFs. Les portefeuilles optimisés selon l’approche moyenne-CVaR présentent un ratio de Sortino environ 10% plus élevé que ceux optimisés selon l’approche moyenne-variance traditionnelle, ce qui confirme l’avantage de cibler spécifiquement le risque de queue.

Un résultat particulièrement intéressant est que les bénéfices de l’optimisation du portefeuille proviennent principalement des prévisions précises des rendements des ETFs plutôt que des estimations de la matrice de variance-covariance. Bien que la prise en compte de la dépendance asymétrique améliore la performance, l’ampleur de cette amélioration est relativement faible comparée à celle apportée par de meilleures prévisions de rendement.

Enfin, l’étude montre que l’autorisation de la vente à découvert améliore considérablement les performances des portefeuilles basés sur les réseaux de neurones, avec une augmentation d’environ 38-42% des ratios de Sharpe et de Sortino, tandis que les portefeuilles ARMA n’en bénéficient pas de manière significative. Cette différence souligne l’importance de la précision des prévisions dans un contexte de vente à découvert.

2 Réplication de l’étude

2.1 Données utilisées

Dans notre réplication, nous avons exporté les données de prix de clôture (*last price*) des trois ETFs étudiés par Zhao et al. (2018) : SPDR S&P 500 (SPY), SPDR Dow Jones (DIA) et PowerShares QQQ (QQQ). Cependant, nous avons choisi d’étendre la période d’analyse par rapport à l’article original pour vérifier la robustesse des résultats et nous assurer que les conclusions ne sont pas liées à un effet d’aubaine spécifique à la période étudiée par les auteurs.

Notre ensemble de données couvre ainsi la période du 3 janvier 2011 au 3 mai 2025, englobant la période originale de l’étude (3 janvier 2011 au 13 avril 2015) et l’étendant considérablement. Cette approche nous permet non seulement de répliquer les résultats de l’article, mais aussi d’évaluer si les méthodes proposées conservent leur efficacité sur une période plus longue et dans des conditions de marché différentes.

Comme dans l’article original, nous avons divisé nos données en trois sous-ensembles :

- Un ensemble d’entraînement pour estimer les paramètres des modèles
- Un ensemble de test pour optimiser les hyperparamètres et éviter le surapprentissage
- Un ensemble hors échantillon pour évaluer la performance des modèles

Les rendements quotidiens ont été calculés en utilisant les rendements logarithmiques, conformément à la méthodologie décrite dans l’article.

2.2 Statistiques Générales

Notre analyse des statistiques descriptives des trois ETFs révèle des caractéristiques similaires à celles rapportées par Zhao et al. (2018). Le tableau 1 présente ces statistiques et la matrice de corrélation pour notre période d'étude.

TABLE 1 – Statistiques descriptives et matrice de corrélation

Ticker	SPY US Equity	DIA US Equity	QQQ US Equity
Panel A : Statistiques descriptives			
Moyenne	0.00046	0.00040	0.00062
Écart-type	0.00961	0.00888	0.01054
Asymétrie	-0.55112	-0.50597	-0.39761
Kurtosis	8.21359	7.36236	6.17882
Jarque-Bera (p-value)	0.00000	0.00000	0.00000
ADF (p-value)	0.00000	0.00000	0.00000
Panel B : Matrice de corrélation (Pearson [Spearman])			
SPY US Equity	1.000 [1.000]		
DIA US Equity	0.972 [0.953]	1.000 [1.000]	
QQQ US Equity	0.927 [0.889]	0.872 [0.807]	1.000 [1.000]

Nos résultats confirment dans l'ensemble les observations de l'article original, avec quelques différences notables. En comparant avec les valeurs rapportées l'auteur, nous constatons que les moyennes et écarts-types des rendements sont très proches. Cependant, nous observons une asymétrie négative légèrement plus prononcée pour le SPY et le DIA dans notre échantillon, ainsi qu'un kurtosis plus élevé pour les trois ETFs.

La matrice de corrélation montre également des valeurs de corrélation plus élevées que celles rapportées dans l'article original. Néanmoins, le test de Jarque-Bera rejette l'hypothèse de normalité pour toutes les séries au niveau de confiance de 99%, tandis que le test de Dickey-Fuller augmenté (ADF) confirme la stationnarité des rendements, ce qui concorde avec les résultats de l'article.

La figure 1 présente l'évolution des trois ETFs pendant notre période d'étude.

Notre analyse graphique montre que les trois ETFs présentent des trajectoires généralement similaires avec une tendance haussière dominante sur l'ensemble de la période, mais avec des périodes de volatilité marquée, notamment pendant les corrections de marché. Le QQQ (NASDAQ-100) affiche la performance la plus forte, suivi du SPY (S&P 500) puis du DIA (Dow Jones), reflétant la surperformance des valeurs technologiques sur cette période.

2.2.1 Modèle ARMA-bench

Notre première étape dans la réplcation a consisté à implémenter les modèles ARMA qui servent de référence dans l'étude originale. Selon Zhao et al. (2018), les auteurs ont choisi des modèles ARMA(8,8), ARMA(10,10) et ARMA(7,7) respectivement pour SPY, DIA et QQQ, en se basant sur l'analyse des corrélogrammes de l'échantillon d'entraînement.

Cependant, notre propre analyse des fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF) nous a amenés à remettre en question ces spécifications. Les figures 2, 3 et 4 présentent nos résultats pour chaque ETF, avec des intervalles de confiance à 95%.

Notre première difficulté majeure dans la réplcation concerne la sélection des ordres des modèles ARMA. L'interprétation de nos ACF et PACF ne corrobore pas les choix d'ARMA(8,8), ARMA(10,10) et ARMA(7,7) faits par les auteurs. En effet, nos analyses suggèrent que les

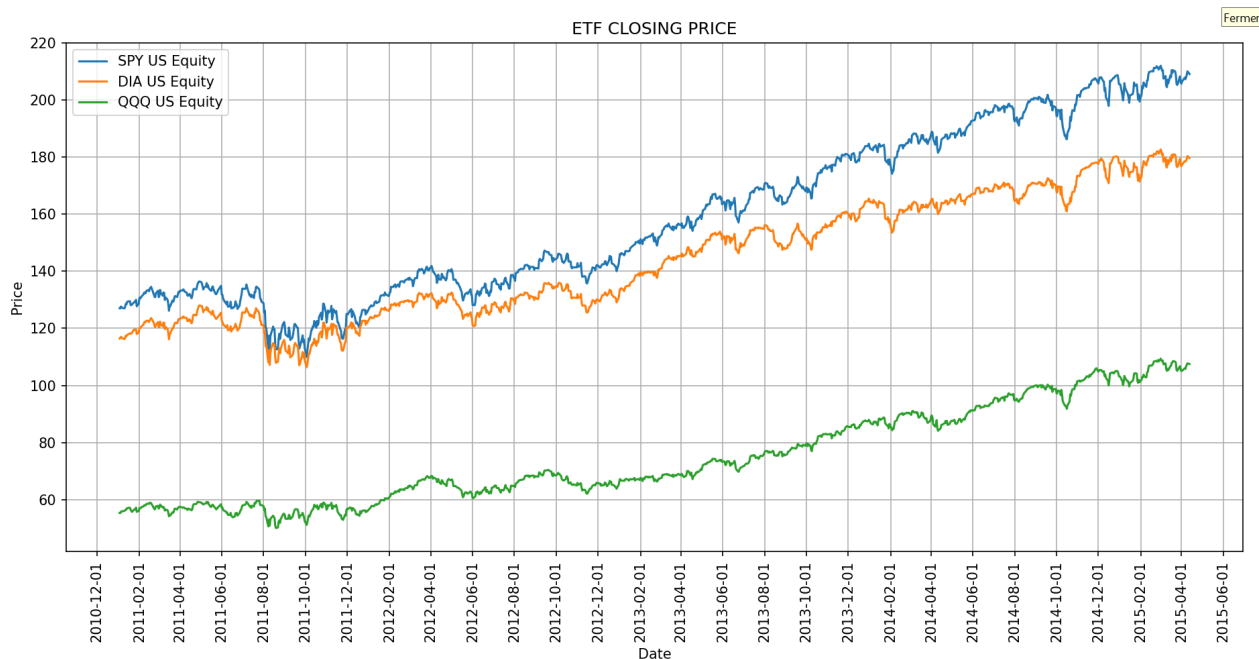


FIGURE 1 – Performance des ETFs pendant la période d'étude

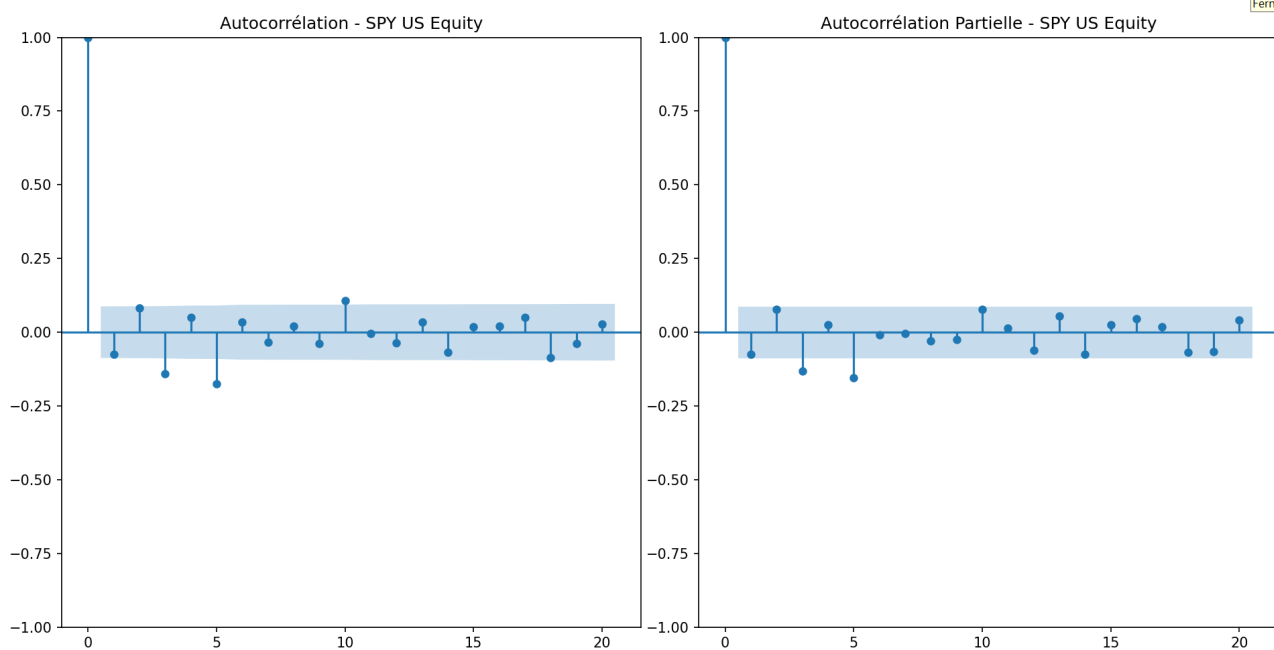


FIGURE 2 – Fonctions d'autocorrélation et d'autocorrélation partielle pour SPY

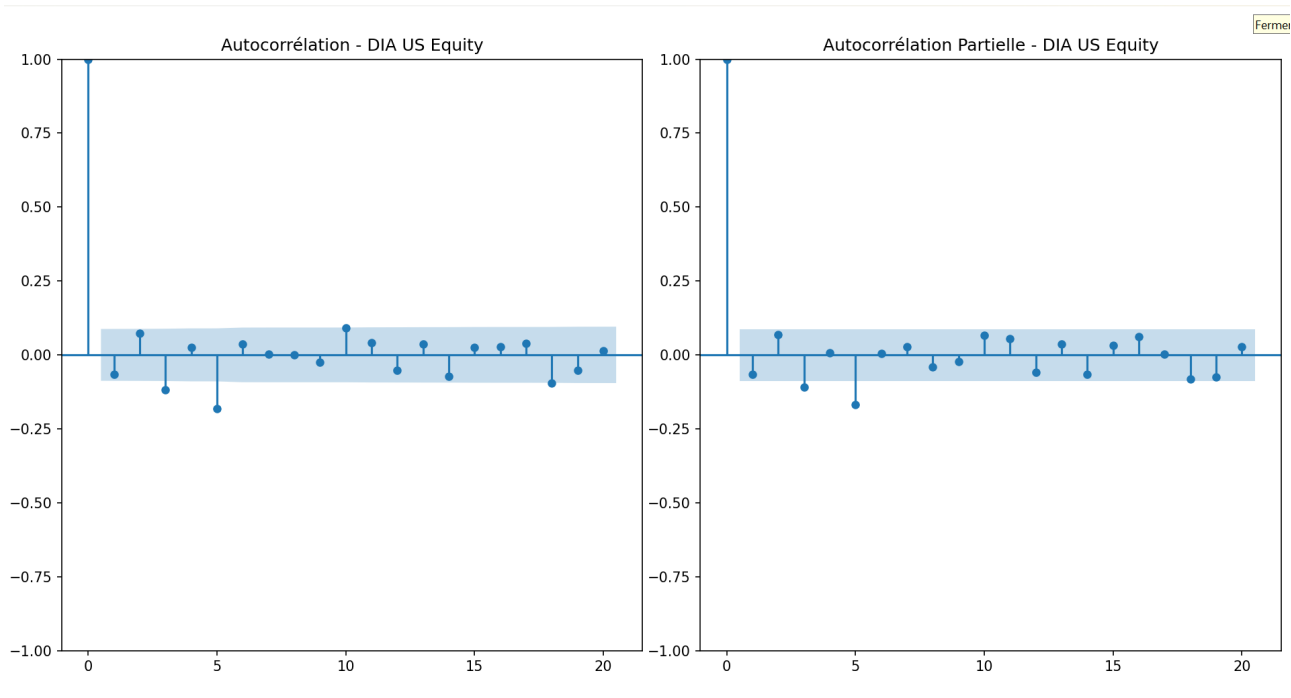


FIGURE 3 – Fonctions d'autocorrélation et d'autocorrélation partielle pour DIA

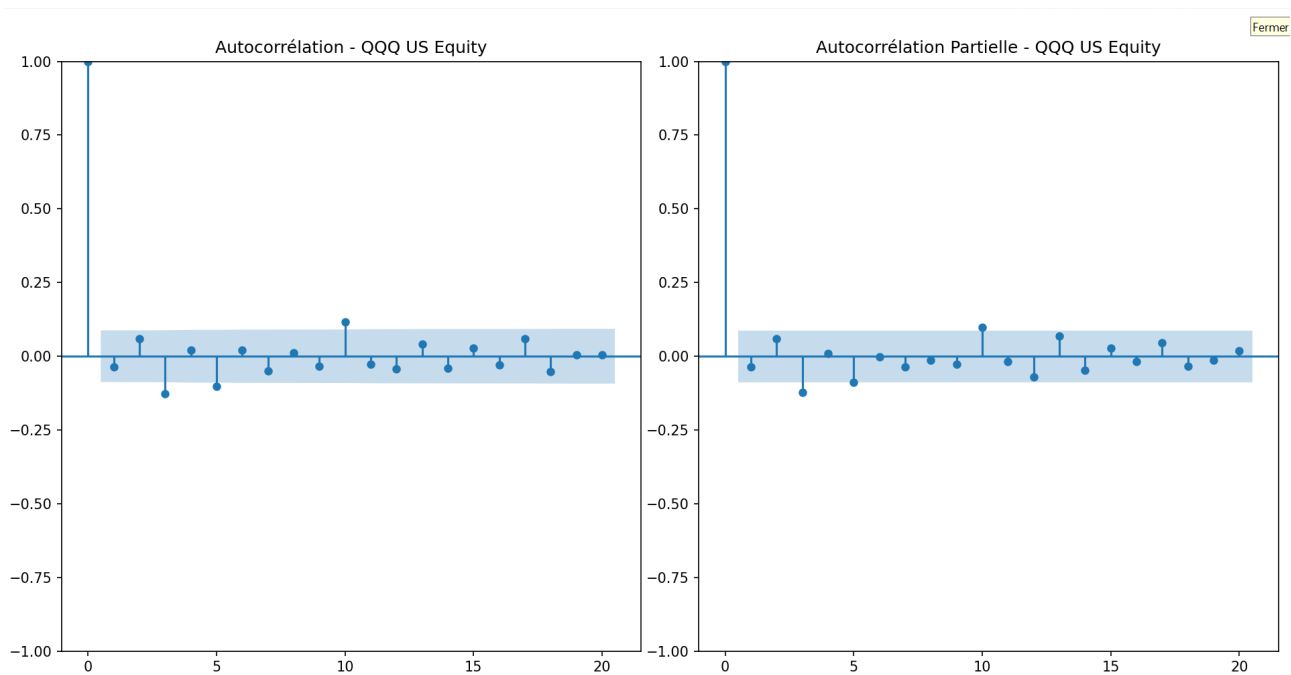


FIGURE 4 – Fonctions d'autocorrélation et d'autocorrélation partielle pour QQQ

fonctions d'autocorrélation et d'autocorrélation partielle montrent très peu de coefficients significativement différents de zéro au seuil de 0.95. Cela indiquerait plutôt des modèles de faible ordre, potentiellement ARMA(1,1) ou même un simple bruit blanc pour certaines séries.

Cette divergence soulève une question méthodologique importante, car l'article ne détaille pas précisément comment ces ordres ont été déterminés. Les auteurs mentionnent s'être basés sur les corrélogrammes de l'échantillon d'entraînement, mais nos propres analyses avec des intervalles de confiance à 0.95 indiquent des structures potentiellement beaucoup plus simples.

Face à cette difficulté, nous avons développé un code d'optimisation des paramètres ARMA basé sur différents critères : le critère d'information d'Akaike (AIC), le critère d'information bayésien (BIC) et l'erreur quadratique moyenne (RMSE). Cette approche systématique nous a permis d'obtenir une sélection objective des ordres des modèles et de comparer nos résultats avec ceux de l'article original.

Le tableau 2 présente les résultats de cette optimisation pour chaque ETF.

TABLE 2 – Résultats de l'optimisation des paramètres ARMA

Critère	SPY US Equity	DIA US Equity	QQQ US Equity
Paramètres de l'article	(8, 8)	(10, 10)	(7, 7)
Meilleurs paramètres AIC	(5, 5)	(5, 0)	(8, 9)
Meilleurs paramètres BIC	(1, 1)	(1, 0)	(1, 0)
Meilleurs paramètres RMSE	(7, 10)	(9, 9)	(8, 9)

En résumé, bien que notre procédure d'optimisation fournisse des résultats plus proches des choix de l'article que notre analyse initiale des ACF et PACF, des différences persistent. Les raisons exactes des choix de paramètres dans l'article original restent encore mystérieuses, les auteurs n'ayant pas détaillé leur méthodologie de sélection. Cette divergence souligne l'importance de la transparence méthodologique dans les études empiriques et les défis inhérents à la reproduction exacte de résultats en finance quantitative.

Pour la suite de notre réplique, nous avons décidé d'utiliser les paramètres définis dans l'article.

2.2.2 Multilayer perceptron

Dans la continuité de l'étude, nous avons implémenté un réseau de neurones Multi-Layer Perceptron (MLP) pour prévoir les rendements journaliers des trois ETFs. Le MLP constitue une architecture de réseau de neurones à propagation avant, composé d'une couche d'entrée, d'au moins une couche cachée et d'une couche de sortie. Comme dans l'article, nous avons utilisé une procédure d'early stopping en divisant notre jeu de données en trois ensembles : entraînement, test (validation) et hors échantillon. L'entraînement s'effectue via la rétropropagation de l'erreur, avec une fonction d'erreur de type MSE (Mean Squared Error).

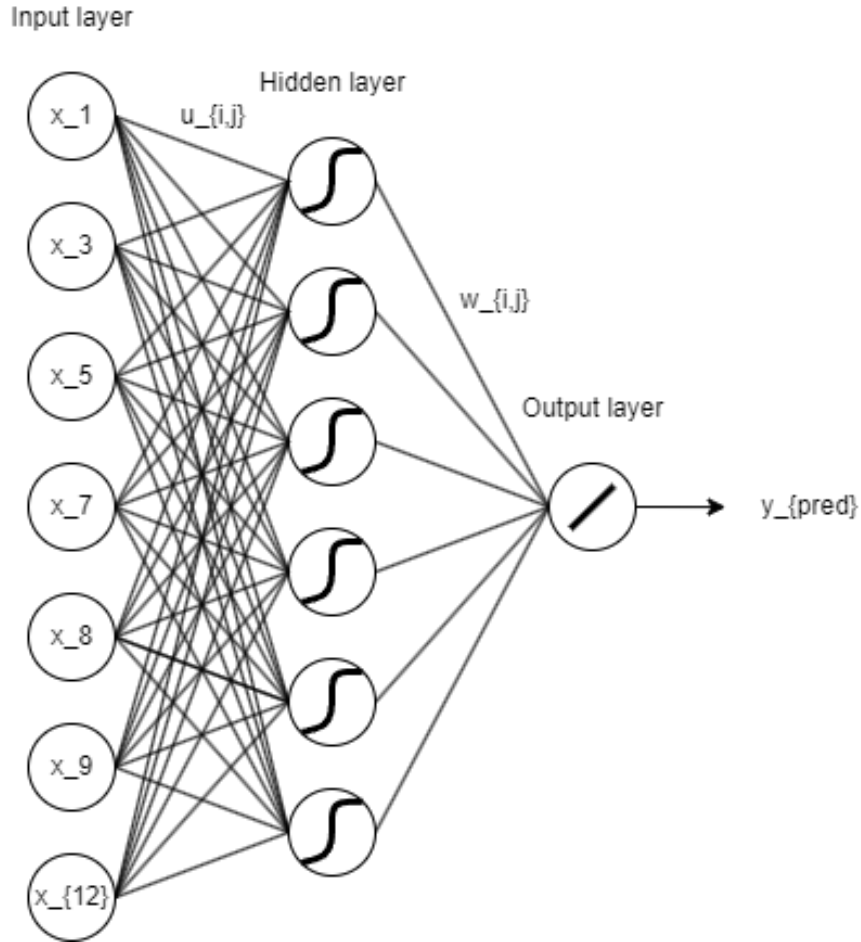


FIGURE 5 – Architecture du réseau de neurones Multi-Layer Perceptron (MLP)

La sélection du nombre de nœuds cachés a été optimisée en maximisant la performance sur l'ensemble de validation, à l'instar de la procédure décrite dans l'article. Nous avons utilisé une fonction d'activation sigmoïde dans la couche cachée et une fonction linéaire dans la couche de sortie, comme recommandé dans les applications de prévision de séries temporelles financières.

2.2.3 Recurrent Neural Network

Nous avons également implémenté un réseau de neurones récurrent (Recurrent Neural Network, RNN) pour la prévision des rendements journaliers des trois ETFs. Contrairement au MLP, l'architecture RNN intègre une mémoire à court terme via des connexions récurrentes, ce qui permet au modèle de capturer des dépendances temporelles plus complexes et d'exploiter l'information contenue dans les observations passées. Cette propriété est particulièrement pertinente dans le cadre des séries financières, qui présentent souvent des dynamiques non linéaires et des effets de mémoire.

Le RNN utilisé dans notre étude suit une architecture simple avec une couche cachée récurrente. À chaque pas de temps, l'entrée est combinée avec l'état précédent du réseau, ce qui permet de conserver une trace de l'information passée. Bien que les RNN soient plus coûteux en temps de calcul que les MLP, ils peuvent produire de meilleures performances prédictives grâce à cette capacité de mémoire de court terme.

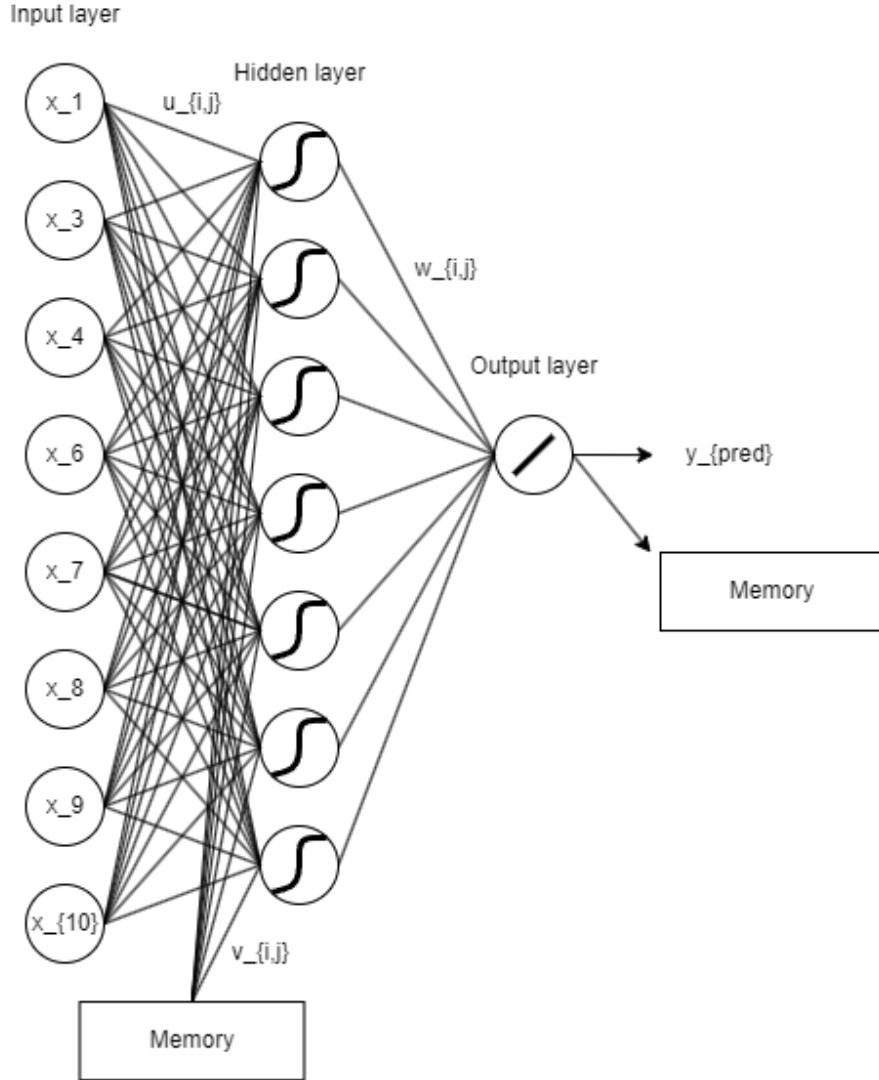


FIGURE 6 – Architecture d'un RNN

Comme pour le MLP, nous avons divisé notre ensemble de données en trois parties (entraînement, test, hors échantillon) et utilisé une procédure d'early stopping pour éviter le surapprentissage. La fonction d'activation utilisée dans la couche cachée est une tangente hyperbolique, tandis que la couche de sortie utilise une activation linéaire pour prédire les rendements.

2.2.4 Psi Sigma Network

Le dernier modèle de prévision implémenté est le Psi Sigma Network (PSN), une classe particulière de réseaux de neurones d'ordre supérieur (Higher Order Neural Networks, HONN). Ce type de réseau a été introduit pour combiner la rapidité d'apprentissage des réseaux simples à une capacité de modélisation puissante, tout en évitant l'explosion combinatoire du nombre de paramètres.

Le PSN est un réseau feed-forward entièrement connecté, où la couche cachée applique une transformation polynomiale aux entrées, capturant ainsi les interactions non linéaires complexes

entre les variables. Contrairement aux MLP ou RNN, le PSN fixe les poids entre la couche cachée et la couche de sortie à 1. Seuls les poids entre la couche d'entrée et la couche cachée sont entraînés. Cette spécificité réduit significativement le temps d'apprentissage tout en maintenant une capacité d'approximation élevée.

Comme pour les autres modèles, nous avons utilisé un découpage des données en trois sous-ensembles (entraînement, test et hors échantillon), avec une procédure d'early stopping fondée sur la performance sur l'ensemble de test. Le PSN a été entraîné via rétropropagation, avec une fonction d'erreur quadratique moyenne (MSE).

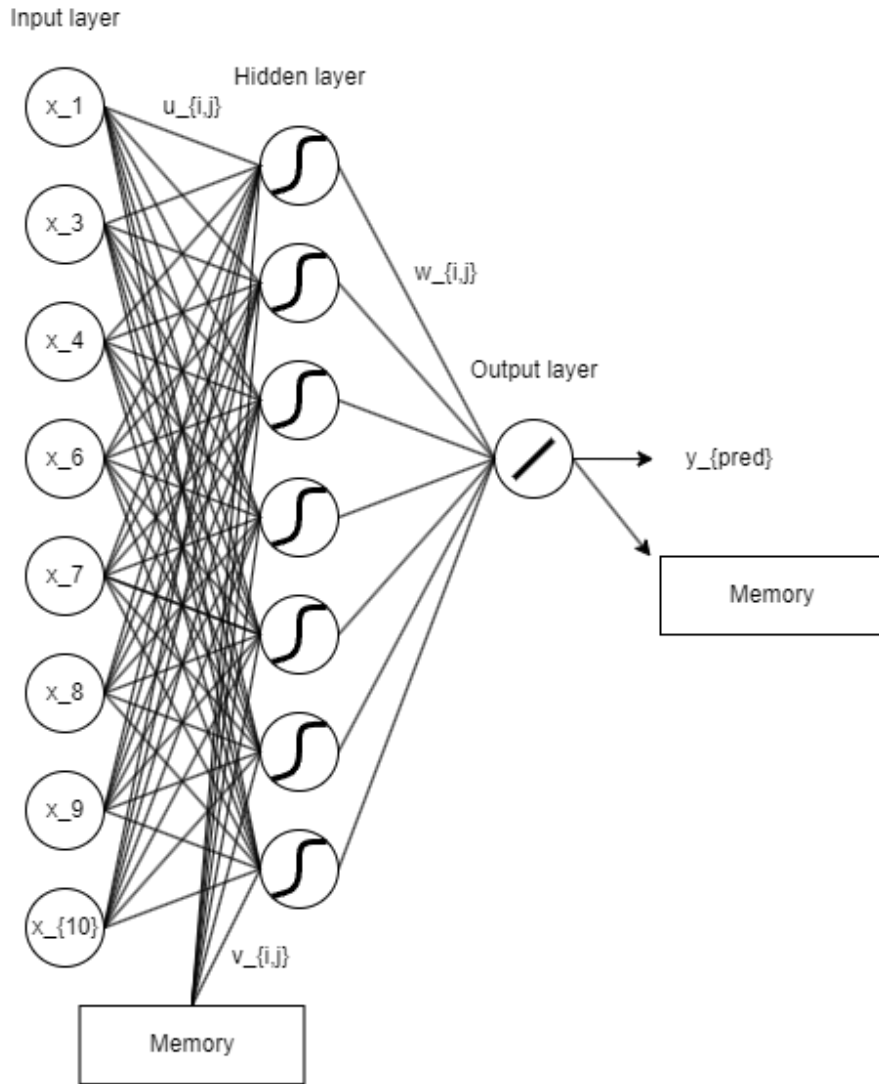


FIGURE 7 – Architecture d'un RNN

2.2.5 Out-of-sample statistical performance

Les résultats ci-dessus montrent que le classement des modèles en termes de performance statistique est cohérent pour toutes les séries d'ETF. En général, les ARMA de référence présentent les moins bons résultats statistiques par rapport à tous les modèles. Le PSN apparaît comme constamment supérieur en termes statistiques face à tous les réseaux de neurones et aux ARMA. Enfin, le RNN est le deuxième meilleur modèle.

TABLE 3 – Out-of-sample statistical performance

ETF	Statistic	ARMA	MLP	RNN	PSN
SPY	MAE	0.0054	0.0058	0.0057	0.0056
	MAPE	150.16%	148.17%	172.37%	167.50%
	RMSE	0.0071	0.0074	0.0073	0.0071
	THEIL-U	0.9286	0.7403	0.8197	0.8690
DIA	MAE	0.0049	0.0056	0.0052	0.0050
	MAPE	106.13%	153.52%	163.54%	162.70%
	RMSE	0.0064	0.0072	0.0069	0.0064
	THEIL-U	0.9283	0.7578	0.8528	0.8381
QQQ	MAE	0.0058	0.0064	0.0062	0.0059
	MAPE	100.32%	143.37%	159.05%	154.94%
	RMSE	0.0078	0.0086	0.0079	0.0077
	THEIL-U	0.9413	0.7525	0.7775	0.8143

2.3 Modélisation des copules

2.3.1 Définition

Une copule est une fonction $C : [0, 1]^d \rightarrow [0, 1]$ qui permet de relier des distributions marginales à leur distribution conjointe. Selon le théorème de Sklar, on a :

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$$

où :

- F est la fonction de répartition conjointe,
- F_i sont les fonctions de répartition marginales,
- C est la copule.

2.3.2 t-copula classique

La t-copula est construite à partir de la loi t multivariée :

$$C(u_1, \dots, u_d) = t_{\nu, R}(t_{\nu}^{-1}(u_1), \dots, t_{\nu}^{-1}(u_d))$$

où :

- $t_{\nu, R}$ est la fonction de répartition de la loi t multivariée,
- t_{ν}^{-1} est la quantile inverse de la loi t univariée.

Avantages

- Modélise la dépendance dans les queues (contrairement à la copule gaussienne)
- Symétrie dans les queues

Limite : symétrie des queues

La t-copula classique suppose une dépendance symétrique entre les extrêmes (queues). En pratique, surtout en finance, on observe une co-dépendance plus forte dans les pertes que dans les gains, ce qui motive l'utilisation d'une copule asymétrique.

2.3.3 La Skewed t-Copula

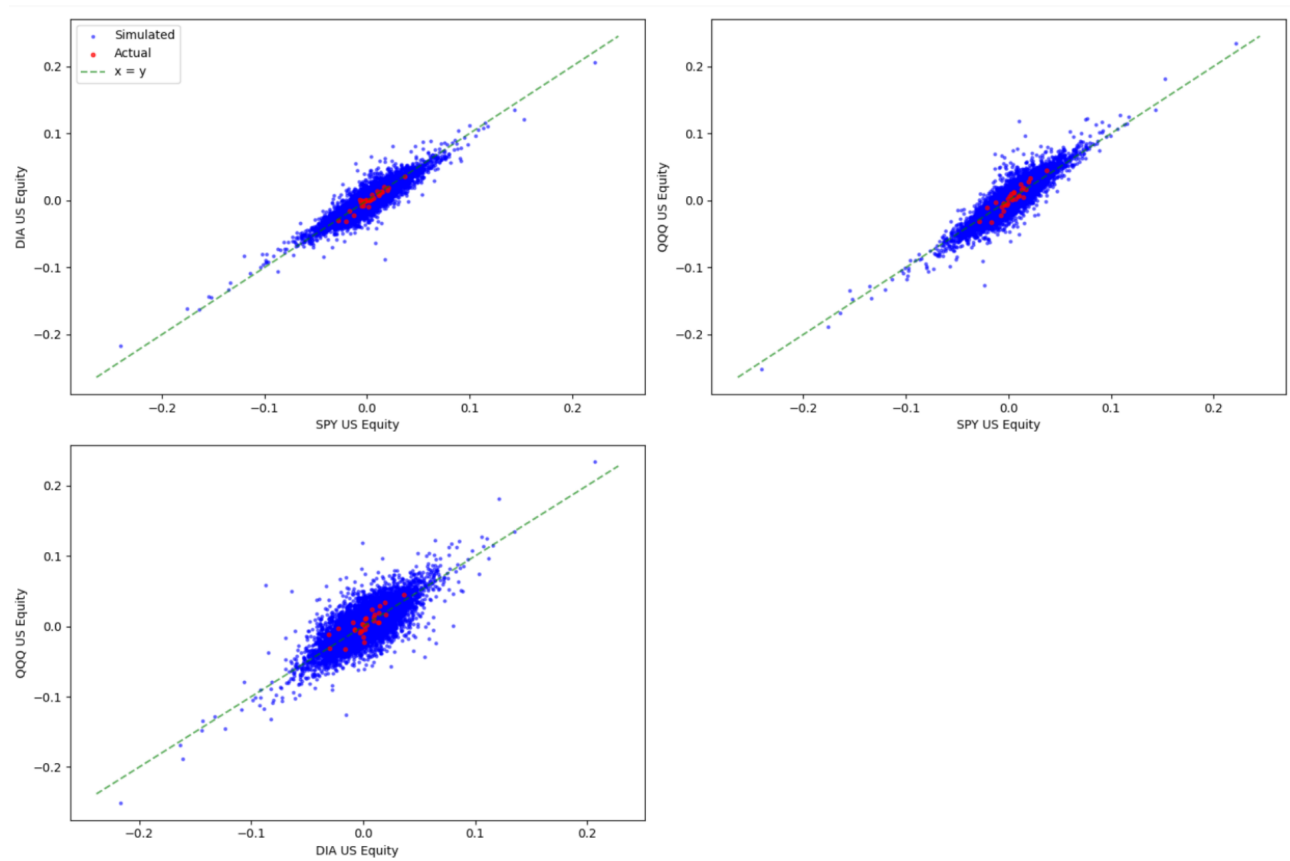
La skewed t-copula est construite à partir d'une loi t multivariée asymétrique. La loi t asymétrique de Hansen (utilisée en GARCH). Utiliser pour modéliser la dépendance asymétrique et les queues épaisses dans les séries financières. Contrairement à la copule t classique, la version "skewed" permet une asymétrie dans les distributions marginales et conditionnelles.

Avantages

- Capture à la fois les queues épaisses et l'asymétrie des rendements financiers.
- Permet une modélisation plus réaliste des co-mouvements extrêmes (contagion financière).
- Généralise la copule t classique, qui est un cas particulier lorsque $\lambda = 0$.

2.3.4 Résultats obtenus

Voici des nuages de points, représentant les rendements des trois ETF, simulés grâce à l'estimation des paramètres de dépendance via la copule de student asymétrique. Nous constatons que les rendements sont concentrés ensemble.



2.4 Optimisation de portefeuille

2.4.1 Approche moyenne-variance traditionnelle

L'approche moyenne-variance traditionnelle, développée par Markowitz en 1952, propose une méthode d'optimisation permettant à un investisseur de sélectionner un portefeuille d'actifs en fonction d'un compromis entre le rendement espéré et le risque, ce dernier étant mesuré par la variance des rendements.

On a :

— Problème d'optimisation classique :

$$\min_{\mathbf{w}} \quad \mathbf{w}^T \Sigma \mathbf{w}$$

$$\text{sous les contraintes : } \mathbf{w}^T \boldsymbol{\mu} = \mu_{\text{cible}} \quad \text{et} \quad \sum_{i=1}^n w_i = 1$$

Avec :

- $\mathbf{w} = (w_1, w_2, \dots, w_n)^T$: vecteur des poids des actifs,
- $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$: vecteur des rendements espérés,
- Σ : matrice de covariance des rendements des actifs.

Hypothèses du modèle

1. Les rendements des actifs sont normalement distribués.
2. L'investisseur est rationnel et averse au risque.
3. Le risque est uniquement mesuré par la variance.
4. Pas de coûts de transaction ni de contraintes de marché.
5. Les corrélations entre actifs sont constantes.

Limites

- La distribution normale des rendements est souvent irréaliste.
- La variance ne fait pas de distinction entre les pertes et les gains.
- Le modèle est sensible aux erreurs d'estimation de $\boldsymbol{\mu}$ et Σ .
- Peu adapté aux contextes de crise (ex. : COVID-19, crises financières).

L'approche moyenne-variance constitue un point de départ pour des modèles plus avancés comme la Value-at-Risk, etc.

2.4.2 Approche moyenne-CVaR avec copules

Selon les auteurs, la variance est largement utilisée dans la pratique financière, mais ne constitue pas une mesure de risque satisfaisante. La variance pénalise de manière égale les gains et les pertes. Or, les pertes tendent à être plus prononcées que les gains à la hausse. Ainsi, cette étude considère le CVaR comme une mesure de risque alternative : il est facilement interprétable et présente plusieurs propriétés mathématiques intéressantes. La minimisation du CVaR d'un portefeuille est liée à celle de la VaR. Les auteurs adopte alors l'approche d'optimisation de portefeuille Mean-CVaR basée sur une copule à réseau de neurones. Voici leurs résultats :

Table 8. Performance of different trading strategies (Mean-95% CVaR).

	Realized return (%)	Return/CVaR	Sortino	Max drawdown (%)
Panel A: ETF indices and 1/N portfolio				
SPY	10.233	2.9636	1.3519	7.590
DIA	6.773	1.8335	0.8995	7.041
QQQ	16.477	4.2314	2.2277	8.822
1/N	11.161	3.4367	1.4937	7.415
Panel B: Mean-CVaR optimization without short selling				
ARMA-DCC- SKT	13.114	4.0400	1.774	6.650
ARMA-ADCC- SKT	12.869	3.9714	1.735	6.650
ARMA-GAS- SKT	13.116	4.0721	1.786	6.650
RNN-DCC- SKT	28.757	9.3672	3.988	6.975
RNN-ADCC- SKT	28.884	9.4085	4.005	6.975
RNN-GAS- SKT	28.940	9.5513	4.013	7.015
PSN-DCC- SKT	30.491	9.8103	4.136	8.312
PSN-ADCC- SKT	30.608	9.8482	4.152	8.317
PSN-GAS- SKT	30.726	9.8860	4.168	8.404
Panel C: Mean-CVaR optimization with short selling				
ARMA-DCC-SKT-S	13.873	3.9422	1.692	7.958
ARMA-ADCC-SKT-S	14.190	4.0384	1.731	7.958
ARMA-GAS-SKT-S	15.022	4.2556	1.828	7.958
RNN-DCC-SKT-S	42.609	13.5624	5.544	7.835
RNN-ADCC-SKT-S	43.151	13.6425	5.507	8.931
RNN-GAS-SKT-S	42.991	13.9732	5.598	7.785
PSN-DCC-SKT-S	43.081	13.6203	5.503	8.890
PSN-ADCC-SKT-S	43.307	13.7847	5.639	7.785
PSN-GAS-SKT-S	44.294	14.0037	5.662	9.033

Notes: The table presents the out-of-sample performances over the period January 2014 to March 2015 (68 weekly observations). Panel A reports performances of the three ETF indices and the 1/N portfolio (equally weighted buy-and-hold portfolio). Panel B reports performances of different mean-CVaR portfolios without short-selling. All the portfolios are weekly rebalanced tangency portfolios obtained by the different mean-CVaR optimization based on various model combinations. For example, ARMA-DCC refers to the performance of the tangency portfolio of the efficient frontier of the three ETF assets, where the expected returns are obtained through ARMA forecasts, while the variance-covariance matrix is predicted by DCC. Panel C reports performances of different different mean-CVaR portfolios with short-selling. 'SKT' represents that the 95% CVaR is predicted using a Monte-Carlo simulation with the skewed t copulas to allow for asymmetric tail dependence '-S' denotes optimizations allowing short-selling.

Nous comprenons donc qu'en termes de prévision et de performance de trading des différents modèles, le modèle PSN surpasse les modèles MLP et RNN. Les résultats confirment que les portefeuilles issus des réseaux de neurones, et en particulier ceux fondés sur le modèle PSN, procurent des avantages significatifs à l'investisseur lorsque l'optimisation repose sur une mesure de risque asymétrique plutôt que sur une mesure symétrique. Par ailleurs, nous notons que l'ensemble des réseaux de neurones fournit des prévisions plus précises et une rentabilité supérieure à celle du modèle ARMA.

2.5 Résultat de la Réplication

Nous constatons une différence importante entre nos résultats et ceux du papier. Par exemple, pour les modèles suivant, nous avons :

- ARMA-ADCC : 3.561 contre 13.114
- RNN-GAS : -0.12 contre 28.9
- PSN-GAS : 0.829 contre 30.726

	Return (%)	Return / CVaR	Sortino	Max drawdown (%)
Panel B : Mean–Variance opt. w/o short-selling				
ARMA-DCC	3.561	0.2735	0.8122	0.360
ARMA-ADCC	3.838	0.3167	0.9356	0.220
ARMA-GAS	4.049	0.3121	0.9234	0.363
RNN-DCC	0.433	0.0196	0.0607	1.057
RNN-ADCC	0.332	0.0116	0.0376	2.458
RNN-GAS	-0.120	-0.005	-0.0156	0.931
PSN-DCC	0.714	0.0995	0.2954	0.273
PSN-ADCC	2.607	1.3888	4.3661	0.062
PSN-GAS	0.829	0.1307	0.3744	0.276
Panel C : Mean–Variance opt. with short-selling				
ARMA-DCC-S	3.773	0.2952	0.9289	0.384
ARMA-ADCC-S	-	-	-	-
ARMA-GAS-S	3.684	0.3155	0.9786	0.328
RNN-DCC-S	-0.207	-0.006	-0.0194	1.471
RNN-ADCC-S	-	-	-	-
RNN-GAS-S	1.637	0.084	0.2479	0.965
PSN-DCC-S	0.626	0.084	0.2369	0.523
PSN-ADCC-S	-	-	-	-
PSN-GAS-S	0.868	0.1244	0.3876	0.302

Les résultats que nous avons obtenus à partir de données présentant des caractéristiques statistiques similaires à celles utilisées dans l'article se sont révélés peu convaincants. Cela nous a conduits à poursuivre ce projet selon plusieurs axes de réflexion, dans le but d'adopter une approche critique vis-à-vis des conclusions proposées par les auteurs.

3 Extension

3.1 Critique 1 : Benchmark

Les résultats présentés dans cette étude montrent que le modèle ARMA utilisé comme benchmark présente des limitations significatives pour challenger efficacement les modèles de réseaux de neurones. En effet, comme le démontrent les figures 2, 3 et 4, les paramètres du modèle ARMA semblent avoir été sélectionnés de manière sous-optimale ou selon une méthodologie insuffisamment explicitée.

Afin de proposer un benchmark économétrique plus robuste, nous avons analysé les rendements logarithmiques des ETFs qui révèlent clairement des clusters de volatilité et de l'hétéroscédasticité, comme illustré dans la Figure 8.

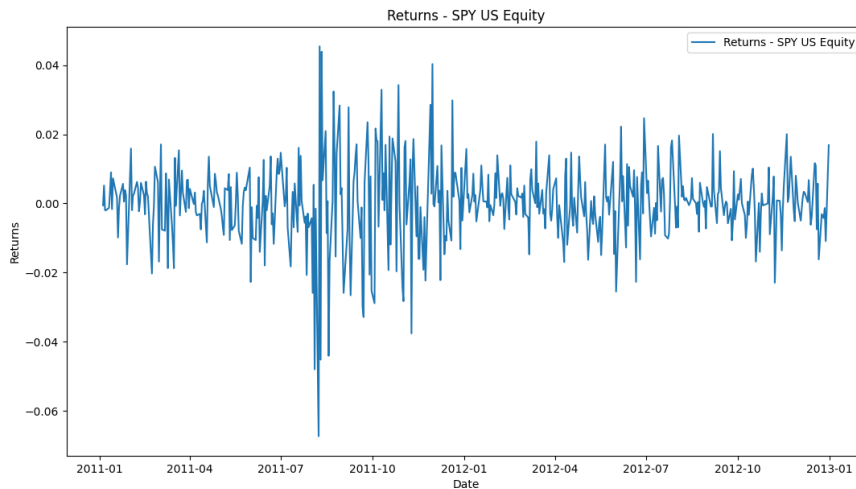


FIGURE 8 – Clusters de volatilité dans les rendements logarithmiques des ETFs

En tenant compte de l'alternance entre périodes haussières et baissières (bull and bear

markets) dans nos données, nous proposons l'utilisation d'un modèle Markov-Switching combiné avec un GARCH GJR (MS-GJRGARCH) avec une matrice de transition 2×2 représentant deux états de marché.

Les résultats de ce modèle alternatif, présentés dans le tableau ci-dessous, démontrent une amélioration significative par rapport au benchmark ARMA original et rivalisent même avec le modèle PSN sur certaines métriques :

ETF	MS-GJRGARCH (Notre modèle)			PSN (benchmark)		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
DIA	0.0054	126.16%	0.0073	0.0051	128.81%	0.0068
SPY	0.0053	113.66%	0.0074	0.0055	141.21%	0.0071
QQQ	0.0066	114.31%	0.0080	0.0062	119.22%	0.0075

Ces résultats nuancent les conclusions de l'étude originale concernant la supériorité systématique des réseaux de neurones. Notre modèle MS-GJRGARCH surpasse le PSN en termes de MAE et MAPE pour SPY, et présente une meilleure MAPE pour DIA, bien que le PSN conserve globalement un léger avantage sur l'ensemble des métriques. Cela suggère qu'un modèle économétrique bien spécifié peut rivaliser avec les approches par réseaux de neurones dans certains contextes de prévision des rendements d'ETFs.

3.2 Critique 2 : Période d'étude

Table 3. The total data-set.

Datasets	Trading days	Start date	End date
Total data-set	1075	03/01/2011	13/04/2015
Training data-set	502	03/01/2011	31/12/2012
Test data-set	252	02/01/2013	31/12/2013
Out-of-sample data-set	321	02/01/2014	13/04/2015

Note: The in-sample period is the sum of the training and test datasets.

La période d'étude choisie par les auteurs, allant de 2011 à 2015, correspond à une phase relativement stable sur le plan économique, sans événements majeurs ou crises soudaines. Cette « période lisse » limite la portée des conclusions du papier, car elle ne permet pas d'évaluer la robustesse des modèles utilisés en contexte de forte incertitude ou de choc exogène, comme cela aurait été le cas lors de la pandémie de COVID-19. Il reste donc incertain si les approches proposées seraient tout aussi efficaces en période de crise ou de volatilité extrême. Par ailleurs, notre tentative de réplique des résultats, en utilisant exactement les mêmes données et protocoles que ceux décrits dans l'article, n'a pas permis de reproduire les performances annoncées. Cela soulève des questions sur la fiabilité des approches proposées, en particulier dans des environnements plus volatiles.