

Neural Network Copula Portfolio Optimization for Exchange Traded Funds

Yang Zhao, Charalampos Stasinakis, Georgios Sermpinis, Yukun
Shi

Gestion Quantitative

Présenté par:

Rhellab · Lin · Montariol · Montagard

Titre complet :

Neural Network Copula Portfolio Optimization for Exchange Traded Funds

Auteurs :

- Yang Zhao – Jiangxi University of Finance and Economics
- Charalampos Stasinakis – University of Glasgow
- Georgios Sermpinis – University of Glasgow
- Yukun Shi – University of Leicester

Publié dans :

Quantitative Finance, Vol. 18, No. 5, 2018, pp. 761–775

- Modern portfolio theory (Markowitz) repose sur des hypothèses de normalité et de corrélation linéaire.
- Mais : les rendements financiers sont souvent non normaux (asymétrie, kurtosis, dépendance asymétrique).
- Les copules permettent de mieux modéliser la dépendance entre actifs, en particulier en queue.
- Les réseaux de neurones (NN) auraient montré de bonnes performances en prévision financière dans d'autre recherche.

Objectifs du papier

- Évaluer si des prévisions issues de modèles NN peuvent améliorer l'optimisation de portefeuille.
- Comparer différents NNs : MLP, RNN et Pi Sigma Network (PSN).
- Introduire une méthode d'optimisation basée sur :
 - des prévisions de rendements via NN ;
 - une copule asymétrique (skewed-t) pour modéliser la dépendance ;
 - le CVaR comme mesure de risque.

Objectif et Méthodologie

- **Objectif:** Évaluer si les prévisions de réseaux neuronaux améliorent significativement la gestion de portefeuille d'ETFs
- **Données:** Trois ETFs américains (SPY, DIA, QQQ) respectivement sur le S&P 500, le Dow Jones et le Nasdaq, de 2011-2015
- **Modèles comparés:**
 - Modèle de référence: ARMA
 - Réseaux neuronaux: Multi-Layer Perceptron (MLP), Recurrent Neural Network (RNN), Psi Sigma Network (PSN)
- **Tâche:** Prédiction des rendements journaliers à horizon d'un jour

Rappels : Réseaux de Neurones

- **MLP** : réseau multicouche classique (input-hidden-output), non-linéaire.
- **RNN** : prend en compte la mémoire temporelle (utile pour les séries temporelles).
- **PSN** : type de réseau à ordre supérieur, entraînement plus rapide, bonnes performances.
- Les NNs apprennent à partir des données, sans hypothèses fortes sur la forme fonctionnelle.

Multi-Layer Perceptron (MLP)

- Réseau de neurones **feedforward** composé d'une couche d'entrée, d'une ou plusieurs couches cachées et d'une couche de sortie.
- Chaque neurone effectue un calcul : somme pondérée des entrées, puis activation non linéaire (via \tanh).
- Entraînement via **rétropropagation de l'erreur** (backpropagation) et descente de gradient (SGD).
- Capable de modéliser des relations complexes non linéaires dans les données financières.

Structure d'un MLP

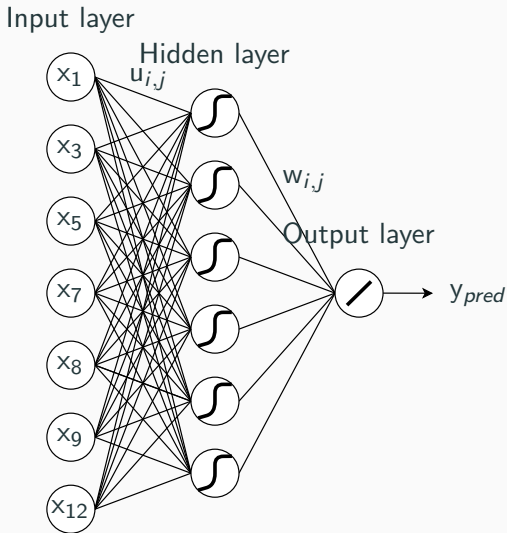


Illustration d'un réseau MLP dans le cas du SPY

Recurrent Neural Network (RNN)

- Réseau de neurones adapté aux données séquentielles, avec une **mémoire interne** pour capturer la dépendance temporelle.
- Chaque neurone reçoit non seulement l'entrée courante, mais aussi son propre état précédent, permettant de modéliser des dynamiques dans le temps.
- Entraînement via **rétropropagation à travers le temps** (Backpropagation Through Time, BPTT).
- En incorporant un effet "mémoire", ce type de modèle pourrait être plus adapté pour reconnaître des patrons dans les séries temporelles.

Structure d'un RNN

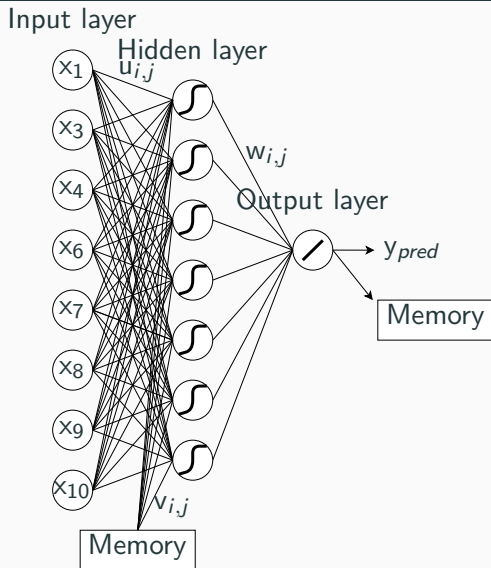


Illustration d'un réseau RNN dans le cas de DIA

Différences entre MLP et RNN

- **MLP (Multi-Layer Perceptron) :**
 - Réseau *feedforward* sans mémoire interne.
 - Traite chaque observation indépendamment.
 - Adapté aux données statiques.
- **RNN (Recurrent Neural Network) :**
 - Possède une mémoire interne via états cachés.
 - Prend en compte la dépendance temporelle entre observations.
 - Particulièrement efficace pour les séries temporelles.
- **Entraînement :**
 - MLP : backprop classique.
 - RNN : backprop à travers le temps (BPTT).

Pi Sigma Network (PSN) - Ghosh and Shin (1991)

- Type de réseau de neurones à ordre supérieur (Higher-Order Neural Network), entraîné comme un RNN.
- Combine la rapidité d'apprentissage des réseaux à une couche avec la capacité de modélisation non linéaire des réseaux plus complexes.
- Structure feedforward où les poids entre la couche cachée et la sortie sont fixes (souvent égaux à 1, également ce qui a été retenu ici).
- Seules les connexions entre l'entrée et la couche cachée sont entraînées, réduisant le nombre de paramètres.
- Performances supérieures aux MLP et RNN dans la prévision des rendements financiers d'après les auteurs.

Structure d'un PSN

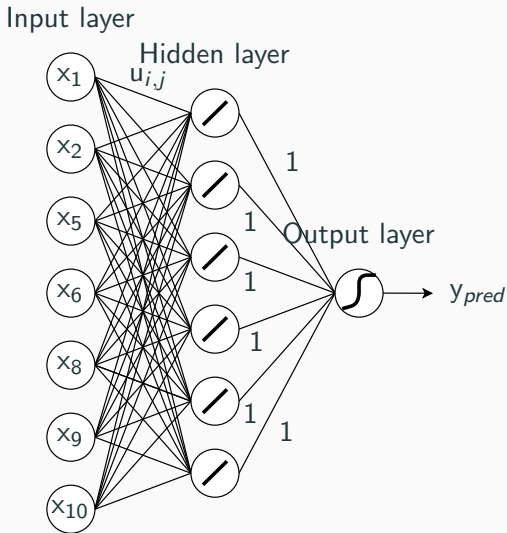


Illustration d'un réseau PSN dans le cas de DIA

Paramétrisation du modèle

- **Moyenne conditionnelle:** Modèle ARMA(p, q) (utilisation des paramètres du benchmark)
- **Variance conditionnelle:** Modèle GJR-GARCH utilisé pour sa capacité à capturer l'effet de levier et mieux réagir aux signaux négatifs du marché

Utilité

- **Construction de la stratégie:** Input essentiel pour l'optimisation de portefeuille
- **Construction des copules dynamiques**

Comparaison des modèles de corrélation dynamique

Objectif des modèles

Ces modèles permettent d'estimer la structure de dépendance temporelle entre les actifs financiers à travers l'estimations de matrices de covarainces et corrélations.

DCC vs ADCC

- **DCC (Engle, 2002)**
 - Modélise les corrélations conditionnelles de façon symétrique
- **ADCC (Cappiello et al., 2006)**
 - Intègre un terme supplémentaire pour les chocs négatifs

Modèle GAS

- **Caractéristiques**
 - Modèle récent qui Utilise le score de la fonction de densité conditionnelle
 - Paramètres évoluant selon la forme de la distribution
 - Performance supérieure dans les environnements volatils

- **Copule de Student :**

- Modélise la dépendance dans les queues
- Dépendance symétrique dans les extrêmes
 - En finance: dépendance plus forte dans les pertes que dans les gains

- **Copule de Student Asymétrique :**

- Capture : queues épaisses et asymétrie des rendements financiers
- Permet une modélisation plus réaliste : contagion financière

Approche d'Optimisation de Portefeuille

- **Approche traditionnelle:** Optimisation moyenne-variance (M-V)
- **Approche proposée:** Neural Network Copula (NNC)
 - Utilisation de la copule t asymétrique pour capturer la dépendance asymétrique entre les rendements des ETFs
 - Optimisation basée sur la moyenne-CVaR (Conditional Value-at-Risk)
 - Estimation dynamique de la matrice de corrélation: DCC, ADCC, GAS
- **Stratégie:** Portefeuilles rééquilibrés hebdomadairement, avec et sans vente à découvert

Résultats et Conclusions

- **Performance des modèles:**
 - Classement: PSN supérieur à RNN supérieur à MLP supérieur à ARMA (statistiquement et trading)
 - PSN: rendement annualisé 17,35%, Sharpe 2,08
- **Optimisation de portefeuille:**
 - Portefeuilles NN significativement supérieurs aux portefeuilles ARMA
 - Approche moyenne-CVaR surpasse l'approche moyenne-variance (amélioration de 10% du ratio Sortino)
 - Vente à découvert: amélioration de 38-42% des ratios Sharpe/Sortino pour les portefeuilles NN
- **Conclusion principale du papier:** L'amélioration des performances est principalement due aux prévisions précises des rendements plutôt qu'aux estimations de la matrice de variance-covariance

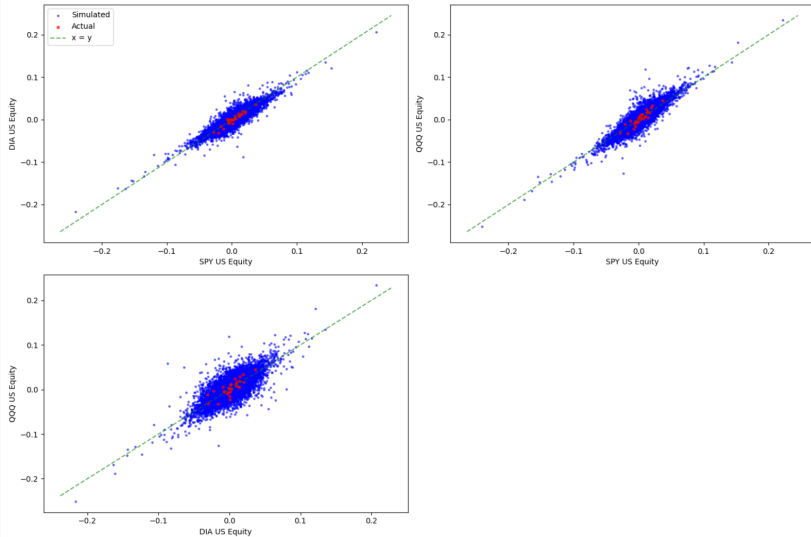
Performances of different trading strategies : Mean-95% CVaR

	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

Résultat de la réplication

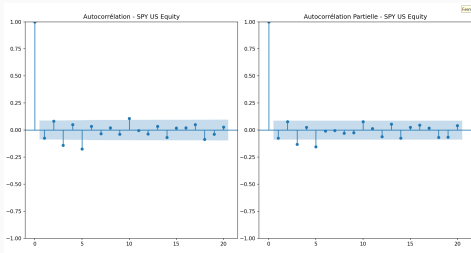
	Return (%)	Return / CVaR	Sortino	Max drawdown (%)
Panel B: Mean-Variance opt. w/o short-selling				
ARMA-DCC	3.561	0.2735	0.8122	0.360
ARMA-ADCC	3.838	0.3167	0.9356	0.220
ARMA-GAS	4.049	0.3121	0.9234	0.363
RNN-DCC	0.433	0.0196	0.0607	1.057
RNN-ADCC	0.332	0.0116	0.0376	2.458
RNN-GAS	-0.120	-0.005	-0.0156	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

Copule

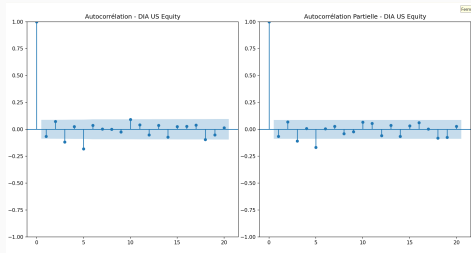


Extension

Critique du benchmark ARMA : Fonctions ACF/PACF (1/2)

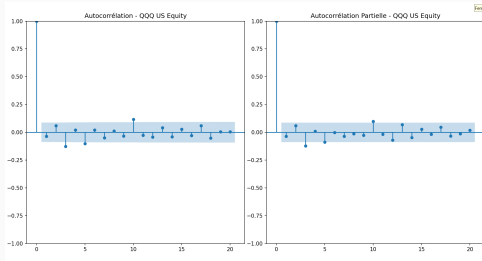


SPY ETF



DIA ETF

Critique du benchmark ARMA : Fonctions ACF/PACF (2/2)

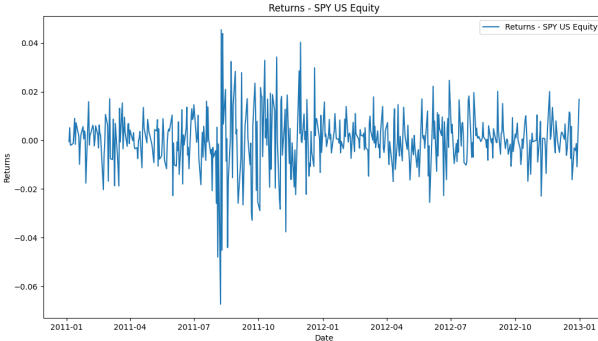


QQQ ETF

Limitations du modèle ARMA

- **Modèle mal spécifié**
- **Non adaptation aux données**
- **Impossibilité de confirmer les paramètres** : Impossible de retomber sur les paramètres du papiers

Vers un benchmark plus pertinent



Rendements du S&P 500

- Remplacer le benchmark ARMA par des modèles plus performants...
- Utiliser un modèle MS GJR GARCH

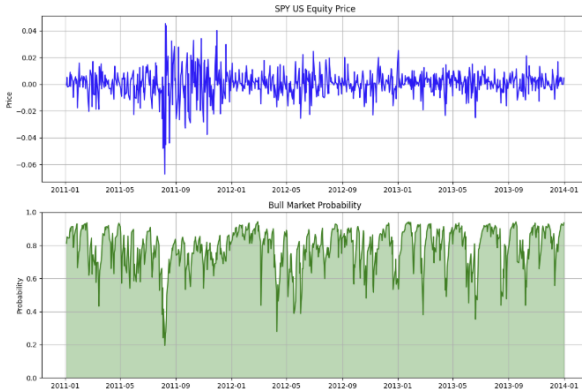
Comparaison des performances de prévision

ETF	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

Implications importantes

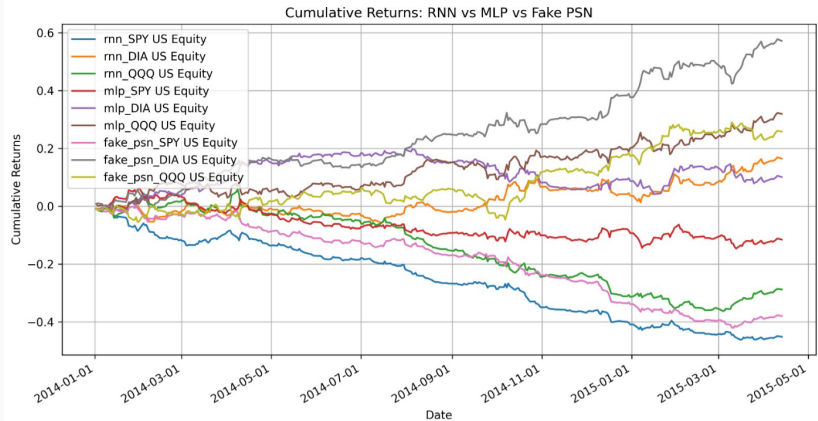
- **Performance comparable:** Notre modèle atteint des performances similaires au PSN
- **Importance du benchmark:** La remise en question du benchmark ARMA était justifiée
- **Efficacité des réseaux de neurones:** La supériorité des réseaux de neurones annoncée doit être nuancée

Analyse visuelle du marché

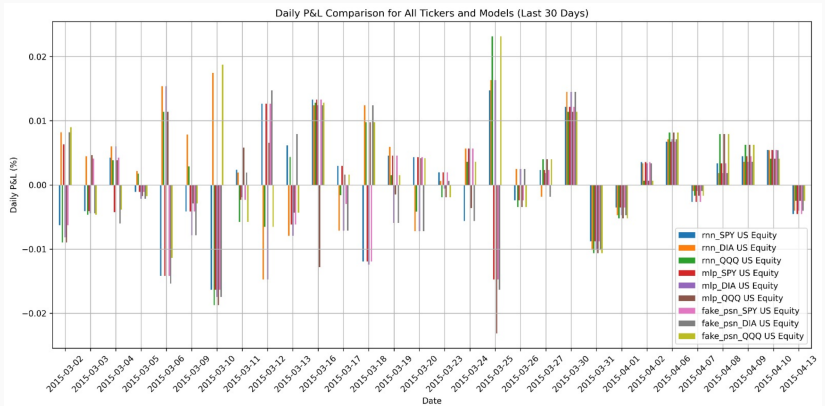


- Marché haussier sur une très grande partie de l'échantillon
- Ce qui pose plusieurs questions

Instabilité des modèles



Mesure de Performance - Trading

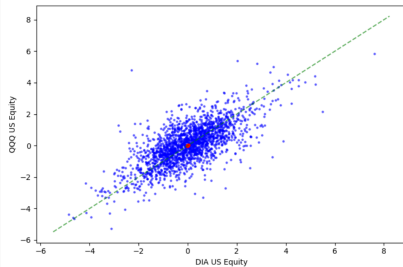
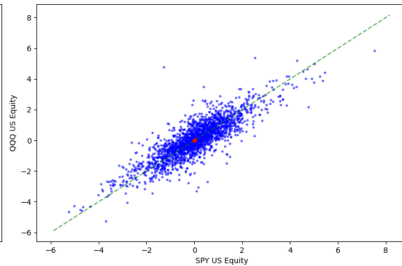
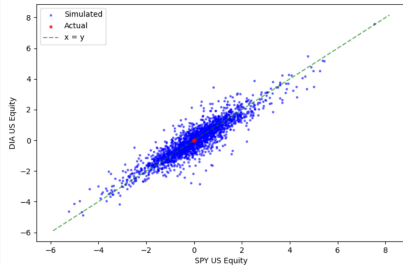


Strategy	Return	Return (A)	Volatility	Sharpe Ratio
SPY	0.4707	0.3531	0.1137	3.1061
DIA	0.4914	0.3679	0.1148	3.2049
QQQ	0.3186	0.2421	0.1390	1.7416
SPY	0.4707	0.3531	0.1173	3.0111
DIA	0.4914	0.3679	0.1150	3.2005
QQQ	-0.3186	-0.2597	0.1392	-1.8656
SPY	0.4707	0.3531	0.1151	3.0683
DIA	-0.4914	-0.4113	0.1128	-3.6468
QQQ	0.3186	0.2421	0.1394	1.7363

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

- Période lisse, pas d'événement particulier
- Extension de la réplication : 10 ans de plus

Copule sur longue période



Annexes

Rétropropagation classique (Backpropagation)

- Algorithme clé pour entraîner les réseaux de neurones feedforward (ex. MLP).
- But : ajuster les poids du réseau pour minimiser l'erreur entre prédiction et vérité.
- Fonctionnement :
 - Calcul de la sortie du réseau (forward pass).
 - Calcul de l'erreur (perte) entre sortie et cible.
 - Propagation de l'erreur en sens inverse à travers les couches.
 - Mise à jour des poids par descente de gradient stochastique.
- Permet d'apprendre des relations complexes non linéaires dans les données.

Rétropropagation à travers le temps (BPTT)

- Extension de la rétropropagation pour réseaux récurrents (RNN).
- Le réseau est “déplié” sur plusieurs pas de temps, créant un réseau feedforward profond.
- Calcul de l'erreur sur toute la séquence temporelle (forward pass sur la séquence).
- Propagation de l'erreur en sens inverse à travers tous les pas de temps (backward pass).
- Mise à jour des poids en tenant compte de la dépendance temporelle.
- Permet au RNN d'apprendre des relations temporelles complexes.

Table 1: Out-of-sample statistical performance

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

- **Modèle ARMA:**

$$\mathcal{L}(\theta) = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^T \varepsilon_t^2$$

- **Modèle DCC:**

$$\mathcal{L}(\theta) = -\frac{1}{2} \sum_{t=1}^T [n \log(2\pi) + 2 \log |D_t| + \log |R_t| + \varepsilon_t' R_t^{-1} \varepsilon_t]$$

- **Modèle ADCC:**

$$\mathcal{L}(\theta) = -\frac{1}{2} \sum_{t=1}^T [n \log(2\pi) + 2 \log |D_t| + \log |R_t| + \varepsilon_t' R_t^{-1} \varepsilon_t]$$

- **Modèle GAS:**

$$\mathcal{L}(\theta) = \sum_{t=1}^T \log f(y_t | f_t, \theta)$$

Copule II

