

Gestion de Portefeuille contenant le Bitcoin, basé sur la Prédiction par les Réseaux Neuronaux Profonds

Junhyuck SONG
Université Paris Nanterre*

Abstract

L'évolution du cours récent du bitcoin montre une nature particulièrement spéculative. La question sur la prédictibilité du cours du bitcoin révèle un défi en raison d'une faible corrélation avec d'autres classes d'actifs. C'est pourquoi la plupart des approches employées dans la littérature optent pour une approche d'apprentissage profond. Pourtant, prévoir le cours du bitcoin à court terme ne suffit pas de se couvrir pleinement contre des pertes importantes sans incorporer des stratégies de la gestion des risques. Cet article s'intéresse principalement à la possibilité d'incorporer cet actif volatile dans un portefeuille traditionnel à partir de la prédiction déduite par des méthodes d'apprentissage automatique.

Mots-clefs: Bitcoin; prédiction des prix; deep learning; allocation d'actifs; classification; LSTM

1 Introduction

Le bitcoin est une monnaie décentralisée sans intermédiaire institutionnel lancée en 2009 par Satoshi Nakamoto[14]. Il représente quasi 70% de la capitalisation sur l'ensemble des marchés de cryptomonnaie. L'évolution récente du bitcoin a ressuscité un vif intérêt chez les

investisseurs. Le cours avait atteint son pic le 16 décembre 2017, dont la valeur s'élève à 19600 en dollar, qui a retrouvé rapidement son niveau initial après 2 mois. La volatilité importante du cours offre une opportunité de spéculation attractive, ce qui explique la popularité auprès des investisseurs institutionnels et privés. Les investisseurs institutionnels s'interrogent notamment sur la possibilité de l'incorporer comme une classe d'actif alternatif pouvant diversifier leurs portefeuilles en raison d'une faible corrélation avec d'autres classes d'actifs classiques. En revanche, cette faible corrélation signifie également que les facteurs déterminants de la variation des prix ne sont pas identiques à ceux des actifs classiques. C'est-à-dire que les cryptomonnaies sont généralement faiblement corrélées avec des indicateurs fondamentaux ou ceux macroéconomiques. La principale difficulté de la valorisation du bitcoin réside également dans le fait que sa valorisation dépende du côté de la demande puisque l'offre de la monnaie est fixée par le système. En effet, la valeur du bitcoin est déterminée exclusivement par la demande de la monnaie.

En tant que classe d'actif, l'évolution du cours montre un comportement similaire comme une commodité sans partager des caractéristiques déterminants de la valorisation de ce dernier. En absence de méthode de valorisation standardisée, la plupart des approches dans la littérature optent pour une technique d'apprentissage profond capable de détecter des tendances cachées afin de prédire le cours du

*L'Université n'entend donner aucune approbation ni improbation aux opinions émises dans cette thèse. Ces opinions doivent être considérées comme propres à leur auteur

bitcoin. Pourtant, de nombreuses recherches focalisent seulement sur la prédiction à court terme voire à très court terme due au fait que la valeur de la cryptomonnaie est quasi indépendante de l'économie réelle. En terme de gestion des risques, le manque d'information sur la perspective à l'horizon d'investissement rend difficile à établir une stratégie d'investissement valable notamment en matière de l'allocation d'actifs.

Cet article s'intéresse à expérimenter une stratégie d'investissement basée sur la prédiction faite par des algorithmes supervisés et non-supervisés en incluant des variables techniques de la blockchain dans le but de vérifier une nouvelle classe d'actif potentielle. Après une brève présentation sur la littérature, la section 3 présente la méthodologie pour construire des modèles. La section 4 présente des résultats obtenus. La section 5 construit un portefeuille fictif dans le but d'expérimenter la performance des portefeuilles.

2 Revue de littérature

Le comportement du cours du bitcoin présente une particularité due à sa volatilité élevée. Les rendements moyens du bitcoin entre 2010 et 2013 sont de 7,14%[17]. En effet, les rendements que pouvait dégager le bitcoin étaient compris entre -41,78% et 136,72% pendant cette période. Les rendements élevés représentent également un risque élevé notamment du au risque de volatilité influencé fortement par des sentiment de marché. Par exemple, L'appréciation du bitcoin est conduite possiblement par des forces artificielles dues à la manipulation des marchés[4]. Toutefois, la volonté de régulation des autorités ou des cyber-attaques peuvent faire baisser significativement le cours comme une baisse du cours importante lors d'une interdiction d'Initial coin offering (ICOs) en Chine et en Corée du sud en 2018. Comme le cours est sensiblement impacté par le sentiment des investisseurs semblable à un comportement

moutonnier, l'analyse des sentiment de marché sur les réseaux sociaux fournit un indicateur prédictif pertinent[7][3].

La blockchain qui assure la fonctionnalité comptable du bitcoin contient également des informations pertinentes sur l'évolution du cours du bitcoin. Pourtant, l'étude porte sur la classification dont les variables explicatives sont basées sur des indicateurs techniques de la blockchain est relativement peu nombreuse[5].

Quant aux méthodes de prédiction, la prédiction basée sur des modèles de séries temporelles classiques comme Auto régressive moyen mobile intégré (ARIMA) est souvent employé comme benchmark pour pouvoir comparer la précision prédictive avec d'autres modèles. L'analyse des données historiques du bitcoin entre 2015 et 2018[1] étudie la performance du modèle ARIMA. La calibration du modèle a été faite de manière glissante qui minimise la somme des carrés des résidus (SCR) et qui minimise le carré moyen des erreurs (MSE) pour la prédiction. Le modèle ARIMA est capable de prévoir le cours mais l'erreur de prédiction est sensible aux changements de régime caractérisés par des multi-tendances lors d'une apprentissage. Lorsqu'il y a une unique tendance, MSE diminue considérablement qui augmente la précision prédictive. Les méthodes traditionnelles montrent bien que les modèles utilisés ne sont pas capable de capter une dépendance temporelle à long terme. Le changement des tendances rapides renforcées par la volatilité du cours rendent les modèles de séries temporelles relativement moins efficaces que les méthodes d'apprentissage automatique, capable de s'ajuster les paramètres optimaux faces aux propriétés non-linéaires de la série.[13][10]

Les méthodes de classification supervisée telles que la forêt aléatoire, la régression logistique et l'analyse discriminant montrent que leurs précisions prédictives sont supérieures à celle des méthodes classiques[15]. Les études simi-

lares en utilisant la forêt aléatoire[8], réseaux de neurones bayésiennes[6][5] support vecteur machine (SVM), XGBoost[2] montrent également les méthodes de classification supervisées ont une précision prédictive supérieure aux modèles traditionnels de séries temporelles dans le cadre de la prévision de la cryptomonnaie. Contrairement aux approches précédentes, les techniques d'apprentissage profond (Deep Learning) sont capables à apprendre la dépendance temporelle. Parmi les méthodes utilisées dans la littérature, la structure LSTM s'avère la plus efficace en raison de ses capacités d'apprendre non seulement la dépendance temporelle à court terme mais aussi celle à long terme. Réseaux de neurones récurrentes (RNN) qui est la structure d'origine de LSTM ne peut pas apprendre la dépendance temporelle à long terme à cause du problème de la disparition du gradient. Par exemple, L'implémentation en utilisant la structure LSTM[9] montre que grâce à sa structure complexe contenant une porte permettant la remise à zéro (forget gate) le réseau de neurones à mémoire court et long terme est capable d'augmenter la précision prédictive. Egalement les travaux sur les variants de la structure LSTM comme LSTM à convolution et AR-LSTM montrent davantage d'amélioration dans la prévision du cours[12][16].

3 Méthodologie

La méthode sélectionnée consiste à classer la variable expliquée étant des rendements transformée en binaire selon l'équation (2) en s'intéressant au pouvoir prédictif des indicateurs techniques de la blockchain. Afin de tester la pertinence de la prévision basée sur des variables de la blockchain, la procédure d'une méthodologie classique[11] s'applique en 6 étapes comme dans la figure 1. L'objectif étant de classer le cours de demain, la variable expliquée est transformée en variables binaires si les rendements d'aujourd'hui sont supérieur à ceux d'hier égaux à 1 sinon 0.

$$S_t = \log(p_t) - \log(p_{t-1}) \quad (1)$$

$$S_t = \begin{cases} 1 & \text{si } S_t \geq 0 \\ 0 & \text{Sinon} \end{cases} \quad (2)$$

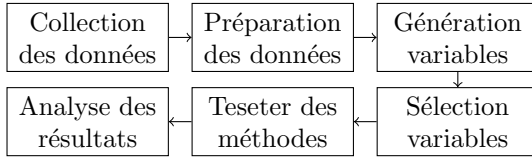
Les données sont téléchargeables sur *Quandl* y compris le cours du bitcoin. Afin de limiter le nombre des observations manquantes sur les variables de la blockchain, la période est fixée entre le 1 septembre 2010 et le 31 décembre 2019 soit 3408 observations. Les données¹ comprennent 24 variables de la blockchain, du cours et des rendements en pourcentage auxquelles rajoutent 4 variables financières. Comme la plupart des données sont non-stationnaires², les variables sont différenciées afin de les rendre stationnaire. Avant l'entraînement par l'algorithme dans le but de faire converger rapidement, les données sont normalisées entre 0 et 1. Enfin, dans l'étape de la préparation des données, la fenêtre m est choisit de manière à pouvoir tester la performance prédictive. C'est-à-dire, les données d'entraînement requises pour avoir une certaine précision prédictive.

Pour la notation, soit S est nos données entières, i et j sont des points de i ème et de j ème observation. La séquence entre i et j est noté $S[i : j]$. Soit X une bloc de donnée découpée contenant m jours, $X[i]$ est i ème observation dans la bloc des données X , $X[i][j]$ représente la séquence dans X . En effet, la séquence commence par $S[1 : m]$ jusqu'à $S[observationtotale - m + 1 : observationtotale]$. Par conséquent, la séquence de $[i : i + m - 1]$ est entraînée par l'algorithme consécutivement pour prévoir $(i + m)$ ème jour.

¹Les variables sont reportées dans l'Annexe

²Le test de la stationnarité sur les 24 variables de la blockchain par Augemented-Dicky Fuller et KPSS tests est réalisé. Si la variable est non-stationnaire, la différence première a été faite afin de la stationnariser. Pour les variables financières, log-rendement transformation a été appliquée

Figure 1: Procédure des données



Pour le choix d'algorithme de classification, la régression logistique, la machine à vecteur de support (SVM) et la forêt aléatoire ont été sélectionnées. Pour les algorithmes d'apprentissage profonds, simple DNN, CNN, et LSTM ont été choisis. L'étape de la sélection des variables se fait en tenant compte des méthodologies qui seront utilisées pour des problèmes de classification. Par exemple, pour la régression logistique, la sélection des variables a été faite par la méthode Bayesian Model Averaging (BMA). Les variables sélectionnées ont ensuite été entraînées par la régression logistique par itération grâce à la Descente graduelle (Gradient Descent). Afin de sélectionner des variables pertinentes pour les algorithmes d'automatique, la sélection des variables par la forêt aléatoire a été employée. Enfin, afin de pouvoir comparer la performance prédictive de différentes méthodes, les différentes variables sélectionnées ont été testées sur l'ensemble des méthodes.

4 Résultats

Les variables sélectionnées sont les suivantes. *ntratl*, *blchs*, *mirev*, *cptra*, *mktp*, (*avbls*)³ étant respectivement le nombre total des transactions du bitcoin, la taille du BTC de la blockchain.api, le revenu des mineurs, le coût du BTC par transaction, la capitalisation du BTC et le volume moyen d'un bloc en MB ont été sélectionnés par la forêt aléatoire. Pourtant, le fait de la nature stochastique de la forêt aléatoire, les variables sélectionnées ont été testées de différentes combi-

³Omission de cette variable a résulté d'une augmentation significative de la précision prédictive

Table 1: Précision de la prévision

m	DNN	CNN	LSTM	SVM
7	60.95%	54.65%	54.22%	50.35%
30	56.35%	54.15%	53.73%	50.12%
90	56.78%	52.92%	47.08%	50.32%
180	61.30%	51.40%	51.41%	50.07%

naisons afin de diminuer l'incertitude du modèle.

La fenêtre m étant le nombre du jour pour l'entraînement est utilisée $m = 7, 30, 90, 180$ puisque les marchés de cryptomonnaie sont ouverts tous les jours. L'intérêt de cette fenêtre de prévision consiste à vérifier l'existence d'une tendance à long terme si les algorithmes fournissent une précision élevée. Les données sont découpées en 70%, 20% et 10% pour l'entraînement, la validation et la prévision. Dans la présentation des résultats, les résultats de la régression logistique ont été omis du fait qu'ils ne satisfont pas une précision requise.

Le tableau 2 correspond aux résultats sur la précision de la classification dont les variables explicatives sont *ntratl*, *blchs*, *mirev*, *cptra*, *mktp*. L'omission d'une des variables explicatives résulte d'une chute à peu près 10% sur la précision prédictive. Contrairement au résultat attendu, la simple DNN structure donne le meilleur résultat. CNN et LSTM donnent des résultats assez similaires. Quant à la fenêtre, le résultat est difficilement interprétable mais au-delà d'une fenêtre plus de 180, la précision de la classification est généralement plus de 60%.

5 Gestion de portefeuille

Pour vérifier la performance basée sur la prévision de la classification du bitcoin dans un portefeuille fictif, les différentes fenêtres de 7, 30, 90, 180 sont employées. Par souci de simplicité, nous considérons 10 actifs dans le portefeuille

Table 2: Statistiques descriptives des actifs fictifs

Variable	Moyen	Std	Skew	Kurt
$CL = F$	-0.03%	0.022	-0.10	2.36
$EURUSD$	-0.01%	0.0053	0.14	3.35
$EURGBP$	-0.01%	0.0058	-1.43	22.66
$GC = F$	0	0.0096	-0.75	9.69
$USDJPY$	0.01%	0.0061	-0.12	5.59
$IXIC$	0.06%	0.01	-0.68	2.35
$N225$	0.04%	0.0134	-0.37	4.50
RUT	0.04%	0.106	-0.48	1.33
TNX	0	0.0205	0.13	2.35
BTC	0.17%	0.0429	0.041	4.10

comprenant le bitcoin. La corrélation de Spearman des actifs est présentée dans la figure 2. Tous les actifs a un poids égal à 10% et qu'il est constant dans le temps. Par contre, le poids alloué au bitcoin est modifié en fonction du signal entrant. Si la prévision donne 0, le poids alloué devient 0. Sinon, nous gardons le bitcoin dans notre portefeuille. Le poids maximal alloué de chaque actif est égal à 10% et que le signal à la hausse consécutive ne modifie pas la configuration générale de notre portefeuille. Le portefeuille composé de 9 actifs en excluant le bitcoin lorsque le signal reçu est négatif reste à 10%. La prévision est faite sur la période entre le 14 mars 2017 et le 31 décembre 2019 soit 1023 observations.

Ces hypothèses stipulent qu'il n'y ait pas de coût de transaction et la parfaite liquidité qui ne sont pas réalistes. Néanmoins, le point central est de montrer que cet actif peut fournir une possibilité comme un actif alternatif capable d'être un moyen de diversification. Le portefeuille de benchmark étant le portefeuille sans incorporer le modèle prédictif sur le bitcoin à cette période montre des rendements de -0.03 % avec l'écart-type de 0.4405.

La figure 3 montre que le portefeuille incluant le bitcoin génère plus de rendements. Pour l'écart-type du portefeuille, il faut l'interpréter avec précaution car le portefeuille sans bitcoin

Figure 2: Rendements cumulatifs normalisés des actifs

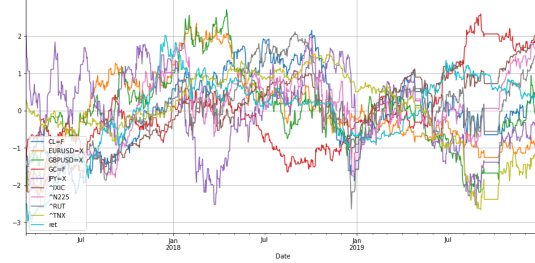


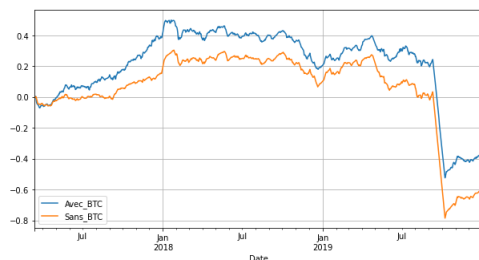
Table 3: Rendements et l'écart-type des portefeuilles

	m = 7	m = 30	m = 90	m = 180
Précision	65.61%	62.55%	64.55%	64.79%
$Moyen^1$	0.004	0.013	-0.017	0.0247
Std^1	0.4396	0.4383	0.4345	0.4384
$Sharpe^1$	0.0001	0.0003	-0.0004	0.00056
$Moyen^2$	-0.0593	-0.0593	-0.0593	-0.0593
Std^2	0.4637	0.4683	0.46371	0.4637
$Sharpe^2$	-0.0013	-0.0012	-0.0012	-0.0013

^{1,2} sont respectivement le portefeuille avec bitcoin et sans bitcoin

contient 9 actifs alors que l'autre portefeuille contient 10 actifs. Grâce à la propriété de sous-additivité, la variance d'un portefeuille diminue si la corrélation des actifs n'est pas significativement corrélée. Mais le portefeuille contenant du bitcoin montre le ratio de Sharpe plus élevé que celui sans bitcoin. Le portefeuille sans bitcoin utilisé comme benchmark montre bien que le portefeuille avec bitcoin connaît un effet de diversification qui génère plus de rendements.

Figure 3: Rendements des portefeuilles Avec et Sans bitcoin



des risques des actifs classiques. Bien que la corrélation entre les actifs de cryptomonnaies et ceux classiques soit faible, les études approfondies sur le comportement des actifs en période de stress ne sont pas analysées dans la littérature. A priori, la difficulté majeure est l'indisponibilité des données des cryptomonnaies relatives à celles d'autres d'actifs car les données de cryptomonnaie ne sont disponible qu'à partir de 2011.

6 Conclusion et Discussion

Cette simple implémentation d'un outil basé sur les réseaux de neurones profonds montre que le bitcoin peut être une nouvelle classe d'actif pouvant offrir un moyen de diversification malgré sa volatilité extrêmement élevée. En revanche, cette simple approche a posée quelques hypothèses au départ comme la parfaite liquidité et les coût de transaction nuls. Pourtant, les coût de transactions sont certainement un élément important puisque cette approche nécessite une vente immédiate lorsque nous recevons un signal négatif. En réalité, la liquidité du bitcoin n'est certainement pas instantanée. La médiane de la confirmation étant de 8 heures dans nos données, le temps de la réalisation d'une transaction dépend sensiblement des périodes de l'euphorie et de la chute brutale du cours, ceux sont des périodes de la illiquidité haute. L'accumulation du coût de transaction influence des rendements totaux du portefeuille contenant le bitcoin si la fréquence de la transaction est courte.

L'autre problème consiste à vérifier l'avantage du portefeuille contenant le bitcoin pour des périodes pendant lesquelles les marchés connaissent une tension extrême telles que la période de la bulle spéculative, l'écartement de cette bulle et celles d'une inflation élevée afin de voir si le portefeuille présente une protection contre

Annexe

Table 4: Liste des variables

Variables	Description
cptrv	les données sur Mt.gox du prix du BTC en USD dollar
diffy	BTC difficulté
avbls	le volume moyen d'un bloc en MB
abrct	Le temps de la confirmation d'une transaction en médiane
hrate	BTC Hash rate
cptrv	Le coût en pourcentage du volume de transactions
etrv	Le volume des transactions estimées du BTC
toutv	Le volume total d'Output
ntrbl	Le nombre de transaction du BTC par bloc
naddu	Le nombre d'adresse unique du BTC utilisé
nrep	Le nombre de transactions en excluant les adresses populaires
ntrat	Le nombre total des transactions du BTC
ntran	Le numéro du BTC en transaction
trfee	La commission des transactions totales du BTC
totbc	BTC total débloqué
mirev	Le revenu des mineurs
cptra	Le coût du BTC par transaction
trvou	Le volume du BTC en échange en USD dollar
etrvu	Le volume des transactions estimées du BTC en USD dollar
trfus	La commission totales du BTC en USD dollar
mktp	La capitalisation du BTC
blchs	La taille du BTC de la blockchain api
mwntd	Le nombre des transactions journalier de "My Wallet"
mwnus	Le nombre des utilisateurs de "My Wallet"
mwtrv	Le volume des transactions du BTC sur "My Wallet"
gold	Les log rendements de l'or sur CME
oil	Les log rendements du pétrole brut sur CME
bond	US 10Y obligation
sp500	Les rendements de Future sp500

Table 5: Corrélation Pearson des log-Rendements des actifs fictifs

	CL=F	EURUSD=X	GBPUSD=X	GC=F	JPY=X	IXIC	N225	RUT	TNX
CL=F	1.0	0.0083	0.0204	0.0559	-0.0233	0.2366	0.0517	0.2817	0.2369
EURUSD=X	0.0083	1.0	0.5386	0.0727	-0.384	0.0092	-0.1111	0.0146	0.0551
GBPUSD=X	0.0204	0.5386	1.0	0.016	-0.1413	0.0594	0.0648	0.0965	0.0698
GC=F	0.0559	0.0727	0.016	1.0	-0.1149	-0.0862	-0.1404	-0.0721	-0.3342
JPY=X	-0.0233	-0.384	-0.1413	-0.1149	1.0	0.001	0.4144	-0.0111	0.0401
IXIC	0.2366	0.0092	0.0594	-0.0862	0.001	1.0	0.1696	0.8757	0.3205
N225	0.0517	-0.1111	0.0648	-0.1404	0.4144	0.1696	1.0	0.1529	0.1258
RUT	0.2817	0.0146	0.0965	-0.0721	-0.0111	0.8757	0.1529	1.0	0.3506
TNX	0.2369	0.0551	0.0698	-0.3342	0.0401	0.3205	0.1258	0.3506	1.0

References

- [1] Azari Amin. “Bitcoin Price Prediction: An ARIMA Approach”. In: (2019).
- [2] Li C et Sun W Chen Z. “Bitcoin Price Prediction using Machine Learning: An approach to Sample Dimension Engineering”. In: *Journal of Computational and Applied Mathematics* 365 (2019).
- [3] Mai F et Bai Q et Shan Z et Wang X et Chiang R. “From bitcoin to big coin: The impacts of social media on bitcoin performance”. In: *SSRN Electronic Journal* (2015).
- [4] Hamrick JT et Moore T et Obermen Tali Gandal N. “Price Manipulation in the Bitcoin Ecosystem”. In: *Journal of Monetary Economics* (2017).
- [5] Lee J Jang G. “An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information”. In: *IEEE Access* 6 (2018), pp. 5427–5437.
- [6] Lee J et Lee W Jang H. “Predicting Bitcoin Prices Using Rolling Windows LSTM model”. In: *In Proceedings of the KDD Data Science in Fintech Workshop* (2018).
- [7] Chen R et Lazar M. “Sentiment Analysis of Twitter Feeds for the Prediction of Stock Market Movement”. In: (2016).
- [8] Saluja S et Zhao A Madan I. “Automated bitcoin Trading via Machine Learning Algorithms”. In: (2015).
- [9] Roche J et Caton S McNally S. “Predicting the Price of Bitcoin Using Machine Learning”. In: *In Processing of the 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing, Cambridge, UK* (2018), pp. 339–343.
- [10] De Guzman R.A.M et Mike K.P.SO. “Empirical Analysis of Bitcoin Prices using Threshold Time Series Models”. In: *Annals of Financial Economics* 13(4) (2018).
- [11] Schutt R et O’Neil C. “Doing Data Science”. In: *Sebastopol: O’Reilly* (2014).
- [12] Shintate T et Pichl L. “Trend Prediction Classification for High Frequency Bitcoin Time Series with Deep Learning”. In: *Journal of Risk and financial Management* 12(1) (2019).
- [13] Nanjiba S et Chakrabarty A Roy S. “Bitcoin Price Forecasting Using Time Series Analysis”. In: *21st International Conference of Computer and Information Technology* (2018).
- [14] Nakamoto S. “Bitcoin: A Peer-to-Peer Electronic Cash System”. In: *Technical Report* (2009).
- [15] Amjad M et Shah D. “Trading Bitcoin and Online Time Series Prediction”. In: *NIPS Time Series Workshop* (2016).
- [16] J He K et Zhang X et Ren S et Sun. “Deep Residual Learning for Image Recognition”. In: *In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA* (2016), 770–778.
- [17] Brière M et Oosterlinck K et Szafarz. “Virtual Currency, Tangible Return: Portfolio Diversification with Bitcoin”. In: *Working Papers CEB* (2015), pp. 1–18.