

HEC MONTRÉAL

**Algorithmes d'apprentissage optimisé via une recherche de paramètres
Baysien dans un contexte d'investissement**

**par
Charles Rousseau**

**Sciences de la gestion
(Option Finance de marché)**

*Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)*

Août 2019
© Charles Rousseau, 2019

Abstract

Nous proposons une stratégie d'investissement grandement inspirée par le « *momentum* » (MOM) de Jegadeesh et Titman (1993) qui vise à prédire le rendement moyen mensuel futur standardisé par la volatilité de chaque titre du S&P500 entre juillet 1998 et juin 2018. La prédiction de cette variable est réalisée à travers des réseaux de neurones dont les paramètres sont optimisés selon une méthode d'optimisation séquentielle par modèle. Cette méthodologie a l'avantage d'être exhaustive et rapide tout en permettant d'omettre certaines hypothèses restrictives quant à la configuration des réseaux de neurones. En tenant compte du risque des titres boursiers du S&P500, en exploitant un énorme éventail de variables explicatives et en prenant avantage de la puissance d'apprentissage par expérience des réseaux de neurones, nous sommes en mesure d'obtenir une stratégie d'investissement 2 fois moins volatile que la stratégie MOM de référence sur le même marché. En termes de performance, les réseaux de neurones génèrent un rendement moyen annuel excédentaire de plus de 6%. Ainsi, la stratégie d'investissement est en mesure d'exploiter des relations qui vont au-delà des prix antérieurs des titres du S&P500.

Keywords: Momentum, Artificial Neural Network, Sequential model-based optimisation, risk premia

Table des matières

ABSTRACT	I
TABLE DES MATIÈRES.....	III
LISTE DES TABLEAUX ET DES FIGURES	V
LISTE DES ABRÉVIATIONS	IX
REMERCIEMENTS	XI
INTRODUCTION.....	1
CHAPITRE 1 REVUE DE LITTÉRATURE.....	5
CHAPITRE 2 MÉTHODOLOGIE.....	16
CHAPITRE 3 RÉSULTATS EMPIRIQUES.....	31
CONCLUSION.....	41
BIBLIOGRAPHIE	43
ANNEXES.....	47

Liste des tableaux et des figures

Figure 1: Dynamique d'un réseau de neurones ANN classique ne possédant qu'une seule couche cachée et un neurone de sortie. Les cercles sur la première couche du réseau Input layer correspondent aux variables explicatives. Les cercles sur les couches subséquentes représentent l'information transformée et transmise aux neurones de la couche suivante. Le graphique du bas illustre en détail le processus de propagation et de transformation de l'information de chaque neurone.

Figure 2: Portefeuille gagnant et perdant entre juillet 1998 et juin 2018. Les rendements cumulés présentés sont : (1) et (2) le rendement respectif du portefeuille du premier et dernier décile selon l'historique de la performance des 12 derniers mois détenus durant 1 mois, sans compter la performance du dernier mois, (3) le rendement de l'indice du marché, le S&P500 et (4) le rendement du taux sans risque, soit le rendement d'une obligation américaine 1 mois. À droite du graphique est présentée la valeur cumulée de chaque portefeuille. (1), (2) et (3) sont présentés en excès par rapport au taux sans risque.

Figure 3: Portefeuille gagnant et perdant entre mars 2009 et juin 2011. Les rendements cumulés présentés sont : (1) et (2) le rendement respectif du portefeuille du premier et du dernier décile selon la performance historique des 12 derniers mois détenus durant 1 mois, sans compter la performance du dernier mois, (3) le rendement de l'indice du marché, le S&P500 et (4) le rendement du taux sans risque, soit le rendement d'une obligation américaine 1 mois. À droite du graphique est présentée la valeur cumulée de chaque portefeuille. (1), (2) et (3) sont présentés en excès par rapport au taux sans risque.

Figure 4: Entre juillet 1998 et juin 2018, les rendements cumulés présentés sont : (1) et (2) le rendement cumulé du 1^{er} et du 9^e décile selon la prédiction ANN. Ces déciles représentent respectivement le décile de portefeuille avec la pire et meilleure performance. (3) le rendement de l'indice du marché, le S&P500 et (4) le rendement du taux sans risque, soit le rendement d'une obligation américaine 1 mois. À droite du graphique est présentée la valeur cumulée de chaque portefeuille. (1), (2) et (3) sont présentés en excès par rapport au taux sans risque.

Figure 5: Les rendements cumulés présentés sont : (1) stratégie MOM sur le marché du S&P500 provenant du portefeuille qui est long et court les titres ayant connu les meilleurs et pires rendements sur les 12 derniers mois, (2) la stratégie WML provenant du portefeuille qui long et court le 9^e et le 1^{er} décile respectivement en fonction de la prédiction des réseaux ANN. Le 9^e et le 1^{er} décile ont respectivement la meilleure et la pire performance empirique. (3) le rendement de l'indice du marché, le S&P500 et (4) le rendement du taux sans risque, soit le rendement d'une obligation américaine 1 mois. À droite du graphique est présentée la valeur cumulée de chaque portefeuille. (1), (2) et (3) sont présentés en excès par rapport au taux sans risque.

Tableau 1: Performance et caractéristiques des portefeuilles d'investissement des déciles MOM entre juillet 1998 et juin 2018. Les portefeuilles 1 et 10 incluent respectivement les firmes ayant connu les pires et meilleurs rendements sur les 12 derniers mois dans l'univers du S&P500, sans compter la performance du dernier mois. MOM correspond au portefeuille autofinancé qui est long et court respectivement le portefeuille 10 et 1. Les rendements moyens excédentaires, écarts types et alpha (MEDAF) sont en pourcentage. SR correspond au ratio de Sharpe, Sk correspond au coefficient d'asymétrie de la distribution des rendements et S&P500 correspond à l'indice du marché de référence. Les portefeuilles sont détenus sur une base mensuelle.

Tableau 2: Performance et caractéristiques des portefeuilles d'investissement ANN par déciles entre juillet 1998 et juin 2018. Les portefeuilles sont classés en décile en fonction de la prédiction de chaque titre de l'univers du S&P500. WML correspond au portefeuille qui est long et court respectivement le portefeuille 9 et 1; le portefeuille est long et court les déciles avec la meilleure et la pire performance. Les rendements moyens excédentaires, écarts types et alpha (MEDAF) sont en pourcentage. SR correspond au ratio de Sharpe, Sk correspond au coefficient d'asymétrie de la distribution des rendements et S&P500 correspond à l'indice du marché de référence.

Tableau 3 : Performance et caractéristiques de la stratégie MOM et de la stratégie d'investissement issue des réseaux de neurones ANN entre juillet 1998 et juin 2018. MOM est le portefeuille autofinancé qui est long et court les titres ayant connu les meilleurs et pires rendements sur les 12 derniers mois, sans compter le dernier mois. WML est le portefeuille autofinancé qui long et court le 9^e et le 1^{er} décile en fonction de la prédiction effectuée sur les titres du S&P500. SR correspond au ratio de Sharpe, Max Pertes Cum correspond au plus faible rendement cumulé réalisé sur la période et % Rend. Positif correspond à la proportion de rendement positif réalisé sur la période. La portion du bas présente les coefficients *Beta* aux trois facteurs de Fama-French (1996) ainsi que le facteur MOM de Cahart (1997). Les coefficients *Beta* des deux stratégies sont présentés accompagnés de leur statistique de test T en italique. Les facteurs de risques sont respectivement Mkt-rf le facteur du marché, SMB le facteur taille, HML le facteur valeur et UMD au facteur de risque MOM.

Liste des abréviations

ANN : Réseau de neurones artificiels (*Artificial Neural Network*)

CRSP: Center for Research in Security Prices

DR: Dropout rate

EI: Expected improvement

LR: Learning rate

MEDAF: Modèle d'évaluation d'actif financier

MLC : Momentum life cycle

MOM : Facteur d'investissement associé au rendements positifs générés par la surperformance et sous-performance des titres financiers ayant respectivement bien et moins bien performé dans les mois derniers.

RMS: Root Mean Squared Propagation

SGD: Stochastic Gradient Descent

SMBO: Sequential model-based optimisation

TPE: Tree-structured Parzen Estimator

TSM: Time series momentum

WML : Portefeuille d'investissement qui est long et court selon une règle d'investissement (*Winners minus Losers*).

Remerciements

Je tiens tout d'abord à remercier ma famille qui m'a grandement appuyé depuis le début de mes études. Un remerciement spécial pour ma mère, mon père et ma copine qui ont été à mes côtés durant tout ce processus. Merci pour votre patience, votre soutien moral et votre positivisme. Je tiens également à remercier mon bon ami Alex qui m'a grandement motivé et aidé à travers plusieurs étapes de mon mémoire. Merci à vous tous, sans vous, la rédaction aurait été beaucoup plus difficile.

Enfin, je remercie mon superviseur Christian pour m'avoir encouragé à entreprendre un mémoire. Cette expérience a été un grand défi à travers lequel j'ai beaucoup appris. Son éthique de travail est très inspirante et m'a motivé tout au long de ce projet.

Introduction

L'investissement algorithmique ou systématique repose entre autres sur la modélisation de signaux et de règles d'investissement bien définis. Dans leurs études sur le facteur « momentum » (MOM), Jegadeesh et Titman (1993) ont mis de l'avant une telle règle d'investissement; l'achat et la vente de titres boursiers se font uniquement sur la base de leurs rendements historiques. Les recherches qu'ils ont effectuées sur le MOM sont à la base de notre stratégie d'investissement. Typiquement, la prime associée au facteur (MOM) provient de l'achat et de la vente respectifs de titres boursiers ayant surperformés et sous-performés. Le rendement excédentaire généré par ce facteur d'investissement a suscité la curiosité de plusieurs chercheurs. Certains se sont penchés sur des explications potentielles à la rentabilité à court terme du facteur d'investissement. Fama et French (1996) ont développé un modèle à trois facteurs dans l'optique d'expliquer le rendement MOM; le modèle n'est qu'en mesure de capter le comportement à long terme du facteur. Cahart (1997) introduit un quatrième facteur, qui réplique explicitement les rendements MOM. D'autres auteurs ont tenté de décortiquer les sources de profits du facteur MOM. Moskowitz et Grinblatt (1999) illustrent qu'une très grande portion des rendements est attribuable aux différents secteurs d'investissement. Cooper et Gutierrez (2004) et Chordia et Shivakumar (2002) illustrent, quant à eux, que la rentabilité de la stratégie d'investissement est dépendante de l'état de l'économie au moment de la formation des portefeuilles d'investissement. Certains chercheurs ont illustré que l'effet MOM n'est pas limité au marché des actions américaines. Au contraire, la méthodologie utilisée par Jegadeesh et Titman (1993) permet de dégager des rendements excédentaires significatifs positifs sur différents marchés internationaux et plusieurs véhicules d'investissement tels que les commodités et les devises. Finalement, plusieurs auteurs se sont intéressés aux diverses améliorations potentielles qu'ils pouvaient apporter au profil rendement risque de la stratégie MOM. Parmi les améliorations proposées, on retrouve celle de Lee et Swaminathan (2000) et Grundy et Martin (2001) qui proposent, entre autres, deux manières alternatives de faire le regroupement des portefeuilles dans la construction du facteur. Le premier propose d'inclure un filtre sur le volume transigé, tandis que le second propose de former les portefeuilles sur la base du rendement excédentaire inexplicé par les facteurs de risque usuel. Dans les deux cas, les chercheurs dégagent un rendement supplémentaire aux rendements MOM. On retrouve également

Barroso et Santa Clara (2015), Kent et Moskowitz (2016) et Moskowitz, Ooi, Pedersen (2012) qui analysent des stratégies alternatives en tenant compte du risque inhérent des rendements MOM; le rendement des actifs considérés est standardisé par la volatilité historique. Ainsi, dans leurs travaux respectifs, les auteurs sont en mesure de réduire le risque des grandes pertes associées au facteur MOM. Les multiples travaux effectués sur le MOM ont suscité la question suivante : est-ce qu'il y aurait une manière d'exploiter conjointement toutes les propriétés et les sources d'amélioration du facteur MOM pour construire une stratégie d'investissement profitable? Pour répondre à cette question, nous nous sommes tournés vers les algorithmes d'apprentissage utilisés dans le développement de l'intelligence artificielle. Plus précisément, nous allons exploiter des réseaux de neurones. Depuis la dernière décennie, les réseaux de neurones ont su se démarquer, non seulement pour leurs gains en performance dans divers domaines, mais également par leur accessibilité; plusieurs plateformes et plusieurs langages informatiques réduisent la barrière à l'entrée pour ses utilisateurs. En finance, les articles de Sirignano, Sadhwani et Giesecke (2018) et Gu, Kelly et Xiu (2018) sont de très bons exemples de recherches qui ont exploités divers algorithmes d'apprentissage sophistiqués. Dans la première recherche, les auteurs exploitent des réseaux de neurones afin d'étudier la relation entre les prêts hypothécaires et le risque de paiements des emprunteurs. Dans la deuxième recherche, ces auteurs exploitent, entre autres, les réseaux de neurones dans une stratégie d'investissement. Dans les deux cas, ils sont en mesure de justifier la valeur ajoutée associée à l'utilisation d'algorithmes d'apprentissage et démontrent que leurs utilisations gagnent à être exploitées davantage dans le domaine de la finance.

De manière générale, un réseau de neurones transforme les variables explicatives qui lui sont initialement fournies à l'aide d'une série d'opérations linéaires qui permettent de modéliser la variable d'intérêt. La capacité d'introduire de la non-linéarité à la résolution de problèmes complexes et sa possibilité d'apprentissage par expérience sont parmi les propriétés qui rendent un réseau de neurones si puissant. Par construction, l'exploitation de réseaux de neurones nécessite plusieurs hypothèses quant à l'ensemble des paramètres qui les décrivent. Chaque paramètre qui peut influencer la dynamique du réseau est un hyperparamètre qui nécessite d'être spécifié. Afin d'émettre aucune hypothèse restrictive sur les différentes valeurs des hyperparamètres de nos réseaux de neurones, nous utilisons un optimisateur d'hyperparamètres. Nous avons opté pour la méthodologie de Bengio et Bergstra (2012) qui propose une méthode d'optimisation séquentielle par modèle (« *Sequential model-based optimisation* »). En soi, cette méthodologie implique de

spécifier un univers complet d'hyperparamètres potentiels à partir desquels les réseaux de neurones seront construits. La configuration des hyperparamètres optimaux est celle qui minimise la fonction de coût du réseau. Cette méthodologie a l'avantage d'être exhaustive et rapide tout en permettant d'omettre certaines hypothèses restrictives quant à la forme des réseaux de neurones.

Dans notre recherche, nous avons analysé l'univers du S&P500 entre juillet 1998 et juin 2018. La stratégie d'investissement consiste à prédire le rendement futur moyen mensuel standardisé par la volatilité de chaque titre de notre univers d'investissement. Le choix de modéliser cette variable est grandement influencé par le ratio de Sharpe, qui mesure l'espérance de rendement attendu d'un titre en excès au rendement sans risque, le tout standardisé par la volatilité de l'actif. La variable que nous modélisons est un bon intermédiaire au ratio de Sharpe qui nous permet de gérer implicitement le risque et de lisser le rendement futur de chaque actif. Pour faire notre prédiction, nous exploitons des réseaux de neurones optimisés selon la méthodologie de Bengio et Bergstra (2012). Ainsi, avec une structure neuronale optimisée et un large éventail de variables descriptives, nous sommes en mesure de faire la prédiction de notre variable d'intérêt. Avec l'ensemble des prédictions effectuées sur chaque titre de l'univers d'investissement, nous utilisons la méthodologie de Jegadeesh et Titman (1993) pour construire un portefeuille équilibré et autofinancé; l'objectif est de créer un portefeuille d'investissement qui est long et court respectivement un portefeuille de titres avec un grand et faible signal d'achat. Sur notre période d'investissement, la stratégie de référence MOM sur le marché du S&P500 dégage un rendement annuel excédentaire moyen de 0.99%, avec une volatilité annuelle moyenne de 33.63% et une exposition au marché (« *beta* ») de -1.011. Bien que positive sur l'ensemble de la période, cette stratégie a été énormément affectée par la crise financière de 2008; à la suite de la crise, la prime est inversée. En ce qui a trait à notre stratégie d'investissement issue des réseaux de neurones, nous sommes en mesure de générer un rendement annuel moyen excédentaire de 6.09% avec une volatilité annuelle moyenne de 15.54% tout en ayant une exposition au marché (« *beta* ») de 0.315. En termes de rapport rendement risque, les réseaux de neurones sont en mesure de surperformer largement le MOM de référence et l'indice du marché avec un ratio de Sharpe annualisé de 0.392 contre 0.029 pour les rendements MOM et 0.33 pour le marché du S&P500.

Notre période d'analyse affecte grandement la performance de la stratégie MOM¹ de référence. De plus, par rapport au facteur de risque usuel (Fama-French et Cahart), les réseaux de neurones génèrent un rendement alpha moyen de 6.26%. Une très grande portion de ces rendements provient du facteur valeur. Ainsi, en exploitant l'apprentissage par expérience, nos réseaux de neurones ont été capables de générer un rendement positif très peu corrélé avec le marché qui s'avère à être expliqué par les rendements du facteur valeur. Ceci s'explique, entre autres, par le fait que le facteur valeur a connu de bonnes performances sur notre période d'analyse. Il est important de reconnaître que notre stratégie d'investissement est grandement inspirée par le MOM mais n'est pas limitée aux interactions pure MOM; nos réseaux de neurones ont le loisir d'apprendre des relations qui vont au-delà des prix antérieurs.

Cette recherche est divisée comme suit : tout d'abord, le chapitre 1 présente une revue de la littérature sur le facteur de risque MOM, qui est le fondement de notre stratégie. Ensuite, le chapitre 2 présente en détail toute la méthodologie exploitée dans la construction de notre stratégie d'investissement. Finalement, le chapitre 3 présente des résultats quant à la performance de la stratégie d'investissement.

¹ Kent et Moskowitz (2016) identifient un ratio de Sharpe annualisé de 0.60 pour les rendements MOM sur l'ensemble de firmes du NYSE, Amex et Nasdaq entre 1927 et 2013. Barroso et Santa Clara (2015) rapportent un ratio de Sharpe annualisé de 0.53 pour les rendements MOM sur l'ensemble des firmes du NYSE, Amex et Nasdaq entre 1926 et 2011.

Chapitre 1

Revue de littérature

1.1 Le facteur de risque « momentum ».

Les premiers à avoir mis de l'avant la présence du facteur de risque MOM sont Jegadeesh et Titman. En 1993, ils présentent une étude approfondie de la profitabilité de ce facteur sur le marché des actions américaines du NYSE entre 1965 et 1989. Leur stratégie d'investissement vise à acheter et à vendre mensuellement un portefeuille d'actions qui a respectivement surperformé et sous-performé au cours des derniers mois. En faisant varier la période de formation et de détention entre 3 et 12 mois, ils sont en mesure de générer des rendements mensuels positifs significatifs. La stratégie d'investissement la plus profitable est obtenue avec une période de formation et de détention respective de 12 et 3 mois, ce qui permet d'obtenir un rendement mensuel moyen de 1.31%. Ils observent cependant qu'avec une période de détention supérieure à un an, la stratégie produit des rendements négatifs; les auteurs font référence à ce phénomène comme étant un renversement des rendements. Selon eux, ce phénomène est causé par un retour aux valeurs fondamentales des titres boursiers. Les rendements issus du facteur MOM seraient en grande partie attribuables aux comportements des investisseurs qui contribuent à éloigner le prix des titres boursiers de leurs valeurs fondamentales. En 1993, aucune personne n'avait encore documenté la rentabilité de cette stratégie d'investissement. Au contraire, DeBondt et Thaler (1985) rapportent des résultats opposés; les actions des firmes ayant sous-performées (surperformées) leurs pairs, dans les mois récents, vont en moyenne surperformées (sous-performées) dans les mois à venir. Toutefois, plusieurs autres recherches ont démontré plus tard que la prime MOM est bel et bien présente et génère de très bons rendements. Les résultats trouvés sur le facteur de risque MOM ont ouvert les vannes à plusieurs autres questions de recherches, dont celle de l'efficience des marchés. En effet, plusieurs se sont intéressés aux explications possibles de la rentabilité de cette stratégie, d'autres sur l'étendue des effets du facteur et certains ont proposé des améliorations à la stratégie d'investissement initialement documentée par Jegadeesh et Titman (1993).

1.1.1 Quelques explications au facteur de risque MOM

Fama et French (1996) développent un modèle de tarification d'actifs financiers à multiples facteurs. Le modèle développé ajoute deux autres facteurs de risque au modèle d'évaluation des actifs financiers « MEDAF ». En incluant un facteur pour la taille et un facteur pour la valeur de l'entreprise, il est maintenant possible de tenir compte de la surperformance des actions issue d'entreprises de petite taille et de bonne qualité. Inclure ces deux nouveaux facteurs de risques permet d'expliquer l'effet de renversement à long terme des rendements MOM documentés par Jegadeesh et Titman (1993); les rendements positifs à moyen terme de la stratégie d'investissement restent toujours inexpliqués. Par le fait même, les auteurs ont démontré que les rendements générés par le MOM sur le marché des actions américaines ne peuvent pas être expliqués entièrement par des mesures de covariance et des facteurs de risques. Cahart (1997) introduit un 4^e facteur de risque au modèle de Fama et French (1996) dans son étude sur la performance des fonds d'investissements américains. Ce nouveau « facteur de risque » réplique les rendements d'un portefeuille d'investissement du facteur MOM sur un an. À l'aide de maintenant quatre facteurs, il analyse en détail la rentabilité d'une stratégie consistant à acheter et à vendre respectivement les meilleurs et les pires fonds d'investissements en fonction de leur performance sur les mois passés. Tout comme Jegadeesh et Titman (1993), cette stratégie d'investissement est profitable. De plus, il est en mesure d'attribuer plus de 50% des rendements réalisés au facteur valeur et MOM.

Moskowitz et Grinblatt (1999) font également des recherches sur les sources de profit du facteur MOM sur le marché des actions américaines. Leurs contributions sont multiples. Tout d'abord, ils illustrent que l'effet du facteur en question est également très présent dans différentes industries du marché. Ce résultat est d'autant plus important dans la mesure où les auteurs contrôlent pour le facteur taille, valeur et MOM de chaque action. D'autre part, ils démontrent que les rendements de la stratégie initialement proposée par Jegadeesh et Titman (1993) sont grandement atténués lorsqu'on contrôle pour l'effet MOM propre à chaque industrie; les rendements sont plus faibles, mais demeurent positifs et significatifs. Ainsi, les rendements générés par le facteur MOM des titres boursiers sont grandement attribuables aux performances des industries. Finalement, les auteurs illustrent qu'il est possible de générer un rendement positif et significatif sur la base des performances antérieures des industries. En effet, le portefeuille

composé d'actions provenant d'industries ayant connues les meilleures performances surperforme le portefeuille composé d'actions provenant d'industries ayant connues les pires performances. Tout comme Jegadeesh et Titman (1993), cette stratégie présente également un renversement des rendements à long terme. Toutefois, sa profitabilité est plus importante à très court terme. Moskowitz et Grinblatt attribuent la profitabilité de leur stratégie à un effet de diversification qui n'est pas nécessairement observable dans la recherche Jegadeesh et Titman (1993); agréger les actions de différentes industries permet de réduire le risque spécifique des firmes d'une même industrie.

Jegadeesh et Titman (2001) étendent leur recherche effectuée en 1993 pour la période de 1990 à 1998. Leur étude permet notamment de confirmer leurs résultats initialement rapportés : la stratégie d'investissement demeure profitable sur le marché des actions américaines entre 1990 et 1998. D'autre part, ils présentent des preuves empiriques en faveur des modèles comportementaux suggérant que les profits du facteur MOM sont causés par une réaction décalée des investisseurs aux informations du marché. Bien qu'ils ne puissent pas expliquer entièrement toutes les relations complexes de la profitabilité de leur stratégie d'investissement, les modèles comportementaux semblent apporter le plus de pouvoir explicatif aux anomalies entourant le phénomène.

Cooper et Gutierrez (2004) étudient la relation entre l'état de l'économie et la profitabilité d'une stratégie d'investissement MOM sur le marché des actions américaines. Ils analysent explicitement l'impact de l'état du marché, récession ou expansion, sur la rentabilité à court terme et le renversement à long terme des rendements documentés par Jegadeesh et Titman (1993). De plus, les auteurs sont en mesure de démontrer que la profitabilité d'une telle stratégie est conditionnelle à l'état de l'économie. Sur le marché des actions américaines, une stratégie d'investissement basée sur ce facteur de risque est uniquement profitable à court terme à la suite d'une économie en expansion; la période de formation des portefeuilles doit être une période d'expansion. Être en mesure d'identifier l'état de l'économie dans lequel nous évoluons est indispensable à la profitabilité de la stratégie d'investissement. Ce résultat est étroitement relié aux travaux de Chordia et Shivakumar (2002) qui documentent la propension de plusieurs facteurs macroéconomique à expliquer les rendements du facteur de risque MOM. Aussi, l'effet de renversement des rendements du facteur MOM documentés par Jegadeesh et Titman (1993) est observable, peu importe l'état de l'économie de la période de formation. Ainsi, l'état de l'économie a un impact significatif sur les caractéristiques anormales du facteur MOM.

Être en mesure de bien comprendre les relations complexes des rendements du facteur MOM sur le marché des actions n'est pas une tâche facile. Les modèles factoriels, l'état de l'économie et la performance des industries ne permettent d'expliquer qu'une portion des rendements anormaux du facteur MOM.

1.1.2 Un facteur de risque omniprésent

Les rendements d'investissements positifs associés au facteur MOM sont loin d'être limités aux titres boursiers. Asness, Liew et Stevens (1997) sont parmi les premiers à avoir documenté l'effet du facteur MOM à l'extérieur du marché des actions américaines. Dans une analyse transversale des rendements de différents pays, ils sont en mesure de conclure que ce facteur est non seulement présent à l'international, mais qu'il possède une prime d'investissement comparable à celle initialement documentée par Jegadeesh et Titman (1993). Essentiellement, les auteurs reprennent la méthodologie utilisée par Jegadeesh et Titman (1993) en l'appliquant sur des indices de pays MSCI. Le facteur de risque MOM-pays possède également des propriétés comparables à celles initialement documentées par Jegadeesh et Titman (1993); le renversement des rendements est observable après 3 ans.

Rouwenhorst (1998) démontre quant à lui que l'effet du facteur MOM sur le marché des actions américaines est également observable sur d'autres marchés d'actions internationaux. L'étude est réalisée sur un énorme échantillon de firmes provenant de 12 pays européens entre 1980 et 1995. L'auteur démontre qu'un portefeuille d'actions diversifiées composé des plus récents gagnants surperforme d'environ 1% mensuellement un portefeuille d'actions diversifiées composé des plus récents perdants. Tel que documenté sur le marché américain, ce rendement mensuel persiste jusqu'à un an après la formation initiale des portefeuilles d'investissement et ne peut être attribué aux facteurs de risque usuels. En effet, une fois contrôlé pour le risque du marché et la taille des entreprises, le rendement anormal mensuel rapporté par cette stratégie augmente. Cette étude confirme que les résultats de Jegadeesh et Titman (1993) ne sont ni une coïncidence ni un phénomène exclusif au marché des actions américaines.

Le marché des devises est un marché très attrayant pour les investisseurs sophistiqués qui savent comment exploiter de plus en plus des indicateurs techniques pour générer un profit. Okunev et White (2003) font une étude empirique de la profitabilité de plusieurs stratégies

d'investissement, dont une s'intéressant spécifiquement au facteur MOM sur les devises étrangères. Leur étude se concentre sur huit devises de pays développés sur la période de 1970 à 2000. À l'aide de plusieurs combinaisons de moyenne mobile à court et à long terme, ils font un classement des bonnes et des mauvaises devises. À l'aide de leur classement et de la méthodologie de Jegadeesh et Titman (1993), ils sont en mesure de générer un rendement mensuel de 6% qui n'est pas sensible à l'intervalle de temps et aux devises considérées. Cette analyse empirique démontre qu'introduire une composante MOM à une stratégie d'investissement sur les devises peut générer des rendements significatifs positifs. Menkhoff et al. (2012) s'intéressent également à la profitabilité de cette stratégie sur le marché des devises. Les auteurs approfondissent davantage les travaux de Okunev et White (2003) en tenant compte d'un plus grand nombre de devises et en élargissant le spectre des sources d'explication potentielle des rendements de la stratégie. Tels que documentés par Jegadeesh et Titman (1993 et 2001), les rendements présentent un renversement à long terme. Toutefois, sur le marché des devises, cette stratégie est davantage profitable sur un très court horizon. En se fiant aux performances des devises du dernier mois et en investissant sur un horizon futur d'un mois, ils sont en mesure de générer un rendement annuel significatif de 10%. Ici, les frais de transactions permettent d'expliquer une très grande portion des rendements de la stratégie; ajuster la stratégie pour inclure le prix de vente et d'achat diminue significativement les rendements du facteur MOM. Cependant, même après avoir ajusté pour les frais de transactions, la stratégie demeure profitable. De plus, les rendements de la stratégie sur les devises ne sont pas corrélés significativement avec le cycle des affaires, la liquidité, le facteur du « carry trade » et les facteurs de risques usuels.

Tout comme le marché des devises, celui des commodités présente un excellent véhicule d'investissement pour les investisseurs sophistiqués. Dans une étude des caractéristiques fondamentales de 31 commodités entre 1971 et 2010, Gorton, Hayashi, Rouwenhorst (2012) illustrent que la stratégie MOM est également un outil d'investissement puissant permettant de générer des rendements annuels moyen de plus de 11%. Dans leur étude, ils examinent un modèle de prédictions de prix de commodités. Les signaux tels que la base (différence entre le contrat à terme et son sous-jacent), les rendements historiques du contrat futur et les rendements historiques des actifs sous-jacents sont empiriquement reliés au niveau d'inventaire de la commodité. En ce sens, suivre les signaux associés au MOM correspond à acheter et à vendre respectivement des commodités qui ont un niveau d'inventaire plus faible et plus élevé que leur moyenne historique.

Sur le marché des commodités, le MOM semble être un bon intermédiaire pour profiter davantage des niveaux d'inventaires des commodités.

Gutierrez et Kelley (2008) illustrent quant à eux que les rendements significatifs positifs du facteur MOM présents sur le marché des actions américaines ne sont pas seulement contraints à l'investissement sur un horizon à moyen terme. Les auteurs démontrent qu'une stratégie visant à acheter et à vendre respectivement les portefeuilles d'actions ayant connu la plus grande et la plus faible croissance sur la dernière semaine est en mesure de générer des profits significatifs sur les 52 semaines à suivre. Ce résultat est d'autant plus surprenant étant donné le fort renversement que subissent les titres financiers à la suite d'un rendement hebdomadaire extrême. La stratégie est en mesure de largement compenser la première semaine de pertes. Il est également rapporté que l'effet du facteur MOM hebdomadaire est un phénomène dominant dans la stratégie initialement rapportée par Jegadeesh et Titman (1993). Dans ce sens, ils concluent que le MOM long terme est un phénomène omniprésent et continu. De plus, les rendements de leur stratégie d'investissement hebdomadaire ne sont pas sensibles aux types de nouvelles encourues sur le marché. Plus précisément, les rendements des actions suivants des nouvelles dites explicites (mouvement du prix de l'action suivant une nouvelle publique) et implicites (mouvement de prix de l'action sans la présence de nouvelle publique) se comportent de la même manière. Dans les deux cas, leur stratégie d'investissement hebdomadaire subit un renversement à très court terme, mais également à long terme telle que rapportée par Jegadeesh et Titman (1993). Ce résultat relance le débat quant aux théories comportementales qui visent à attribuer le facteur MOM à des réactions des investisseurs aux nouvelles du marché.

Par ailleurs, Asness, Moskowitz et Pedersen (2013) illustrent l'envergure du facteur MOM. Dans un énorme travail de synthèse, les auteurs analysent le comportement et les caractéristiques du facteur MOM et du facteur « valeur » dans 8 marchés différents et à travers plusieurs classes d'actifs. Leur objectif était de déterminer comment réagissent ces facteurs de risques dans une même classe d'actif, entre classe d'actif, dans un même marché et entre différents marchés. Tout d'abord, ils déterminent que les facteurs MOM et « valeur » sont présents sur tous les marchés et classes d'actifs considérées, ce qui inclut la présence du facteur de risque MOM sur le marché des obligations et un facteur de risque « valeur » sur les taux de change et les commodités. Ensuite, ils documentent la présence d'une corrélation significative entre ces deux facteurs de risques à travers les différentes classes d'actifs. Les rendements issus du facteur « valeur » provenant de marché

non corrélé entre eux sont positivement corrélés. La même chose est rapportée sur les rendements du facteur de risque MOM provenant de deux marchés non corrélés. Cependant, les deux facteurs de risques sont négativement corrélés entre eux dans chaque classe d'actifs et entre les classes d'actifs. La dynamique de toutes ces corrélations suggère la présence d'une structure de risque commune aux deux facteurs de risques. Le cycle économique, la consommation, le risque de défaut et quelques facteurs macroéconomiques sont en mesure d'expliquer partiellement cette structure de risque commune. La plus grande partie de la structure de risque est attribuable aux facteurs de liquidité et de financement; le risque de liquidité est négativement corrélé au facteur « valeur » et positivement corrélé facteur MOM. De plus, le risque de financement est positivement corrélé avec le facteur MOM sur tous les marchés et les classes d'actifs. L'analyse conjointe de ces facteurs de risques révèle une dynamique complexe entre deux facteurs de risques très différents.

En somme, le facteur de risque MOM initialement documenté par Jegadeesh et Titman (1993) est loin d'être dû au hasard. Leurs conclusions sont omniprésentes, observables à travers plusieurs véhicules d'investissement, différents marchés et sur plusieurs périodes d'investissement.

1.1.3 Une stratégie sujette à de lourdes pertes

Depuis sa première publication en 1993, plusieurs recherches ont été menées dans le but de comprendre, mais également d'améliorer le profil rendement risque du facteur de risque MOM. L'article de Charles Lee et Swaminathan (2000) analyse la relation entre le prix et le volume transigé d'une action et l'impact de ce dernier sur les rendements de la stratégie issue du facteur MOM. Plus précisément, les auteurs s'intéressent à l'utilité du volume transigé passé d'une action dans la prédiction de ses rendements futurs à moyen et à long terme. Ils ont montré que les firmes avec un plus faible volume transigé surperforment les firmes avec un volume transigé plus élevé. De plus, acheter un portefeuille d'action à faible volume ayant bien performé au cours des derniers mois et vendre un portefeuille d'action à fort volume ayant mal performé dans les derniers mois génèrent un minimum de 4% supplémentaire aux rendements issus du facteur MOM classique. Le rendement de cette stratégie est non seulement significatif, mais ne peut pas être expliqué par les facteurs de risques usuels de Fama-French (1996). Les auteurs illustrent également que les rendements significatifs générés ne sont pas attribuables à une prime de liquidité associée au

volume transigé des actions. Au contraire, les rendements d'une telle stratégie sont attribuables à ce qu'ils nomment le « *Momentum life cycle* » (MLC). Cette hypothèse soutient que les firmes passent à travers plusieurs périodes d'attraction et de négligence qui sont fonction du rendement et du volume transigé historique. Ainsi, à l'aide d'un filtre supplémentaire appliqué sur les portefeuilles d'actions associés au facteur MOM, ils sont en mesure de dégager un rendement supplémentaire significatif.

Grundy et Martin (2001) analysent également une stratégie alternative à celle présentée par Jegadeesh et Titman (1993). L'essence de leur recherche provient de la dynamique variable de l'exposition au marché (« *beta* ») du facteur MOM. Suivant un marché en expansion, les rendements du facteur MOM sont positivement corrélés avec le marché, tandis qu'à la suite d'un marché baissier, les rendements du facteur MOM sont négativement corrélés au marché. Ils tentent d'exploiter cette dynamique variable de l'exposition au marché (« *beta* ») pour générer des rendements supplémentaires au facteur de risque MOM. Pour se faire, ils analysent le comportement du facteur en question face aux trois autres facteurs de risques de Fama-French (1996). Plutôt que de faire un classement des titres boursiers uniquement sur la base de leurs rendements historiques, ils construisent des portefeuilles d'actions en fonction du rendement excédentaire inexplicable par les facteurs de risques usuels. D'une part, cette stratégie est en mesure de générer des rendements plus profitables que celui du facteur MOM classique. D'autre part, ils observent que les portefeuilles d'actions gagnantes (perdantes) selon leur critère d'investissement sont majoritairement les mêmes que les gagnantes (perdantes) formés sur la base du rendement historique; les deux groupes se chevauchent. Il n'est donc pas surprenant de constater que faire un classement des actions sur la base du rendement spécifique de chaque firme est une stratégie alternative et profitable.

Barroso et Santa Clara (2015) proposent quant à eux une manière intéressante de gérer le risque des grandes pertes inhérentes associé au facteur MOM. Parmi les facteurs de risques traditionnels, celui du MOM offre le plus grand rapport rendement risque. Toutefois, ce facteur de risques est également accompagné de très grandes pertes potentielles. Telle qu'établie par Grundy et Martin (2001), la corrélation variable au marché de ce facteur de risque est à l'origine de ses gains, mais également de ses lourdes pertes. Ici, les auteurs proposent une méthode efficace pour gérer le risque des grandes pertes. Tout d'abord, ils suggèrent de standardiser les rendements des actions par leur volatilité historique avant de faire le classement des portefeuilles d'actions

gagnantes et perdantes. De cette façon, les rendements du facteur MOM ne sont plus dominés par des actions ayant connu une très forte volatilité. D'autre part, ils proposent une manière de couvrir l'exposition au marché (« *beta* ») du facteur de risque MOM. L'idée est de dynamiquement couvrir la stratégie d'investissement MOM en fonction de sa volatilité anticipée. Ces deux alternatives permettent d'éliminer pratiquement toutes les grandes pertes que subit le facteur de risque MOM classique et de doubler le rapport rendement risque de la stratégie.

Kent et Moskowitz (2016) analysent également le comportement du facteur de risque MOM dans différents états de l'économie. Leur travail rejoint de près celui des deux articles précédents. Historiquement, ce facteur de risque génère des rendements moyens positifs. Toutefois, en période de crise ou de turbulence, les portefeuilles d'actions ayant le moins bien performés dans le passé vont surperformés ceux qui ont bien performés dans le passé; la prime de risque devient négative. Pour être en mesure de mitiger ces lourdes pertes en période de haute volatilité, ils proposent un modèle de prédiction de l'état futur de l'économie. Ainsi, les poids des portefeuilles d'actions gagnantes et perdantes sont ajustés en fonction de l'état économique anticipé. Tout comme Barroso et Santa Clara (2015), ils sont en mesure de pratiquement doubler le rapport rendement risque du facteur MOM. De plus, les résultats qu'ils rapportent ne sont pas limités uniquement au marché des actions américaines; la stratégie est également profitable sur 8 différents marchés et plusieurs classes d'actifs.

Moskowitz, Ooi, Perdersen (2012) étudie la profitabilité d'une stratégie d'investissement qui vise à exploiter le facteur MOM sur l'actif lui-même; on parle ici de « *Time series momentum* » (TSM). Dans leur recherche, ils analysent 58 actifs liquides: des titres boursiers, des obligations et des contrats d'achat sur devise et des commodités. Plutôt que de comparer la performance de chaque classe d'actifs entre elles ou même comparer les actifs d'une même classe entre eux, chaque actif est comparé avec sa performance antérieure; la décision d'achat ou de vente d'un actif est prise indépendamment des autres. Pour se faire, les rendements historiques de tous les titres sont étudiés en relation avec leur propre volatilité historique. La standardisation des rendements historiques de chacun des titres permet entre autres d'avoir un point de référence comparable entre les actifs issus de différentes classes. Cette méthodologie leur permet de dégager des rendements positifs et significatifs non explicables par les facteurs de risques usuels. La profitabilité de cette stratégie est entre autres attribuable à la très grande diversification dont elle bénéficie; le fait

d'inclure 58 classes d'actifs avec des caractéristiques et des profils de risque différents réduit grandement le risque de la stratégie d'investissement.

La stratégie d'investissement initialement proposée par Jegadeesh et Titman (1993) est loin d'être sans risque. La plupart des améliorations ou des alternatives proposées dans la littérature permet de mieux contrôler pour le risque inhérent de la stratégie et offre un meilleur rapport rendement risque.

1.2 Algorithmes d'apprentissages et la finance de marché

L'utilisation de régressions linéaires est largement utilisée dans plusieurs domaines de la finance. Toutefois, cet outil implique plusieurs hypothèses qui peuvent être contraignantes. Les réseaux de neurones permettent notamment de contourner quelques limitations des régressions linéaires telles que la multicollinéarité dans les variables explicatives. De plus, ils permettent d'introduire une non-linéarité dans les relations étudiées et une analyse d'apprentissage par expérience. Avec son gain en popularité et ses résultats non équivoques, l'utilisation des réseaux de neurones est de plus en plus exploitée dans plusieurs sphères de recherche.

En finance, nous pouvons penser à Sirignano, Sadhwani et Giesecke (2018) qui développent un modèle d'apprentissage sophistiqué afin d'étudier la relation non linéaire entre les prêts hypothécaires et le risque de paiements des emprunteurs. La majorité des recherches sur le sujet se base sur des modèles linéaires pour faire l'analyse comportementale des emprunteurs. Le résultat le plus important dans leur recherche est justement de démontrer que plusieurs variables comportementales décrivant l'emprunteur et l'économie dans son ensemble ne dépendent pas linéairement du risque de paiement du prêt hypothécaire. Ce résultat est également observable au sein des variables explicatives; les interactions entre les variables explicatives sont également fortement non linéaires. Ici, les auteurs utilisent des réseaux de neurones pour exploiter les interactions complexes entre le taux de chômage, le ratio prêt-valeur, le taux hypothécaire et l'appréciation du prix de la propriété dans ses prévisions. Leur recherche illustre bien la valeur ajoutée de l'utilisation d'algorithmes d'apprentissages dans la résolution de problèmes complexes.

Gu, Kelly et Xiu (2018) font une analyse exhaustive de plusieurs méthodes quantitatives pour faire de la tarification d'actifs financiers. Les auteurs analysent entre autres la performance des modèles de régressions linéaires simples et généraux, des modèles de régressions à réduction

de dimensions, les arbres de décisions et les réseaux de neurones. Pour mesurer le pouvoir prédictif de chaque modèle, ils présentent la profitabilité d'une stratégie d'investissement issue de chaque méthode appliquée sur le marché des actions du NYSE, AMEX et NASDAQ entre 1957 et 2016. La stratégie vise à acheter un portefeuille d'actions ayant les plus grands rendements prédits et à vendre un portefeuille d'actions avec les plus faibles rendements prédits; la méthodologie est comparable à celle de Jegadeesh et Titman (1993). La tarification d'actifs financiers repose grandement sur la propension à bien prédire le rendement futur d'un titre à l'aide d'un ensemble de variables. Tel que démontré par les auteurs, ce type de problématiques peut grandement bénéficier de l'utilisation d'algorithmes d'apprentissages sophistiqués. En effet, ils sont en mesure d'illustrer que les réseaux de neurones et les arbres de décisions ont un meilleur pouvoir prédictif sur le rendement futur des titres financiers; les réseaux de neurones jumelés à la méthodologie de Jegadeesh et Titman (1993) génèrent un rendement mensuel moyen de plus de 3%. Les auteurs illustrent également que le pouvoir prédictif des modèles complexes non linéaire est majoritairement attribuable à des variables de tendance, d'élan et de renversement de rendement.

De leur côté, Benoit et Raffinot (2018) analysent les différents cycles économiques observés au cours des 30 dernières années sur le marché américain et la zone européenne. Plus précisément, ils s'intéressent à la prédiction du changement de l'état économique en exploitant des arbres de décisions et de techniques de « *boosting* ». Les chercheurs sont en mesure d'utiliser un énorme éventail de variables financières et macroéconomiques pour bien détecter le moment précis du changement de l'état de l'économie. La capacité des algorithmes quantitatifs sophistiqués à pouvoir exploiter une panoplie de variables explicatives sans compromettre le modèle est ce qui rend l'outil très puissant. Leur modèle est en mesure de détecter rapidement et précisément les cycles de croissance sur le marché américain et la zone européenne, ce qui leur permet d'exploiter précisément le bon moment d'entrée et de sortie du marché. Ainsi, les auteurs sont en mesure de dégager un rendement moyen annuel de plus de 10% sur le marché des actions américaines en exploitant les arbres de décisions et des techniques de « *boosting* ».

En somme, l'utilisation d'algorithmes d'apprentissages sophistiqués gagne à être utilisée davantage dans le domaine de la finance. La capacité d'introduire de la non-linéarité et à exploiter un énorme éventail de données pour résoudre une problématique est définitivement un atout important.

Chapitre 2 Méthodologie

2.1 Données

Afin d'avoir le plus grand nombre de données possible, nous avons utilisé une fréquence d'observations quotidiennes sur l'ensemble des firmes ayant fait partie de l'indice du S&P500 entre juillet 1998 et juin 2018. Le choix de faire l'analyse sur cette période est en grande partie justifié par l'accessibilité des données sur les variables explicatives exploitées; plusieurs indices ont des données qui ne remontent qu'au début des années 1990. Le nombre de firmes à travers cet échantillon est évalué à 1939. Les données relatives aux prix et aux volumes transigés de chaque firme sont fournies par *Center for Research in Security Prices (CRSP)*. Tels que mentionnés à la section précédente, plusieurs autres facteurs comme les coefficients d'expositions (« *beta* ») aux différents facteurs de risques usuels, l'état de l'économie et des indicateurs de tendance influencent la profitabilité et le pouvoir prédictif des rendements du facteur MOM. Nous présentons en annexe une liste exhaustive de l'ensemble des variables utilisées pour l'apprentissage des réseaux de neurones. Pour introduire une certaine temporalité dans les données fournies aux réseaux de neurones, nous prenons également plusieurs variables sur des fenêtres mobiles de 1, 3, 6 et 12 mois et en niveau d'accélération, soit le changement observé du niveau de la variable sur 1, 3 et 6 mois.

2.2 Indice de référence : Jegadeesh et Titman (1993)

L'analyse de la performance de notre stratégie d'investissement est comparée avec la stratégie d'investissement de Jegadeesh et Titman (1993) dans notre univers d'investissement, soit le S&P500. Pour former les portefeuilles de référence, chaque mois, nous effectuons tout d'abord un classement des actions en fonction de leur rendement cumulé réalisé au cours des 12 derniers mois. Pour être conformes à la littérature de Jegadeesh et Titman (1993), Asness (1994) et Fama et French (1996), nous excluons le dernier mois dans le calcul du rendement cumulé. Ceci permet d'éviter le renversement des rendements démontrés par Jegadeesh (1990) et Lehmann (1990). Nous imposons également que les firmes présentent un minimum de 3 ans de rendements historiques pour s'assurer que l'univers d'actions de notre indice de référence soit le même que

celui employé dans l'élaboration de notre stratégie d'investissement. En fonction du classement des titres du S&P500, l'indice de référence pour le facteur MOM est déterminé en prenant une position longue et courte respectivement sur les déciles des meilleures et des pires firmes. Les poids des titres en portefeuilles sont tous équipondérés pour former un portefeuille à coût nul (ou autofinancé); les positions à découvert financent les positions d'achat. Ce portefeuille est formé sur une base mensuelle.

2.3 Réseaux de neurones artificiels

Plusieurs méthodes quantitatives d'apprentissage sont exploitées dans le développement de l'intelligence artificielle. Pour notre recherche, le choix de la méthode s'arrête sur les réseaux de neurones artificiels, où *Artificial neural networks* (ANN), qui sont largement utilisés dans plusieurs domaines connexes à la finance.

2.3.1 Une brève description

Les réseaux de neurones artificiels portent leur nom, car ils tentent de répliquer la dynamique du système neuronal humain dans la résolution de problèmes complexes. Il existe plusieurs types de réseaux de neurones. Le réseau de neurones classique ANN est typiquement composé de quelques couches successives, sur lesquelles se trouvent plusieurs neurones. La figure 1 illustre la dynamique typique de ce genre de réseaux. Ici, le ANN possède cinq variables explicatives, une seule couche cachée composée de trois neurones et un seul neurone de sortie. Pour un ANN typique, chaque couche du système prend comme information les sorties de la couche précédente. La seule exception est pour la toute première couche qui prend son information des variables explicatives initialement fournies au réseau.

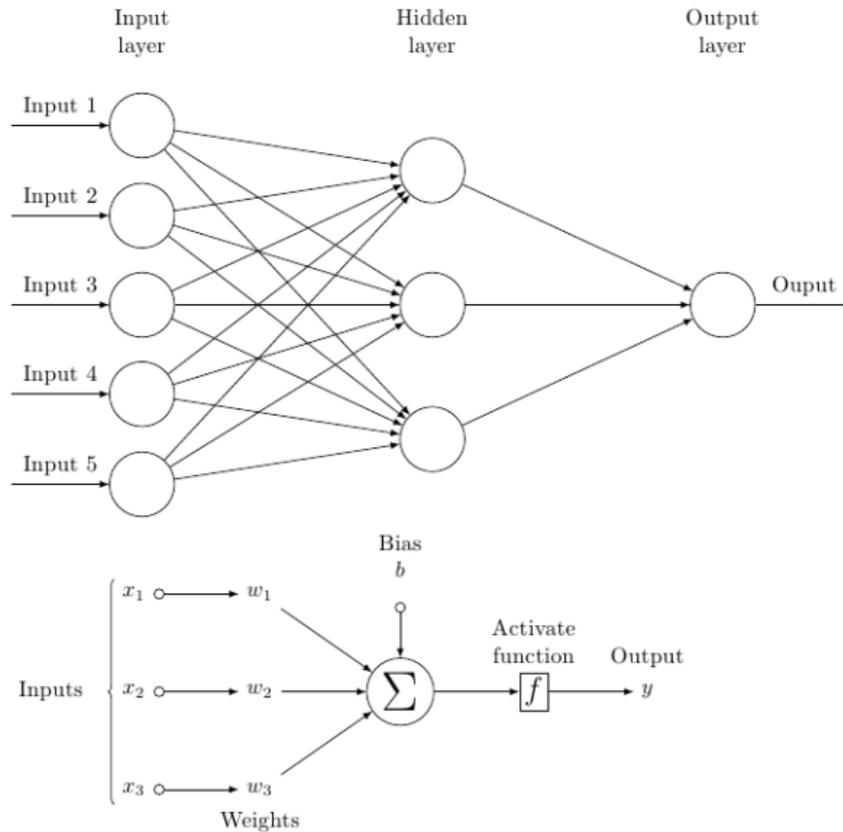


Figure 1: Dynamique d'un réseau de neurones ANN classique ne possédant qu'une seule couche cachée et un neurone de sortie. Les cercles sur la première couche du réseau Input layer correspondent aux variables explicatives. Les cercles sur les couches subséquentes représentent l'information transformée et transmise aux neurones de la couche suivante. Le graphique du bas illustre en détail le processus de propagation et de transformation de l'information de chaque neurone.

La première couche du système neuronal est appelée la couche des variables, ou « *Inputs layer* ». Toutes les couches subséquentes sont dites cachées, ou « *Hidden layer* ». Pour une couche i , l'information contenue dans chaque neurone est pondérée par une matrice de poids aléatoire, additionnée, puis ajoutée à une constante (ou biais) également aléatoire. Ce nouveau signal passe ensuite à travers une fonction d'activation non linéaire qui permet de déterminer si le neurone de la prochaine couche sera activé. Cet exercice est répété pour chaque connexion allant vers un neurone de la couche suivante. Ce processus de propagation de l'information fait partie des raisons pour lesquelles un réseau de neurones est puissant dans la résolution de problèmes: il permet d'introduire une non-linéarité. Une régression linéaire à multiples variables peut être représentée par un réseau de neurones sans couche cachée et un seul neurone de sortie. Une des propriétés fondamentales des réseaux de neurones est sa capacité d'apprentissage par expérience. Lorsque l'information initialement contenue dans les variables explicatives se rend jusqu'à la dernière

couche du système, celle-ci est comparée à la variable à modéliser. À partir de ce point, le réseau modifie les poids de chaque connexion pour tenter de répliquer la variable à modéliser. Le ANN complète ainsi une époque. À chaque époque, l'objectif est de trouver les poids optimaux à accorder à chaque connexion du système pour minimiser la fonction objective du réseau. Cette étape d'optimisation porte le nom de rétropropagation.

En somme, un réseau de neurones transforme les variables explicatives de la première couche en appliquant une série de multiplications linéaires envoyée vers des fonctions d'activation non linéaires avant de rejoindre le(les) neurone(s) de sortie. À partir de ce point, la rétropropagation du système ajuste les poids du réseau pour optimiser sa fonction objective.

2.3.2 *Les paramètres des réseaux de neurones : les hyperparamètres*

L'objectif ultime d'un algorithme d'apprentissage A est de trouver une fonction F qui minimise une fonction de coût quelconque $L(\mathbf{x}; F)$ à l'aide d'un sous-échantillon de données \mathbf{x} provenant d'une distribution $G(\mathbf{x})$ inconnue. En fait, l'algorithme d'apprentissage A est une fonction qui tente d'associer un sous-ensemble de données \mathbf{x}^{train} provenant de $G(\mathbf{x})$ à une fonction F . L'algorithme d'apprentissage produit F à travers un processus d'optimisation sur un ensemble de paramètres θ . Cependant, l'algorithme d'apprentissage lui-même dépend d'un ensemble de paramètres δ qui doit être spécifié et ajusté. L'ensemble de paramètres θ est souvent spécifié et optimisé à travers la rétropropagation, tandis que l'ensemble de paramètres δ , doit être optimisé par la fonction de coût elle-même. Par exemple, on peut penser aux poids des connexions d'un réseau de neurones comme un paramètre de θ et le nombre de couches d'un réseau de neurones comme un paramètre de δ . L'ensemble des paramètres δ à optimiser est considéré comme un hyperparamètre. En soi, tous les paramètres qui vont influencer l'apprentissage du réseau de neurones sont des hyperparamètres. Idéalement, il faut déterminer la valeur des hyperparamètres qui permettent de minimiser l'erreur globale de la fonction de coût :

$$F = A_{\delta}(x^{train}) \quad (1)$$

$$\delta^* = \underset{\delta \in \Lambda}{\operatorname{argmin}} E_{x \sim G_x} [L(x; A_{\delta}(x^{train}))] \quad (2)$$

La résolution de ce problème consiste à l'optimisation des hyperparamètres. Parmi les hyperparamètres les plus importants, on retrouve ceux de régularisation, ceux d'optimisation et ceux qui appartiennent à la structure globale du réseau.

Les hyperparamètres de régularisation sont nombreux et jouent un rôle-clé dans tous les algorithmes d'apprentissages. De manière générale, entraîner parfaitement un réseau de neurones sur un ensemble de données n'est pas désirable; le modèle tente d'expliquer le bruit et la variance provenant des données plutôt que la distribution à partir de laquelle proviennent les données. Ceci correspond au surentraînement et cause de très grandes erreurs de prédictions. Au contraire, un réseau de neurones simple capable de généraliser un ensemble de données sera en mesure de faire de bien meilleures prédictions sur un tout nouvel ensemble de données; le modèle capture les régularités des données et possède de meilleures chances de pouvoir généraliser l'information sur de nouvelles données. Un excellent moyen d'éviter le surentraînement est d'introduire plus de contraintes sur l'optimisation des poids du réseau de neurones. Thikinov propose la régularisation générale des poids suivante :

$$\min_F \sum_{i=1}^n L(F(x_i), y_i) + \lambda \phi(F) \quad (3)$$

L est la fonction de coût qui minimise la prédiction de $F(x)$ sur y , λ est un paramètre mesurant l'importance de la régularisation et $\phi(F)$ est ajoutée pour augmenter la pénalité sur la complexité de F . Deux variantes très répandues de cette régularisation généralisée sont la régression Lasso et Ridge, ou respectivement la régularisation L1 et L2 :

$$L1 : \min_{\mathbf{w}} (\mathbf{y} - \mathbf{X}\mathbf{w})^T (\mathbf{y} - \mathbf{X}\mathbf{w}) + \lambda \sum_i |w_i| \quad (4)$$

$$L2 : \min_{\mathbf{w}} (\mathbf{y} - \mathbf{X}\mathbf{w})^T (\mathbf{y} - \mathbf{X}\mathbf{w}) + \lambda \sum_i w_i^2 \quad (5)$$

Dans ces formules, \mathbf{y} , \mathbf{x} et \mathbf{w} sont respectivement la variable à expliquer, la variable explicative et l'estimation de la régression linéaire. L2 est le standard de référence en termes de régularisation pour sa solution analytique et sa stabilité. L1 permet quant à elle de rendre le réseau plus parcimonieux. Cette régularisation accorde beaucoup de poids aux facteurs importants et peu aux facteurs non importants, ce qui rend le réseau robuste aux données aberrantes. Un autre excellent moyen d'éviter le surentrainement du réseau de neurones est de faire appel à la désactivation aléatoire de neurones, ou « *dropout rate* » (DR), dans la phase d'entraînement du réseau. Dans leur étude, Srivastava et al. (2014) ont démontré qu'il était possible d'obtenir de meilleures prédictions à l'aide d'un réseau de neurones s'il exploite un DR. L'idée est de limiter l'apprentissage des mouvements conjoints des variables en désactivent aléatoirement un certain pourcentage de neurones à chaque couche.

Dans son apprentissage par expérience, le réseau de neurones ajuste la valeur des poids du système de chaque connexion. À chaque itération (ou époque), le système modifie les poids dans le but de minimiser la fonction objective du réseau. L'ampleur de l'ajustement est déterminée par une fonction d'optimisation. Il est notable que la très grande majorité des fonctions d'optimisation font appel à des algorithmes de gradient stochastiques ou le pas d'ajustement du gradient, « *learning rate* » (LR), qui est un paramètre primordial pour le développement des réseaux ANN. Ainsi, le choix de la fonction d'optimisation et la spécification du pas d'ajustement sont deux hyperparamètres importants pour des algorithmes d'apprentissages sophistiqués.

Une autre classe très importante d'hyperparamètres pour un réseau de neurones est celle qui a trait à la structure même du réseau. Plus précisément, tous les hyperparamètres qui vont dicter comment l'information est intégrée, assimilée et distribuée à travers le réseau ANN. Pour ce qui est de l'intégration et de la propagation de l'information, on peut penser à la forme du réseau en soi. La dynamique d'un réseau rectangulaire n'est pas la même qu'un réseau ou le nombre de neurones par couche diminue géométriquement avec le nombre de couches. Le nombre de couches et de neurones par couche est d'ailleurs très important. On peut également penser aux diverses

fonctions d'activations de chaque neurone, au nombre de variables ou à la quantité de données à fournir au réseau dans sa phase d'apprentissage.

En somme, tous les paramètres qui peuvent influencer la conception d'un réseau de neurones sont des hyperparamètres. La question peut maintenant se poser : comment allons-nous choisir les hyperparamètres de nos réseaux de neurones dans l'élaboration de notre stratégie d'investissement? Est-ce qu'il est préférable de spécifier un pas d'ajustement dans le processus d'optimisation? Est-ce qu'un réseau de neurones à plusieurs couches, étroit avec des fonctions d'activation linéaire est préférable à un réseau peu profond, large avec des fonctions d'activation exponentielle? Est-ce que la structure optimale du réseau change à travers le temps? La réponse se trouve dans l'optimisation séquentielle de modèles.

2.3.3 *Optimisation des hyperparamètres*

Le choix et le réglage des hyperparamètres sont des étapes fastidieuses et longues. Toutefois, il s'agit d'étapes très importantes qui peuvent grandement affecter la performance des réseaux de neurones; il faut être en mesure de déterminer la combinaison d'hyperparamètres qui va minimiser la fonction du coût elle-même. La résolution de ce problème consiste à l'optimisation des hyperparamètres.

Bergstra, Bengio et al (2011) et Bergstra et Bengio (2012) ont évalué la performance de plusieurs méthodes d'optimisation d'hyperparamètres sur plusieurs types de réseaux de neurones et sur diverses bases de données. Dans leurs études, ils comparent la performance de l'optimisation d'hyperparamètres avec un réglage manuel, des grilles de recherche complète et aléatoire et à l'aide d'algorithmes automatisés. Le réglage d'hyperparamètres manuel repose en grande partie sur l'intuition de l'utilisateur et n'offre pas de justification scientifique. Dans la majorité des cas, les utilisateurs vont se fier aux recherches antérieures pour justifier les hyperparamètres de leurs algorithmes d'apprentissages. Ce type d'optimisation d'hyperparamètres laisse place à une grande marge d'erreurs et de tâtonnements avant d'obtenir une combinaison d'hyperparamètres qui semblent bien performer. Dans tous les cas, il ne s'agit pas d'une méthode exhaustive de recherche d'hyperparamètres. Pour sa part, l'optimisation d'hyperparamètres par grille de recherche complète offre une méthode de sélection extrêmement exhaustive. L'utilisateur crée un répertoire contenant tous les hyperparamètres potentiels, puis, l'optimisateur évalue la performance du réseau

de neurones sous chacune d'entre elles. Ce type d'optimisation est extrêmement coûteux en temps de calcul, surtout pour des résolutions de problèmes à très grande dimensionnalité. En effet, pour chaque combinaison d'hyperparamètres, il faut entraîner l'algorithme d'apprentissage sur un ensemble de données, faire des prédictions sur des données de validation puis évaluer la fonction de coût du modèle. Un moyen efficace de contourner le fardeau du temps de calcul est d'évaluer aléatoirement certaines combinaisons d'hyperparamètres. Les chercheurs illustrent que l'optimisation par grille de recherche complète passe beaucoup trop de temps à évaluer la performance des réseaux de neurones configurés à l'aide des « mauvaises » combinaisons d'hyperparamètres. En n'évaluant qu'un sous-ensemble aléatoire d'hyperparamètres, le temps de calcul est grandement réduit tout en permettant d'obtenir de très bonnes estimations sur les hyperparamètres optimaux à prioriser. Finalement, Bergstra et Bengio bonifient la performance de l'optimisation des hyperparamètres en introduisant un critère de recherche Baysien. Plutôt que d'évaluer la fonction objective du réseau de neurones en diverses combinaisons d'hyperparamètres, l'optimisation séquentielle par modèle, ou « *Sequential model based optimisation* » (SMBO), permet l'analyse des combinaisons des hyperparamètres pertinents. Contrairement à la recherche par grille, aléatoire ou complète, l'optimisation SMBO détermine les prochains hyperparamètres à évaluer en fonction de ceux précédemment choisis. Ceci est réalisé à l'aide d'un modèle probabiliste \mathbf{P} et d'un critère de sélection d'hyperparamètre \mathbf{C} . Le modèle probabiliste \mathbf{P} est un substitut à la fonction de coût du réseau de neurones qui est beaucoup plus simple à optimiser. L'idée est de trouver la combinaison d'hyperparamètres qui performe le mieux sur la densité \mathbf{P} , pour ensuite les appliquer sur la fonction de coût de l'algorithme d'apprentissage. De son côté, le critère de sélection d'hyperparamètre \mathbf{C} guide le choix des hyperparamètres à utiliser dans les évaluations du modèle \mathbf{P} . Il existe plusieurs fonctions probabilistes et plusieurs critères de sélection d'hyperparamètres qui peuvent modéliser la fonction de coût du réseau de neurones. Les auteurs proposent d'utiliser conjointement le modèle probabiliste « Tree-structured Parzen Estimator », ou TPE, et le critère de sélection « *Expected Improvement* » (EI).

$$EI_{l^*}(h) = \int_{-\infty}^{l^*} \max(l^* - l)P(l|h)dl \quad (6)$$

Où l^* est le seuil optimal de la fonction de coût de l'algorithme d'apprentissage, h est l'ensemble d'hyperparamètres évalués, l est la valeur de la fonction de coût de l'algorithme d'apprentissage sous l'ensemble d'hyperparamètres proposés et $P(l|h)$ est la fonction substitut de la fonction de coût de l'algorithme d'apprentissage qui exprime la probabilité d'observer l sachant h . L'approche TPE modélise \mathbf{P} à l'aide du théorème de Bayes, où $P(h|l)$ représente la probabilité d'observer l'ensemble d'hyperparamètres h sachant la valeur de la fonction de coût l .

$$EI_{y^*}(x) = \int_{-\infty}^{y^*} \max(y^* - y)p(y|x)dy$$

$$P(l|h) = \frac{P(l \cap h)}{P(h)} = \frac{P(h|l) * P(l)}{P(h)} \quad (7)$$

$$p(h|l) = \begin{cases} q(h) & \text{si } l < l^* \\ g(h) & \text{si } l \geq l^* \end{cases} \quad (8)$$

L'idée est de construire deux distributions distinctes pour l'ensemble d'hyperparamètres, soit $q(h)$ et $g(h)$. La fonction de densité $q(h)$ est formée en utilisant les hyperparamètres $\{h^{(l)}\}$ pour lesquelles la fonction de coût du réseau de neurones est plus faible que le seuil optimal l^* . De son côté, $g(h)$ est formé à l'aide des hyperparamètres restants. Le TPE détermine le seuil optimal l^* comme étant un quantile γ des valeurs observées de la fonction de coût; aucune distribution n'a besoin d'être spécifiée pour $P(l)$.

$$\gamma = p(l < l^*), \quad p(h) = \int P(h|l) * p(l)dl = \gamma q(h) + (1 - \gamma)g(h) \quad (9)$$

$$EI^*_{l^*}(h) \propto \left(\gamma + \frac{g(h)}{q(h)}(1 - \gamma) \right)^{-1} \quad (10)$$

L'équation (10) illustre que pour maximiser le critère EI, l'optimisateur doit trouver les valeurs des hyperparamètres qui ont une plus grande probabilité sous $q(h)$ et une plus faible probabilité sous $g(h)$; maximiser le critère de sélection d'hyperparamètres correspond à maximiser le ratio $q(h)/g(h)$. En pratique, l'approche TPE fait une pige d'hyperparamètres sous $q(h)$ et retourne l'ensemble d'hyperparamètres produisant le plus grand rapport $q(h)/g(h)$ qui correspond au plus

grand EI. Ces hyperparamètres sont ensuite évalués sur la fonction de coût du réseau de neurones. Si la fonction probabiliste \mathbf{P} est un bon substitut à la fonction de coût, ces hyperparamètres devraient améliorer la performance des réseaux de neurones.

Pour ce travail de recherche, nous allons exploiter l'optimisation d'hyperparamètres proposée par Bengio et Bergstra, soit le TPE jumelé au EI. Cette approche est non seulement puissante et rapide, mais permet de poser moins d'hypothèses restrictives quant au choix des hyperparamètres de nos réseaux de neurones.

2.4 Stratégie d'investissement

Dans cette section, nous présentons en détail les étapes du processus de la stratégie d'investissement. Pour se faire, il est important de décrire explicitement comment les réseaux de neurones font leurs prédictions à chaque mois.

2.4.1 Univers des hyperparamètres

Le plus grand avantage de nos réseaux de neurones vient de l'absence d'hypothèse que nous émettons quant à leur structure. En effet, nous déterminons uniquement l'univers possible des hyperparamètres disponibles. Le choix des hyperparamètres est dicté par le critère EI et l'approche TPE qui évalue itérativement quelle combinaison minimise la fonction de coût du réseau. Ici, la fonction de coût à minimiser est la somme des erreurs au carré. Par opposition à la classification, nous utilisons des réseaux de neurones en régression.

Le premier hyperparamètre à définir est le nombre de couches cachées. Plusieurs recherches, telles que Gu, Kelly et Xiu (2018), démontrent que la performance de leurs réseaux de neurones est optimale lorsque ceux-ci sont moins profonds; des réseaux de neurones ayant 3 à 5 couches de neurones semblent avoir la meilleure performance. Dans notre analyse, le nombre de couches des systèmes neuronaux est défini entre 1 et 5. Un deuxième hyperparamètre touchant à la structure même des réseaux neuronaux est sa forme. Nos réseaux de neurones ont deux formes potentielles: triangulaires et inverses exponentielles. Dans les deux cas, le nombre de neurones diminue à travers les couches en convergeant vers le seul neurone de sortie. De plus, la première

couche cachée de nos réseaux ANN peut posséder un nombre de neurones allant du nombre de variables explicatives, soit de 122 à 61; ce dernier représente 50% du nombre de variables explicatives. Le troisième hyperparamètre est la fonction d'activation de chaque neurone du réseau ANN. Ce paramètre est crucial. En effet, les fonctions d'activation dictent la manière dont l'information est dispersée à travers le réseau. Dans notre univers d'hyperparamètres, nous incluons les fonctions d'activation suivantes : *ReLU*, *tanh*, *sigmoid*, *softsign*, *exponentiel*, *linéaire*, *softmax* et *softplus*. L'ensemble de ces fonctions est présenté en annexe. Nous n'imposons qu'une seule restriction : le choix de la fonction d'activation pour le dernier neurone doit être une fonction *linéaire*. Le quatrième hyperparamètre est le taux de désactivation aléatoire (*DR*) des neurones du réseau. Plus précisément, il s'agit du pourcentage de neurones par couche qui sera désactivé de manière aléatoire dans le système. Dans notre univers, le taux de désactivation est discret et peut prendre une valeur de 0%, 5%, 10%, 25% et 50%. Ce paramètre permet d'éviter le surapprentissage des ANN. Le cinquième hyperparamètre est la fonction de régularisation à exploiter et le sixième hyperparamètre est le coefficient de pénalisation des poids. Pour notre univers d'hyperparamètres, nous incluons la régularisation *Lasso* et *Ridge* telle que présentée en (4) et (5). Dans les deux cas, le coefficient de pénalisation des poids du réseau est défini entre 0.0001% et 1%. Tout comme le taux de désactivation aléatoire, ces deux hyperparamètres limitent le surapprentissage du réseau en pénalisant les poids extrêmes associés à certaines variables; nous cherchons à favoriser des structures neuronales plus simples. Le septième hyperparamètre est la fonction d'optimisation des poids du ANN. Dans notre univers, les fonctions d'optimisation possibles sont les suivantes : *Stochastic Gradient Descent* (SGD), le *Stochastic Gradient Descent* ajusté pour l'élan (SGD-Momentum), *Nesterov accelerated gradient* (NAG), *Root Mean Squared Propagation* (RMS), *Adam*, *Adadelta* et *Adagrad*. L'ensemble de ces fonctions est présenté en annexe. Finalement, le dernier hyperparamètre est la taille des données sur lesquelles le réseau optimise les poids. Ce paramètre communément appelé *Batch size* correspond au nombre d'échantillons de données utilisé dans une itération, ou époque, de la phase d'entraînement. Au bout d'une itération, le modèle ajuste les poids des connexions du réseau à l'aide de la fonction d'optimisation. Ainsi, plus le nombre d'échantillons est grand, plus l'ensemble des données fournies aux réseaux est divisé en petits groupes sur lesquels l'algorithme établit les poids des connexions du réseau. Dans notre univers d'hyperparamètres, le *Batch size* peut prendre des valeurs de 10%, 15% et 20% du nombre de données fournies.

2.4.2 Algorithme d'investissement

À chaque début de mois, notre algorithme d'apprentissage optimisé pour cette période fait une prédiction sur le rendement moyen futur standardisé par la volatilité de chaque action faisant partie du S&P500, tel que décrit à l'équation (11). Pour avoir un historique de données suffisamment grand, seules les firmes ayant 3 ans de données antérieures sont considérées dans l'univers du S&P500 à chaque pas de temps. Cette approche nous permet d'obtenir une vraie mesure de performance de notre stratégie d'investissement à travers le temps.

$$\frac{\bar{r}_t^i}{\sigma_t^2}, \quad \forall i \in [1,500] \quad (11)$$

Le choix de prédire les rendements moyens standardisés de chaque actif est grandement inspiré par la méthodologie de Moskowitz, Ooi, Pedersen (2012) et de Barroso and Santa Clara (2015), qui font la prédiction du rendement futur standardisé par la volatilité. D'une part, cette technique permet de gérer implicitement le risque inhérent de la stratégie d'investissement et rend les actifs plus comparables entre eux. D'autre part, en introduisant le rendement moyen, nous tentons de lisser la prédiction effectuée sur la variable d'intérêt. Le rendement et la volatilité sont estimés sur une base mensuelle. Intuitivement, notre variable d'intérêt se rapproche du ratio de Sharpe. Concrètement, pour faire la prédiction de cette variable, il faut tout d'abord estimer un réseau ANN sur ladite variable d'intérêt. Ceci implique de diviser l'ensemble de données fournies au réseau en trois groupes : les données d'entraînement, les données de validation et les données de test. Les données d'entraînement sont utilisées pour estimer le ANN pour un certain ensemble d'hyperparamètres. Les données de validation permettent d'ajuster les poids des connexions du réseau ANN. Finalement, les données de test sont utilisées pour estimer la performance du réseau de neurones à chaque mois; la performance est mesurée sur un ensemble de données jamais encourues par le réseau de neurones dans la phase d'entraînement et dans l'optimisation des hyperparamètres. Pour faciliter l'apprentissage du réseau ANN, nous centrons et nous réduisons l'ensemble des variables fourni au réseau par rapport aux données d'entraînement. Nous introduisons également le concept du « *shuffling* » qui permet essentiellement de briser la temporalité de nos données; la temporalité est déjà introduite dans les variables descriptives. Du

coup, à chaque mois, l'ensemble des données sur l'historique des firmes du S&P500 est séparé en trois ensembles de manière aléatoire selon une pondération 60%, 20% et 20%. Explicitement, 60% des données fournies sont utilisées pour l'ensemble d'entraînement et 20% pour l'ensemble de validation et de test. Cette technique permet d'avoir trois ensembles de données plus comparables, ce qui augmente le pouvoir prédictif des réseaux de neurones. Il est notable que certaines variables explicatives fournies aux réseaux de neurones se chevauchent, ce qui peut créer un léger biais dans les observations de test. Aussi, pour éviter le surapprentissage du réseau de neurones, nous utilisons un « *Early stop* » (ES), qui permet de mettre fin à l'entraînement du réseau ANN « prématurément ». Typiquement, la fonction de coût est convexe décroissante en fonction du nombre d'époques. Lorsque le nombre d'itérations ou d'époques devient trop important, le réseau ANN est en surapprentissage. L'objectif du ES est de stopper l'apprentissage du ANN avant d'entrer dans cette phase de surapprentissage. Plus précisément, lorsque la fonction de coût (MSE) n'a pas diminué après 50 itérations, ES met fin à l'apprentissage. Le modèle ANN optimal retenu pour une période quelconque sera celui qui aura le plus grand pouvoir prédictif, soit le plus grand coefficient de détermination dans l'ensemble de données de validation. Finalement, le réseau optimisé pour cette période émet des prédictions sur la variable d'intérêt décrite dans l'équation (11) de chaque titre du S&P500 à l'aide des plus récentes observations jamais analysées par nos réseaux de neurones. Tout comme la méthodologie de Jegadeesh et Titman (1993), nous construisons 10 portefeuilles en fonction de la prédiction émise sur chaque titre du S&P500. La stratégie d'investissement vise à être long et court deux portefeuilles en fonction des prédictions émises sur chaque titre. Pour avoir un meilleur point de comparaisons avec la stratégie de référence, les poids des titres en portefeuilles sont tous équipondérés pour former un portefeuille à coût nul; les positions à découvert financent les positions d'achat.

En somme, à chaque pas de temps, nous analysons les trois années de données historiques pour chaque firme du S&P500. L'ensemble des données est ensuite séparé en trois ensembles distincts puis centré et réduit selon l'ensemble d'entraînement. L'ensemble de données d'entraînement et de validation est utilisé pour optimiser une structure ANN qui nous permet de prédire la variable décrite en (11). La prédiction effectuée sur chaque titre du S&P500 est utilisée pour former des portefeuilles d'investissement.

Chapitre 3

Résultats empiriques

3.1 Performance de l'indice de référence : Jegadeesh et Titman (1993)

Pour former les portefeuilles de référence, à chaque mois, nous effectuons tout d'abord un classement des titres du S&P500 en fonction de leur rendement cumulé sur les 12 derniers mois, en excluant le rendement du dernier mois. Les firmes ayant le mieux performées sur cette période sont placées dans le portefeuille 10, les firmes *gagnantes*, et les firmes avec la plus faible performance sont placées dans le portefeuille 1, les firmes *perdantes*. Les titres en portefeuilles sont équipondérés pour former un portefeuille autofinancé et comparé avec le rendement du marché, soit l'indice du S&P500, et le rendement sans risque, soit le rendement d'une obligation 1 mois disponible sur le site de Kenneth French. Le tableau 1 présente la performance des portefeuilles des différents déciles entre juillet 1998 et juin 2018. La figure 2 présente le rendement mensuel cumulé sur cette période pour un investissement au taux sans risque, sur l'indice du marché, soit le S&P500, sur le portefeuille du premier et sur le dernier décile. La performance des portefeuilles risqués est ajustée pour le rendement sans risque.

Statistiques	Déciles de portefeuille MOM										MOM	S&P500
	10	9	8	7	6	5	4	3	2	1		
<i>Juillet 1998 - Juin 2018</i>												
r-f	8.33%	6.15%	3.09%	3.95%	5.81%	7.74%	6.30%	6.00%	5.05%	5.38%	0.99%	6.21%
σ	20.40%	16.28%	16.00%	16.00%	17.13%	17.41%	20.38%	21.54%	26.06%	36.71%	33.63%	18.84%
α	3.72%	1.43%	-1.75%	-0.98%	0.40%	2.10%	-0.15%	-0.80%	-2.90%	-5.12%	7.33%	0
$t(\alpha)$	3.765	2.777	-4.803	-2.964	1.246	7.209	-0.412	-2.065	-5.097	-4.717	4.064	0
β	0.723	0.755	0.796	0.806	0.872	0.893	1.039	1.099	1.303	1.736	-1.011	1
SR	0.408	0.378	0.193	0.247	0.339	0.445	0.309	0.278	0.194	0.147	0.029	0.330
Sk	-0.267	-0.593	-0.377	-0.576	-0.417	-0.049	0.035	-0.008	0.304	1.000	-1.495	-0.226

Tableau 1: Performance et caractéristiques des portefeuilles d'investissement des déciles MOM entre juillet 1998 et juin 2018. Les portefeuilles 1 et 10 incluent respectivement les firmes ayant connu les pires et meilleurs rendements sur les 12 derniers mois dans l'univers du S&P500, sans compter la performance du dernier mois. MOM correspond au portefeuille autofinancé qui est long et court respectivement le portefeuille 10 et 1. Les rendements moyens excédentaires, écarts types et alpha (MEDAF) sont en pourcentage. SR correspond au ratio de Sharpe, Sk correspond au coefficient d'asymétrie de la distribution des rendements et S&P500 correspond à l'indice du marché de référence. Les portefeuilles sont détenus sur une base mensuelle.

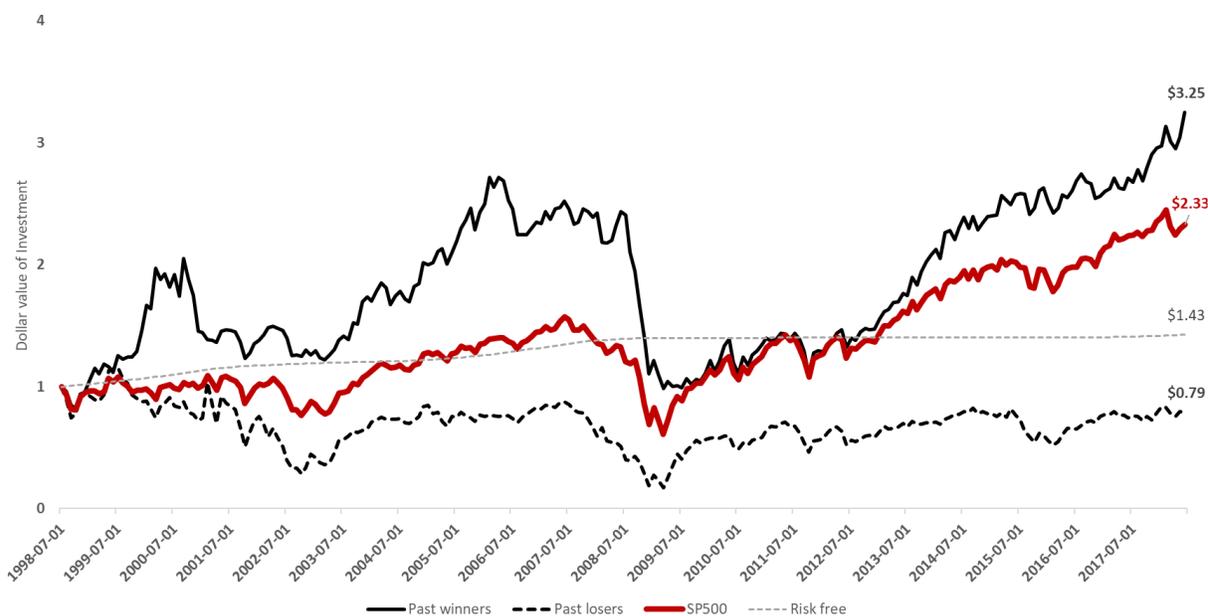


Figure 2: Portefeuille gagnant et perdants entre juillet 1998 et juin 2018. Les rendements cumulés présentés sont : (1) et (2) le rendement respectif du portefeuille du premier et du dernier décile selon la performance historique des 12 derniers mois détenus durant 1 mois, sans compter la performance du dernier mois, (3) le rendement de l'indice du marché, le S&P500 et (4) le rendement du taux sans risque, soit le rendement d'une obligation américaine 1 mois. À droite du graphique est présentée la valeur cumulée de chaque portefeuille. (1), (2) et (3) sont présentés en excès par rapport au taux sans risque

Sur l'ensemble de la période, le portefeuille des récents gagnants surperforme largement le portefeuille des récents perdants. En effet, le portefeuille *gagnant* possède une valeur cumulée ajustée pour le taux sans risque de 3.25\$ contre 0.79\$ pour le portefeuille perdant. Dans ces deux portefeuilles, nous remarquons l'impact de la crise financière d'octobre 2008; le portefeuille gagnant est grandement affecté par la crise. Le portefeuille MOM correspond au portefeuille autofinancé qui est long et court respectivement le portefeuille 10 (les gagnants) et le portefeuille 1 (les perdants). Ainsi, durant la bulle technologique des années 2000, exploiter une stratégie MOM s'est avéré très profitable. Ces résultats vont de pair avec ceux présents dans la littérature. Toutefois, chaque décile ne semble pas présenter les mêmes caractéristiques que celles rapportées dans les articles de références. Ici, le rendement moyen excédentaire n'est pas linéairement croissant à travers les déciles. Le tableau 1 présente en plus amples détails les caractéristiques de chaque décile des portefeuilles MOM sur cette période. Le dixième décile rapporte en moyenne un rendement annuel excédentaire de 8.33% comparativement à un rendement excédentaire annuel moyen de 5.38% pour le portefeuille de titres du premier décile. Le rendement excédentaire moyen positif du premier décile est notamment attribuable à la crise financière d'octobre 2008. À la suite de la crise, ce portefeuille a largement surperformé celui des récents *gagnants*. Ceci provient de la

grande corrélation du portefeuille avec le marché : le coefficient d'exposition au marché (« *beta* ») pour ce décile est de 1.736. Afin de mieux comprendre les performances du premier et du dernier décile, nous présentons à la figure 3 le rendement mensuel cumulé entre mars 2009 et juin 2011 des quatre investissements décrits plus haut.

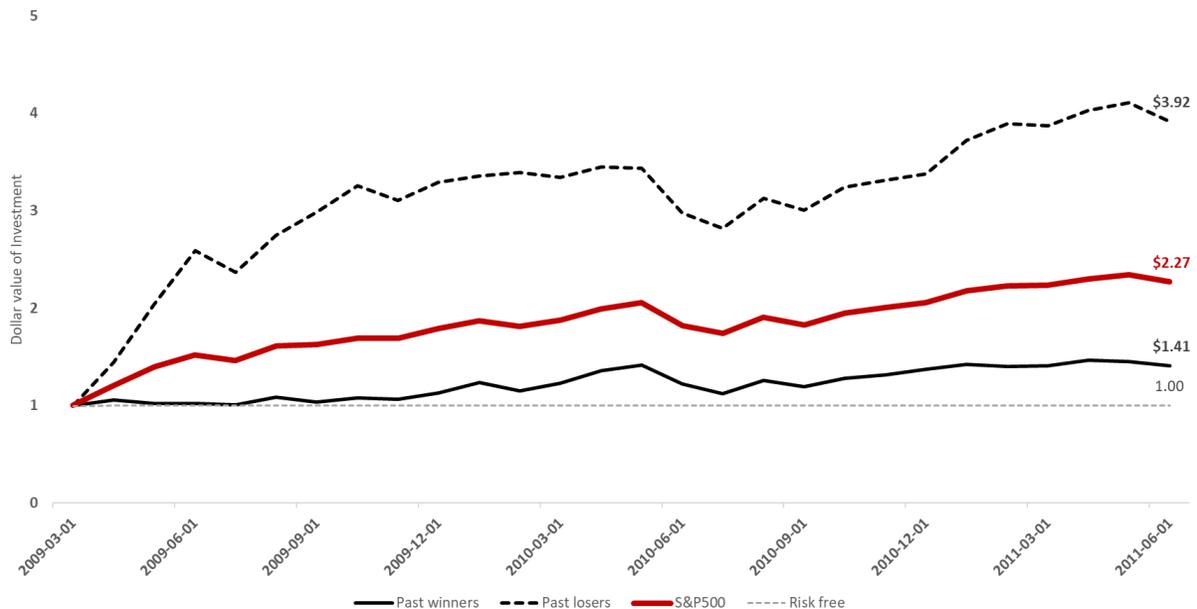


Figure 3: Portefeuille gagnant et perdants entre mars 2009 et juin 2011. Les rendements cumulés présentés sont : (1) et (2) le rendement respectif du portefeuille des récents gagnants et perdants selon la performance historique des 12 derniers mois détenus durant 1 mois, sans compter la performance du dernier mois, (3) le rendement de l'indice du marché, le S&P500 et (4) le rendement du taux sans risque, soit le rendement d'une obligation américaine 1 mois. À droite du graphique est présentée la valeur cumulée de chaque portefeuille. (1), (2) et (3) sont présentés en excès par rapport au taux sans risque.

Le portefeuille des récents perdants surperforme plus de 2 fois celui des récents gagnants. Ces observations sont également rapportées dans la recherche de Kent et Moskowitz (2016) qui attribue cette performance aux coefficients d'expositions au marché (« *beta* ») des déciles, qui varie non seulement à travers le temps, mais également entre les déciles. Ainsi, la crise financière affecte largement le rendement d'une stratégie MOM sur l'ensemble de la période. Sur la totalité de la période étudiée, ce portefeuille MOM ne génère qu'un rendement annuel de 0.99% tout en possédant une volatilité moyenne annuelle de 33.63% pour un faible rapport rendement risque de 0.029. Le coefficient d'exposition au marché (« *beta* ») de ce portefeuille est négatif à -1.011, ce qui lui permet de dégager un alpha (MEDAF) significatif de 7.33%. De son côté, le marché a dégagé un rendement moyen annuel excédentaire de 6.21%, une volatilité moyenne annuelle de

18.84% ainsi qu'un rapport rendement risque de 0.33. De plus, le coefficient d'asymétrie de chacun des portefeuilles est relativement croissant monotone à travers les déciles.

Somme toute, les caractéristiques des déciles des portefeuilles MOM semblent toutes différentes et intimement reliées à la période d'observation. Ainsi, le choix d'inclure ou non la crise financière d'octobre 2008 est très important dans l'analyse de la performance d'une stratégie MOM sur le marché du S&P500.

3.2 Performance de la stratégie issue des réseaux de neurones

À la fin de chaque mois, nos réseaux de neurones optimisés font une prédiction sur le rendement moyen du prochain mois standardisé par la volatilité propre à chaque titre du S&P500 décrit à l'équation (11). Ensuite, chaque titre du S&P500 est placé dans un décile en fonction du signal prédit. Les titres ayant le plus fort signal sont placés dans le dixième décile et les titres avec le plus faible signal sont placés dans le premier décile. Chaque titre d'un décile est équipondéré et détenu pour 1 mois; cette approche facilite la comparaison avec la stratégie de référence, soit le MOM classique de Jegadeesh et Titman (1993). Le tableau 2 rapporte la performance de chaque décile par rapport à l'indice du marché et au taux sans risque entre juillet 1998 et juin 2018. Les caractéristiques des déciles sont différentes à celles rapportées pour MOM sur le marché du S&P500.

Statistiques	Déciles de portefeuilles ANN										WML	S&P500
<i>Juillet 1998 - Juin 2018</i>	10	9	8	7	6	5	4	3	2	1		
r-rf	4.16%	8.99%	7.38%	8.24%	7.25%	5.46%	7.15%	6.85%	5.92%	0.93%	6.09%	6.21%
σ	25.49%	21.81%	21.44%	19.81%	18.99%	19.12%	18.68%	19.06%	18.40%	18.06%	15.54%	18.84%
α	-3.30%	1.97%	0.54%	1.82%	1.14%	-0.57%	1.15%	0.79%	0.40%	-3.80%	4.10%	0
$t(\alpha)$	-4.749	4.747	1.384	5.592	3.549	-1.706	3.625	2.216	0.811	-5.769	4.381	0
β	1.229	1.107	1.093	1.017	0.973	0.978	0.957	0.969	0.888	0.794	0.315	1
SR	0.163	0.412	0.344	0.416	0.382	0.286	0.383	0.359	0.322	0.051	0.392	0.330
Sk	0.675	-0.058	-0.382	-0.203	-0.230	-0.212	-0.136	-0.137	-0.214	-0.622	0.374	-0.226

Tableau 2 : Performance et caractéristiques des portefeuilles d'investissement ANN par déciles entre juillet 1998 et juin 2018. Les portefeuilles sont classés en décile en fonction de la prédiction de chaque titre de l'univers du S&P500. WML correspond au portefeuille autofinancé qui est long et court respectivement le portefeuille 9 et 1; le portefeuille est long et court les déciles avec la meilleure et la pire performance. Les rendements moyens excédentaires, écarts types et alpha (MEDAF) sont en pourcentage. SR correspond au ratio de Sharpe, Sk correspond au coefficient d'asymétrie de la distribution des rendements et S&P500 correspond à l'indice du marché de référence.

Tout comme pour la stratégie de référence MOM, le premier décile sous-performe les autres portefeuilles avec un rendement annuel moyen d'uniquement 0.93%. Ce portefeuille présente également une volatilité annuelle moyenne de 18.06%, un coefficient d'exposition au marché (« *beta* ») de 0.794 et un rapport rendement risque d'uniquement 0.051. En plus de sous-performer les autres déciles, ce portefeuille sous-performe également son comparable immédiat issu de la stratégie MOM. Un autre contraste intéressant entre les deux méthodologies provient du décile le plus performant. En exploitant les réseaux de neurones, le neuvième décile offre la meilleure performance avec un rendement moyen annuel excédentaire de 8.99%, une volatilité de 21.81%, un coefficient d'exposition au marché (« *beta* ») de 1.107, un alpha significatif (MEDAF) de 1.97% ainsi qu'un rapport rendement risque de 0.412. Quant à lui, le dixième décile ANN possède des caractéristiques très similaires au portefeuille composé des récents perdants de la stratégie MOM. En effet, le dixième décile ANN génère un rendement annuel excédentaire moyen de 4,16%, une volatilité annuelle moyenne élevée à 25.49%, un alpha (MEDAF) significatif négatif annuel de -3.30% et un grand coefficient d'exposition au marché (« *beta* ») de 1.229. Ainsi, le portefeuille « *Winners minus Losers* » (WML) issu des réseaux de neurones est long et court le neuvième et premier décile ANN respectivement; la stratégie vise à former un portefeuille autofinancé qui est long et court respectivement le décile avec la meilleure et la pire performance. En termes de performances, cette stratégie génère un rendement annuel excédentaire moyen de 6.09%, une volatilité annuelle moyenne de 15.54%, un alpha (MEDAF) significatif de 4.10%, un rapport rendement risque supérieur au marché de 0.392 et un coefficient d'exposition au marché (« *beta* ») de 0.315. La plus grande valeur ajoutée de cette stratégie d'investissement provient de sa volatilité : la volatilité moyenne est environ deux fois plus faible que celle issue du MOM de référence. Il est également intéressant d'observer qu'en excluant le dixième décile, le rendement excédentaire moyen et l'alpha (MEDAF) moyen semblent croissants à travers les déciles restants. La figure 4 présente le rendement mensuel cumulé entre juillet 1998 et juin 2018 pour un investissement au taux sans risque, sur l'indice du marché, soit le S&P500, et sur le portefeuille du neuvième et du premier décile. La performance des portefeuilles risqués est ajustée pour le rendement sans risque. Concrètement, à chaque pas de temps t , nos réseaux de neurones utilisent l'historique de données sur les 3 dernières années de chaque titre du S&P500 pour faire une prédiction sur la variable d'intérêt au temps $t+1$ décrite à l'équation (11). Ainsi, à chaque mois, le réseau de neurones fait une prédiction en temps réel. Le rendement cumulé mensuel du premier

décile illustre très bien sa faible volatilité. De plus, depuis la crise d'octobre 2008, ce portefeuille d'investissement présente une faible volatilité, mais également une faible corrélation avec l'indice du marché; les rendements sont restés pratiquement nuls à la suite de cette crise. Encore une fois, l'effet de la crise financière de 2008 est largement ressenti dans le portefeuille du neuvième décile. Toutefois, ce portefeuille semble en mesure de bien capter toutes les reprises de marché, soit entre janvier 2003 et juillet 2007 et entre mars 2009 à juin 2018.

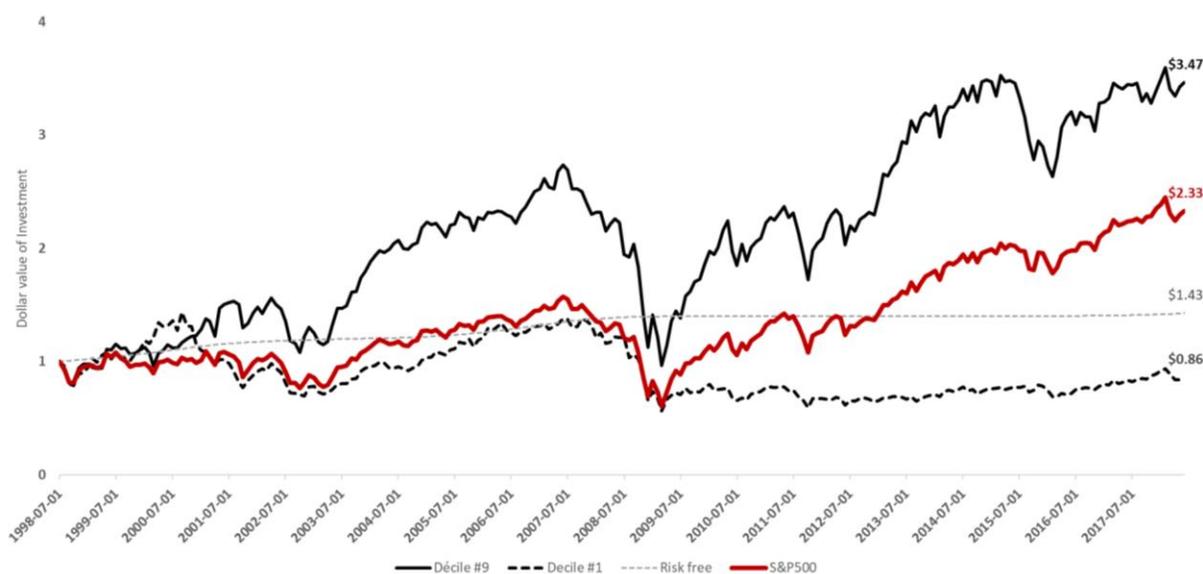


Figure 4: Entre juillet 1998 et juin 2018, les rendements cumulés présentés sont : (1) et (2) le rendement cumulé du 1er et du 9e décile selon la prédiction ANN. Ces déciles représentent respectivement le décile de portefeuille avec la pire et meilleure performance. (3) le rendement de l'indice du marché, le S&P500 et (4) le rendement du taux sans risque, soit le rendement d'une obligation américaine 1 mois. À droite du graphique est présentée la valeur cumulée de chaque portefeuille. (1), (2) et (3) sont présentés en excès par rapport au taux sans risque.

En somme, la stratégie d'investissement WML qui vise à construire un portefeuille d'investissement qui est long et court respectivement le décile avec la meilleure et la pire performance, soit le neuvième et le premier décile, semble offrir de bonnes performances. À la section suivante, nous présentons en plus amples détails la performance des rendements de la stratégie WML.

3.3 Caractéristiques des rendements de la stratégie d'investissement

Dans cette section, nous rapportons plusieurs mesures de performance de la stratégie MOM issue de la méthodologie de Jegadeesh et Titman (1993) sur le marché du S&P500 et celle obtenue grâce à nos réseaux de neurones ANN sur le même marché. Les métriques de performances sont présentées dans le tableau 3. La portion du haut présente des caractéristiques propres aux rendements générés par ces deux stratégies d'investissement, tandis que la portion du bas présente les coefficients d'exposition (« *beta* ») aux trois facteurs de Fama-French (1996) ainsi que le facteur MOM de Cahart (1997). Nous présentons également le rendement mensuel cumulé des deux stratégies d'investissement à la figure 5.

Juillet 1998 - Juin 2018	Stratégies	
	MOM	WML
<u>Statistiques</u>		
r-rf	0.99%	6.09%
σ	33.63%	15.54%
SR	2.95%	39.19%
Min Rend.	-44.2%	-15.7%
Max Rend.	29.0%	15.6%
Max Pertes Cum.	-91.3%	-37.5%
% Rend. Positif	56.72%	52.10%
<u>FF3 + UMD</u>		
α	-4.04%	6.26%
	(1.14)	2.09
β Mkt-rf	-0.180	0.095
	(2.43)	1.59
β SMB	-0.051	0.088
	(0.54)	1.16
β HML	-0.105	0.424
	(1.06)	5.35
β UMD	1.526	-0.346
	25.14	(7.09)
R^2	78.2%	33.9%

Tableau 3 : Performance et caractéristiques de la stratégie MOM et de la stratégie d'investissement issue des réseaux de neurones ANN entre juillet 1998 et juin 2018. MOM est le portefeuille autofinancé qui est long et court les titres ayant connu les meilleurs et pires rendements sur les 12 derniers mois, sans compter le dernier mois. WML est le portefeuille autofinancé qui long et court le 9e et le 1er décile en fonction de la prédiction effectuée sur les titres du S&P500. Le 9e et le 1er décile ANN ont respectivement la meilleure et pire performance empirique. SR correspond au ratio de Sharpe, Max Pertes Cum correspond au plus faible rendement cumulé réalisé sur la période et % Rend. Positif correspond à la proportion de rendement positif réalisé sur la période. La portion du bas présente les coefficients Beta aux trois facteurs de Fama-French (1996) ainsi que le facteur MOM de Cahart (1997). Les coefficients Beta des deux stratégies sont présentés accompagnés de leur statistique de test T en italique. Les facteurs de risques sont respectivement Mkt-rf le facteur du marché, SMB le facteur taille, HML le facteur valeur et UMD au facteur de risque MOM.

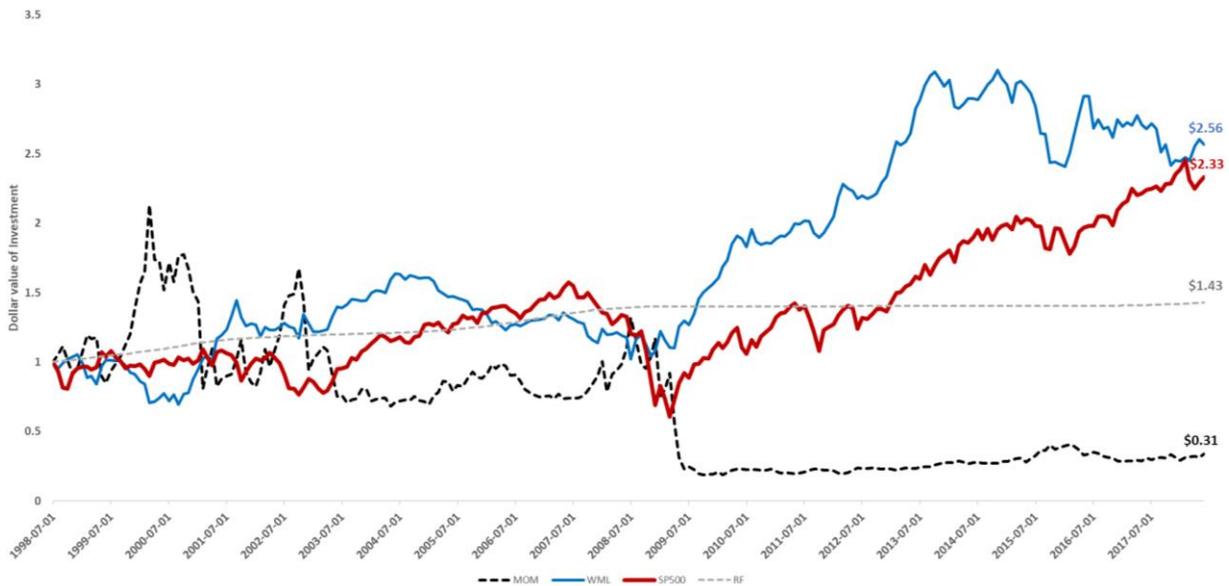


Figure 5: Les rendements cumulés présentés sont : (1) stratégie MOM sur le marché du S&P500 provenant du portefeuille qui est long et court les titres ayant connu les meilleurs et pires rendements sur les 12 derniers mois, (2) la stratégie WML provenant du portefeuille qui long et court le 9e et 1er décile respectivement en fonction de la prédiction des réseaux ANN. Le 9e et le 1er décile ont respectivement la meilleure et la pire performance empirique. (3) le rendement de l'indice du marché, le S&P500 et (4) le rendement du taux sans risque, soit le rendement d'une obligation américaine 1 mois. À droite du graphique est présentée la valeur cumulée de chaque portefeuille. (1), (2) et (3) sont présentés en excès par rapport au taux sans risque.

La stratégie d'investissement exploitant les réseaux de neurones surperforme largement la stratégie MOM de référence; les réseaux de neurones ANN génèrent un rendement annuel excédentaire moyen de plus de 6% comparativement à un peu moins de 1% pour la stratégie MOM. Avec un rendement mensuel minimal de -15.7%, un rendement mensuel maximal de 15.6% et un rendement mensuel cumulé minimal de -37.5%, les réseaux de neurones produisent une stratégie d'investissement qui est largement moins volatile que celle du MOM. De son côté, la stratégie de référence MOM possède une volatilité moyenne annuelle de 33.63% et un rendement cumulé minimal de -91.3%. Cette perte drastique se matérialise majoritairement à la suite de la bulle technologique des années 2000 et de la crise financière d'octobre 2008. Encore une fois, cette stratégie est très sensible aux périodes de turbulence sur le marché du S&P500. Au niveau des coefficients d'exposition (« *beta* ») aux facteurs de risques usuels d'investissement, sans surprise, la stratégie MOM sur le marché du S&P500 est grandement corrélée avec le facteur UMD de Cahart (1997). De plus, lorsqu'elle est ajustée pour les différents facteurs de risque, cette dernière

gène un rendement mensuel moyen alpha non significatif de -4.04%. Avec un coefficient de détermination de plus de 78%, la majorité des rendements de la stratégie MOM sont explicables par le facteur de risque UMD et le facteur de marché. De son côté, la stratégie WML, issue des ANN, génère un rendement alpha moyen significatif de plus de 6.26%. En termes de coefficient d'exposition (« beta ») aux différents facteurs de risques, les rendements ne sont corrélés significativement qu'avec les facteurs HML et UMD. Toutefois, les coefficients d'exposition (« beta ») sont respectivement positifs et négatifs aux facteurs HML et UMD. Ce résultat illustre très bien la propriété d'apprentissage par expérience des réseaux de neurones. Telle que mentionnée plus haut, la stratégie MOM n'a pas bien performé entre juillet 1998 et juin 2018. De son côté, le facteur valeur a démontré de bonnes performances et a dégagé un rendement positif sur l'ensemble de la période. Ainsi, nos réseaux de neurones ont été en mesure de générer un rendement positif qui s'explique en partie par un aspect de valeur des firmes de l'univers. Toutefois, l'ensemble des trois facteurs de Fama-French n'est qu'en mesure d'expliquer 33.9% de la variabilité totale des rendements de notre stratégie ANN. Il est important de reconnaître, encore une fois, que notre stratégie d'investissement est grandement inspirée par le MOM, mais n'est pas limitée aux interactions pure MOM; nos réseaux de neurones ont le loisir d'apprendre des relations qui vont au-delà des prix passés. En utilisant uniquement les variables en lien avec les prix des titres du S&P500, la stratégie est d'autant plus corrélée avec la stratégie MOM de référence. Du coup, la performance du modèle est grandement attribuable à la diversité des données fournies aux réseaux de neurones. Bien que tous justifiés par leur pouvoir explicatif sur une stratégie de MOM classique, les variables explicatives fournies aux réseaux de neurones complètent les variables qui atraient seulement au prix des titres boursiers.

Somme toute, nos réseaux de neurones produisent une stratégie profitable sur l'ensemble de la période considérée. De plus, les rendements issus de notre stratégie ne sont pas grandement affectés par la crise financière de 2008; la performance est stable. Cependant, nos réseaux de neurones ne sont pas en mesure de bien capter la bulle technologique des années 2000 et la crise du pétrole de 2015.

Conclusion

L'utilisation des réseaux de neurones s'avère à être une excellente méthode pour capter les multiples interactions présentes dans les rendements historiques des actions du S&P500. Alors que les rendements de la stratégie MOM classique de Jegadeesh et Titman (1993) n'offrent pas une bonne performance entre juillet 1998 et juin 2018, les réseaux de neurones optimisés ANN surperforment non seulement la stratégie MOM, mais également l'indice du marché, soit le S&P500. Ainsi, la stratégie qui vise à acheter et à vendre respectivement un portefeuille d'actions en fonction du signal de prédiction émis sur chaque titre du S&P500 par les réseaux de neurones est profitable. La profitabilité de cette stratégie d'investissement est entre autres attribuable à notre méthodologie. D'une part, nous faisons la gestion implicite du risque grâce à la variable que nous modélisons, soit le rendement moyen mensuel standardisé par la volatilité de chaque titre de notre univers; la variable de prédiction est un bon intermédiaire au ratio de Sharpe. Cette approche a l'avantage de lisser les rendements futurs tout en prenant compte du risque de chaque actif. La stratégie d'investissement génère des rendements environ 2 fois moins volatiles que la stratégie de référence et un rendement moyen 6 fois plus élevé. D'autre part, notre choix d'implémenter l'optimisation d'hyperparamètres des réseaux de neurones de Bengio et Bergstra (2012) semble également être judicieux. Plutôt que de construire des réseaux de neurones naïfs, tels que plusieurs autres recherches en finance de marché semblent le faire, nous optimisons la structure de nos réseaux afin de minimiser notre fonction de coût, soit la moyenne des erreurs au carré. Ainsi, avant chaque prédiction sur notre variable d'intérêt, nous nous assurons que la structure même du réseau de neurones soit optimale, ce qui nous permet de n'émettre aucune hypothèse restrictive quant à leur forme. Le succès général de cette stratégie vient sans doute de la capacité d'apprentissage par expérience des réseaux de neurones. Bien que toutes les variables descriptives fournies aux réseaux soient justifiées pour leur pouvoir prédictif ou explicatif des rendements MOM, la stratégie d'investissement est négativement et significativement corrélée au facteur de risque MOM de Cahart (1997). En revanche, les rendements de la stratégie sont positivement corrélés avec le facteur valeur de Fama-French (1996), qui a connu une bonne performance durant notre période d'analyse; les réseaux de neurones sont en mesure de capter des caractéristiques de firmes qui ont bien performées.

Bibliographie

Asness, Cliff S., John M. Liew, and Ross L. Stevens. "Parallels between the cross-sectional predictability of stock and country returns." *Journal of Portfolio Management* 23.3 (1997): 79.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen. "Value and momentum everywhere." *The Journal of Finance* 68.3 (2013): 929-985.

Asness, Clifford S., et al. "Fact, fiction and momentum investing." *Journal of Portfolio Management, Fall* (2014).

Barroso, Pedro, and Pedro Santa-Clara. "Momentum has its moments." *Journal of Financial Economics* 116.1 (2015): 111-120.

Bergstra, James S., et al. "Algorithms for hyper-parameter optimization." *Advances in neural information processing systems*. 2011.

Bergstra, James, and Yoshua Bengio. "Random search for hyper-parameter optimization." *Journal of Machine Learning Research* 13.Feb (2012): 281-305.

Bergstra, James, Dan Yamins, and David D. Cox. "Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms." *Proceedings of the 12th Python in science conference*. 2013.

Bergstra, James, Daniel Yamins, and David Daniel Cox. "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures." (2013).

Carhart, Mark M. "On persistence in mutual fund performance." *The Journal of finance* 52.1 (1997): 57-82.

Chordia, Tarun, and Lakshmanan Shivakumar. "Momentum, business cycle, and time-varying expected returns." *The Journal of Finance* 57.2 (2002): 985-1019.

Conrad, Jennifer, and Gautam Kaul. "An anatomy of trading strategies." *The Review of Financial Studies* 11.3 (1998): 489-519.

Cooper, Michael J., Roberto C. Gutierrez Jr, and Allaudeen Hameed. "Market states and momentum." *The Journal of Finance* 59.3 (2004): 1345-1365.

Daniel, Kent, and Tobias J. Moskowitz. "Momentum crashes." *Journal of Financial Economics* 122.2 (2016): 221-247.

Datar, Vinay T., Narayan Y. Naik, and Robert Radcliffe. "Liquidity and stock returns: An alternative test." *Journal of Financial Markets* 1.2 (1998): 203-219.

Eugene, Fama, and R. French Kenneth. "Multifactor explanations of asset pricing anomalies." *Journal of Finance* 51.1 (1996): 55-84.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

Gorton, Gary B., Fumio Hayashi, and K. Geert Rouwenhorst. "The fundamentals of commodity futures returns." *Review of Finance* 17.1 (2012): 35-105.

Grundy, Bruce D., and J. Spencer Martin. "Understanding the nature of the risks and the source of the rewards to momentum investing." *The Review of Financial Studies* 14.1 (2001): 29-78.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. *Empirical asset pricing via machine learning*. No. w25398. National Bureau of Economic Research, 2018.

Gutierrez Jr, Roberto C., and Eric K. Kelley. "The long-lasting momentum in weekly returns." *The Journal of finance* 63.1 (2008): 415-447.

Israel, Ronen, and Tobias J. Moskowitz. "The role of shorting, firm size, and time on market anomalies." *Journal of Financial Economics* 108.2 (2013): 275-301.

Jegadeesh, Narasimhan, and Sheridan Titman. "Returns to buying winners and selling losers: Implications for stock market efficiency." *The Journal of finance* 48.1 (1993): 65-91.

Jegadeesh, Narasimhan, and Sheridan Titman. "Profitability of momentum strategies: An evaluation of alternative explanations." *The Journal of finance* 56.2 (2001): 699-720.

Lee, Charles MC, and Bhaskaran Swaminathan. "Price momentum and trading volume." *the Journal of Finance* 55.5 (2000): 2017-2069.

Menkhoff, Lukas, et al. "Currency momentum strategies." *Journal of Financial Economics* 106.3 (2012): 660-684.

Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Heje Pedersen. "Time series momentum." *Journal of financial economics* 104.2 (2012): 228-250.

Moskowitz, Tobias J., and Mark Grinblatt. "Do industries explain momentum?." *The Journal of finance* 54.4 (1999): 1249-1290.

Okunev, John, and Derek White. "Do momentum-based strategies still work in foreign currency markets?." *Journal of Financial and Quantitative Analysis* 38.2 (2003): 425-447.

Patel, Jigar, et al. "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques." *Expert Systems with Applications* 42.1 (2015): 259-268.

Raffinot, Thomas, and Sylvain Benoit. "Investing through Economic Cycles with Ensemble Machine Learning Algorithms." *Available at SSRN 2785583* (2018).

Rouwenhorst, K. Geert. "International momentum strategies." *The journal of finance* 53.1 (1998): 267-284.

Ruder, Sebastian. "An overview of gradient descent optimization algorithms." *arXiv preprint arXiv:1609.04747*(2016).

Sharpe, William F. "Capital asset prices: A theory of market equilibrium under conditions of risk." *The journal of finance* 19.3 (1964): 425-442.

Sirignano, Justin, Apaar Sadhwani, and Kay Giesecke. "Deep learning for mortgage risk." *arXiv preprint arXiv:1607.02470*(2016).

Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." *The Journal of Machine Learning Research* 15.1 (2014): 1929-1958.

Wang, Jian-Zhou, et al. "Forecasting stock indices with back propagation neural network." *Expert Systems with Applications* 38.11 (2011): 14346-14355.

Wang, Kevin Q., and Jianguo Xu. "Market volatility and momentum." *Journal of Empirical Finance* 30 (2015): 79-91.

Annexes

Liste détaillée des variables utilisées

Nom	Caractéristiques de la firme	Auteur	Année	Journal de publication
1 Beta	Coefficient Beta au S&P500 *	Fama & Macbeth	1973	The Journal of Political Economy
2 BetaSQ	Beta ² *	Fama & Macbeth	1973	The Journal of Political Economy
3 BAspread	Écart bid-ask	Amihud & Mendelson	1989	The journal of Finance
4 ChCSHO	Changement nombre de titre	Pontiff & Woodgate	2008	The journal of Finance
5 DolVol	Volume en dollar transigé	Chordia, Subrahmanyam & Anshuman	2001	The journal of Finance
6 DP	Ratio dividende sur price	Litzenberger & Ramaswamy	1982	The journal of Finance
7 Idiovol	Volatilité idiosyncratique *	Ali, Hwang & Trombley	2003	The journal of Finance Economic
8 Ill	Liquidité	Amihud	2002	The journal of finance market
9 Indmom	MOM 12 mois par industrie	Moskowitz & Grinblatt	1999	The journal of Finance
10 Maxret	Rendement maximum	Bali, Cakici & Whitelaw	2011	The journal of Finance Economic
11 Mom12m	MOM 12 mois	Jegadeesh	1990	The journal of Finance
12 Mom1m	MOM 1 mois	Jegadeesh & Titman	1993	The journal of Finance
13 Mom3m	MOM 3 mois	Jegadeesh & Titman	1993	The journal of Finance
14 Mom6m	MOM 6 mois	Jegadeesh & Titman	1993	The journal of Finance
15 Chmom	Changement MOM 6 mois	Gettleman & Marks	2006	Working paper
16 Mvel1	Capitalisation boursière	Banz	1981	The journal of Finance Economic
17 RetVol	Volatilité *	Ang, Hodrick, Xing & Zhang	2006	The journal of Finance
18 Std_DolVol	Volatilité sur volume en dollar transigé	Chordia, Subrahmanyam & Anshuman	2001	The journal of Finance
19 Std_turn	Volatilité sur roulement des titres	Chordia, Subrahmanyam & Anshuman	2001	The journal of Finance
20 Turn	Roulement des titres	Datar, Naik & Radcliffe	1998	The journal of finance market
21 B_UMD	Coefficient Beta au facteur MOM	Cahart	1997	The journal of Finance
22 B_HML	Coefficient Beta au facteur valeur	Fama & French	1996	The journal of Finance
23 B_SMB	Coefficient Beta au facteur taille	Fama & French	1996	The journal of Finance
24 B_Mkt	Coefficient Beta au facteur du marché	Fama & French	1996	The journal of Finance
25 Mkt	Rendement excédentaire du marché	William F. Sharpe	1964	The journal of Finance
* : rendement calculé sur 12, 6, & 3 mois				
** : rendement calculé sur 12, 6, 3 & 1 mois				

Nom	Caractéristiques de la firme	Auteur	Année	Journal de publication
26 UMD	Rendement du facteur MOM	Cahart	1997	The journal of Finance
27 HmL	Rendement du facteur valeur	Fama & French	1996	The journal of Finance
28 SMB	Rendement du facteur taille	Fama & French	1996	The journal of Finance
29 Alpha	Rendement excédentaire (3 FF + MOM)	Cahart	1997	The journal of Finance
30 R_Rbar	Rendement excédentaire à la moyenne des rendements **	Jegadeesh & Titman	2002	The journal of Finance
31 MA	Moyenne mobile **	Okunev & White	2003	Journal of Financial and Quantitative Analysis
32 UP_DOWN	Signe du rendement 3 ans mobile du marché (Bull or Bear market)	Cooper, Gutierrez, Hameed	2004	The journal of Finance
33 Bear	Signe du rendement 2 ans mobile du marché	Kent & Moskowitz	2016	The journal of Finance Economic
34 DGS3MO	Rendement obligataire gouvernemental 3 mois	Benoit & Raffinot	2018	SSRN
35 DGS6MO	Rendement obligataire gouvernemental 6 mois	Benoit & Raffinot	2018	SSRN
36 DGS1	Rendement obligataire gouvernemental 1 an	Benoit & Raffinot	2018	SSRN
37 DGS2	Rendement obligataire gouvernemental 2 ans	Benoit & Raffinot	2018	SSRN
38 DGS5	Rendement obligataire gouvernemental 5 ans	Benoit & Raffinot	2018	SSRN
39 DGS10	Rendement obligataire gouvernemental 10 ans	Benoit & Raffinot	2018	SSRN
40 SP6M_3M	DGS6MO - DGS3MO	Benoit & Raffinot	2018	SSRN
41 SP1Y_3M	DGS1 - DGS3MO	Benoit & Raffinot	2018	SSRN
42 SP1Y_6M	DGS6MO - DGS3MO	Benoit & Raffinot	2018	SSRN
43 SP2Y_3M	DGS2 - DGS3MO	Benoit & Raffinot	2018	SSRN
44 SP2Y_6M	DGS2 - DGS6MO	Benoit & Raffinot	2018	SSRN
45 SP2Y_1Y	DGS1 - DGS1	Benoit & Raffinot	2018	SSRN
46 SP5Y_3M	DGS5 - DGS3MO	Benoit & Raffinot	2018	SSRN
47 SP5Y_6M	DGS5 - DGS6MO	Benoit & Raffinot	2018	SSRN
48 SP5Y_1Y	DGS5 - DGS1	Benoit & Raffinot	2018	SSRN
49 SP5Y_2Y	DGS5 - DGS2	Benoit & Raffinot	2018	SSRN
50 SP10Y_3M	DGS10 - DGS3MO (Term spread)	Benoit & Raffinot	2018	SSRN

* : rendement calculé sur 12, 6, & 3 mois
** : rendement calculé sur 12, 6, 3 & 1 mois

Nom	Caractéristiques de la firme	Auteur	Année	Journal de publication
51 SP10Y_6M	DGS10 - DGS?MO	Benoit & Raffinot	2018	SSRN
52 SP10Y_1Y	DGS10 - DGS1	Benoit & Raffinot	2018	SSRN
53 SP10Y_2Y	DGS10 - DGS2	Benoit & Raffinot	2018	SSRN
54 SP10Y_5Y	DGS10 - DGS5	Benoit & Raffinot	2018	SSRN
55 AAA	Rendement obligataire corporative AAA	Benoit & Raffinot	2018	SSRN
56 AA	Rendement obligataire corporative AA	Benoit & Raffinot	2018	SSRN
57 A	Rendement obligataire corporative A	Benoit & Raffinot	2018	SSRN
58 BAA	Rendement obligataire corporative BAA	Benoit & Raffinot	2018	SSRN
59 HY	Rendement obligataire <i>High yield</i>	Benoit & Raffinot	2018	SSRN
60 AAA_5y	AAA - DSG5	Benoit & Raffinot	2018	SSRN
61 AA_5y	AA - DSG5	Benoit & Raffinot	2018	SSRN
62 A_5y	A - DSG5	Benoit & Raffinot	2018	SSRN
63 BAA_5y	BAA - DSG5	Benoit & Raffinot	2018	SSRN
64 HY_5Y	HY - DSG5	Benoit & Raffinot	2018	SSRN
65 SP500_LC	Rendement moyen 12 mois <i>Large Cap.</i> S&P500	Benoit & Raffinot	2018	SSRN
66 SP500_MID	Rendement moyen 12 mois <i>Mid Cap.</i> S&P500	Benoit & Raffinot	2018	SSRN
67 SP500_SM	Rendement moyen 12 mois <i>Small Cap.</i> S&P500	Benoit & Raffinot	2018	SSRN
68 VIX index	Indice de volatilité VIX	Benoit & Raffinot	2018	SSRN
69 Oil	Prix contrat futur sur le pétrole	Benoit & Raffinot	2018	SSRN
70 NatG	Prix contrat futur sur le gas naturel	Benoit & Raffinot	2018	SSRN
71 Gold	Prix contrat futur sur l'or	Benoit & Raffinot	2018	SSRN
72 Silver	Prix contrat futur sur l'argent	Benoit & Raffinot	2018	SSRN
73 CRB_Index	Indice de commodité <i>Thomson Reuters</i> (CRB Index)	Benoit & Raffinot	2018	SSRN
74 Energy	Rendement mensuel secteur énergie GICS			
75 Financial	Rendement mensuel secteur financier GICS			
76 Material	Rendement mensuel secteur matériaux GICS			
77 Info_Tech	Rendement mensuel secteur technologie GICS			
78 Industrial	Rendement mensuel secteur industriel GICS			
79 Communication	Rendement mensuel secteur communication GICS			
80 Con_Discr	Rendement mensuel secteur consommation discrétionnaire GICS			
81 Utilities	Rendement mensuel secteur utilité GICS			
82 Cons_Sta	Rendement mensuel secteur produit GICS			
83 Health	Rendement mensuel secteur santé GICS			
* : rendement calculé sur 12, 6, & 3 mois				
** : rendement calculé sur 12, 6, 3 & 1 mois				

Liste des fonctions d'activation définies dans notre univers d'hyperparamètres

Relu :

$$f(x) = \begin{cases} x & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$$

LeakyRelu :

$$f(x) = \begin{cases} x & \text{si } x > 0 \\ \alpha x & \text{sinon} \end{cases}$$

Sigmoid:

$$f(x) = \frac{1}{1 + e^{-x}}$$

Softsign :

$$f(x) = \frac{x}{1 + \text{abs}(x)}$$

Tanh :

$$f(x) = \frac{2}{1 + e^{-2x}} - 1$$

Exponentiel :

$$f(x) = e^x$$

Linéaire :

$$f(x) = x$$

Listes des fonctions d'optimisation définie dans notre univers d'hyperparamètres⁴

Stochastic Gradient Descent (SGD)

$$\theta = \theta - \eta * \nabla_{\theta} L(\theta, x_i; y_i)$$

Où L est la fonction objective à minimiser, θ l'ensemble de paramètres de la fonction objective, η le pas d'ajustement des paramètres, ou « *learning rate* », ∇_{θ} le gradient de la fonction objective du système caractérisé par θ .

Stochastic Gradient Descent ajusté pour l'élan (SGD MOM)

$$\begin{aligned} \mathbf{V}_t &= \gamma \mathbf{V}_{t-1} - \eta * \nabla_{\theta} L(\theta) \\ \theta &= \theta - \mathbf{V}_t \end{aligned}$$

Où \mathbf{V}_t est le vecteur d'ajustement des paramètres et γ le coefficient d'ajustement du vecteur d'ajustement.

Nesterov accelerated gradient (NAG)

$$\begin{aligned} \mathbf{V}_t &= \gamma \mathbf{V}_{t-1} - \eta * \nabla_{\theta} L(\theta - \gamma \mathbf{V}_{t-1}) \\ \theta &= \theta - \mathbf{V}_t \end{aligned}$$

Adagrad

$$\begin{aligned} \mathbf{g}_{t,i} &= \nabla_{\theta_t} L(\theta_{t,i}) \\ \theta_{t+1,i} &= \theta_{t,i} - \frac{\eta}{\sqrt{\mathbf{G}_{t,ii} + \epsilon}} * \mathbf{g}_{t,i} \\ \theta_{t+1} &= \theta_t - \frac{\eta}{\sqrt{\mathbf{G}_t + \epsilon}} \odot \mathbf{g}_t \end{aligned}$$

Ce modèle modifie le pas d'ajustement général η à chaque temps t et pour chaque paramètre du système i . Où $\mathbf{G}_{t,ii}$ est une matrice diagonale dont les éléments i,i sont les sommes des gradients carrés de $\theta_{t,i}$ et ϵ est une constante infiniment petite qui évite la division par zéro.

⁴ Ruder (2016) présente de manière exhaustive l'ensemble des différentes fonctions d'optimisation fréquemment utilisé dans le développement de réseaux de neurones.

Adadelta

$$E(\mathbf{g}^2)_t = \gamma E(\mathbf{g}^2)_{t-1} - (1 - \gamma) \mathbf{g}^2_t$$
$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \frac{\eta}{\sqrt{E(\mathbf{g}^2)_t + \epsilon}} * \mathbf{g}_t$$

Où $E(\mathbf{g}^2)_t$ est la moyenne des gradients carrés historiques pondérée par γ .

Root Mean Squarred Propagation (RMS)

$$E(\mathbf{g}^2)_t = 0.9 E(\mathbf{g}^2)_{t-1} - 0.1 \mathbf{g}^2_t$$
$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \frac{\eta}{\sqrt{E(\mathbf{g}^2)_t + \epsilon}} * \mathbf{g}_t$$

Adam

$$\mathbf{m}_t = \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \mathbf{g}_{t,i}$$
$$\mathbf{v}_t = \beta_2 \mathbf{v}_{t-1} + (1 - \beta_2) \mathbf{g}^2_{t,i}$$

Où \mathbf{m}_t et \mathbf{v}_t estiment le premier et le deuxième moment des gradients passés. Pour corriger le biais imposé sur les valeurs initiales à zéro de \mathbf{m}_t et \mathbf{v}_t , on estime plutôt :

$$\widehat{\mathbf{m}}_t = \frac{\mathbf{m}_t}{1 - \beta_{t,1}}$$
$$\widehat{\mathbf{v}}_t = \frac{\mathbf{v}_t}{1 - \beta_{t,2}}$$
$$\boldsymbol{\theta}_{t+1,i} = \boldsymbol{\theta}_{t,i} - \frac{\eta}{\sqrt{\widehat{\mathbf{v}}_t + \epsilon}} * \widehat{\mathbf{m}}_t$$

