

Louvain School of Management

Naïve Buying Diversification Versus Naïve Portfolio Diversification

Analyse de l'impact de l'approche naïve des investisseurs et des biais comportementaux sur la performance des portefeuilles d'actions

Auteur : Aurélian VANLIER

Promoteur(s) : Pr. Catherine D'HONDT

Année académique 2023-2024

Intitulé du master et de la finalité : Master en Ingénieur de Gestion (120 ECTS)

Déclaration concernant l'utilisation des outils d'IA dans le mémoire de master

Nous reconnaissons que les outils d'IA peuvent être des aides précieuses pendant le travail sur le mémoire de master, mais ils ne sont pas infaillibles. Rappelez-vous que la transparence favorise la confiance et que reconnaître le rôle de l'IA améliore la crédibilité de votre travail.

Par conséquent, lorsque vous décidez d'utiliser un tel outil, vous devez respecter les principes suivants d'utilisation responsable de l'IA.

1. Évaluation Critique :

- Nous avons évalué de manière critique les résultats générés par l'IA, en veillant à ce qu'ils soient en adéquation avec nos objectifs de recherche.
- Toute modification ou correction a été effectuée sur la base de notre expertise et de nos connaissances dans le domaine.

2. Transparence :

- Nous reconnaissons l'utilisation de ChatGPT de manière transparente, en soulignant qu'il a contribué à notre travail mais n'a pas remplacé le jugement humain.
- Notre engagement envers la transparence assure l'intégrité de ce mémoire.

3. Considérations Éthiques :

- Nous avons activement surveillé les biais ou conséquences involontaires introduits par l'outil d'IA.
- Notre responsabilité éthique a guidé nos décisions tout au long du processus de recherche.

Déclaration

Pendant la préparation de ce mémoire de master, l'auteur a utilisé ChatGPT pour les raisons suivantes :

1. Tout d'abord, il a été utilisé pour vérifier l'accord et l'orthographe de certains mots et verbes afin d'assurer l'exactitude linguistique de sections spécifiques du mémoire. Ensuite, il a aidé à améliorer certains codes R déjà en partie développés, en réorganisant le code et en garantissant que les résultats soient affichés correctement.

2. Après avoir utilisé ChatGPT, l'auteur a soigneusement examiné et édité le contenu produit par l'outil. Nous prenons l'entière responsabilité du contenu final présenté dans ce mémoire.

En signant cette déclaration, nous affirmons que le contenu de ce mémoire de master reflète notre travail original, enrichi par l'utilisation responsable de l'IA.

A. Vanlier le 07/08/2024

Résumé

Ce mémoire explore l'utilisation de la Naïve Buying Diversification (NBD) par les investisseurs belges, en se basant sur l'étude de Gathergood et al. Mes résultats montrent que les investisseurs belges ont tendance à utiliser la NBD lors des jours d'achat multiples, en répartissant leurs investissements de manière égale entre les actions achetées, et préfèrent la NBD à la Naïve Portfolio Diversification (NPD). Bien que cette stratégie simplifie la prise de décision, elle entraîne souvent une diversification sous-optimale et de faibles rendements. De plus, mon étude révèle que certaines variables démographiques et financières influencent de manière significative la probabilité d'utiliser la NBD, les investisseurs ayant une meilleure connaissance des marchés financiers sont plus susceptibles d'utiliser la NBD, tandis que ceux ayant un revenu net annuel plus élevé sont moins enclins à y recourir. Ces conclusions confirment certaines observations de Gathergood et al., tout en soulignant l'importance de prendre en compte à la fois les comportements d'investissement spécifiques à chaque région, mais également les contextes économiques et les périodes d'observation spécifiques pour améliorer les stratégies de gestion de portefeuille.

Remerciements

Je tiens à exprimer ma profonde gratitude à ma promotrice, Pr. Catherine D'Hondt, pour sa disponibilité, ses retours rigoureux, ses conseils précieux et son soutien constant tout au long de la rédaction de ce mémoire.

Je souhaite également remercier sincèrement le corps enseignant et le personnel académique de l'UCLouvain FUCaM Mons qui ont contribué à mon parcours universitaire.

Mes remerciements s'étendent également à ma famille et à mes amis qui m'ont supporté et encouragé à travers la rédaction de ce mémoire.

Table des matières

Introduction	1
I. Revue de littérature	2
II. Hypothèses	11
III. Données	13
IV. Tests des hypothèses	16
V. Conclusion	39
Bibliographie.....	41
Annexes	43
I. Revue de littérature	43
Annexe 1 : Résumé de la discussion dans la littérature des avantages et inconvénients de la diversification naïve du portefeuille.....	43
II. Hypothèse 1	45
Annexe 2 : Table tolérance $\pm 2\%$	45
Annexe 3 : Table tolérance $\pm 5\%$	45
Annexe 4 : Table tolérance $\pm 10\%$	45
III. Hypothèse 2	46
Annexe 5 : Table tolérance $\pm 2\%$	46
Annexe 6 : Table tolérance $\pm 5\%$	46
Annexe 7 : Table tolerance $\pm 10\%$	47
IV. Hypothèse 4.....	48
Annexe 8 : Table SP_Q7; Proportion de NBD par revenu net annuel total.....	48
Annexe 9 : Table SP_Q3; Proportion de NBD par niveau de connaissance des marchés financiers.....	49
Annexe 10 : Table Proportion de NBD par Sexe	50
Annexe 11 : Table Proportion de NBD par groupe d'âge	50
V. Codes utilisés & explications :	51
Annexe 12 : Code hypothèse 1	51

Annexe 13 : Code hypothèse 2	61
Annexe 14 : Code hypothèse 3	70
Annexe 15 : Code hypothèse 4	77

Liste des tableaux

Tableau 1 : Hypothèse 1; tolérance $\pm 2\%$	17
Tableau 2 : Hypothèse 1; tolérance $\pm 5\%$	17
Tableau 3 : Hypothèse 1; tolérance $\pm 10\%$	17
Tableau 4 : Hypothèse 2; tolérance $\pm 2\%$	22
Tableau 5 : Hypothèse 2; tolérance $\pm 5\%$	23
Tableau 6 : Hypothèse 2; tolérance $\pm 10\%$	24
Tableau 7 : Tableau des statistiques descriptives	27
Tableau 8 : Résultats de la régression logistique modèle 1	34
Tableau 9 : Résultats de la régression logistique modèle 2	34
Tableau 10 : Proportion de NBD par revenu net annuel total	37
Tableau 11 : Proportion de NBD par niveau de connaissance des marchés financiers	37
Tableau 12 : Proportion de NBD par sexe.....	37
Tableau 13 : Proportion de NBD par groupe d'âge	38

Liste des graphiques et illustrations

Figure 1 : Proportion des jours d'achats où 2 actions sont achetées.....	19
Figure 2 : Proportion des jours d'achats où 3 actions sont achetées.....	19
Figure 3 : Proportion des jours d'achats où 4 actions sont achetées.....	20
Figure 4 : Proportion des jours d'achats où 5 actions sont achetées.....	20
Figure 5 : Graphique des rendements annuels moyens	29
Figure 6 : Graphique des distributions de rendements	30
Figure 7 : Résultats de Gathergood & al. hypothèse 3	31
Figure 8 : Variables significatives impactant l'utilisation de la NBD modèle 2.....	36

Introduction

La diversification est essentielle pour la gestion de portefeuille, permettant de gérer le risque et de maximiser les rendements. Introduite par Harry Markowitz en 1952, la théorie moderne du portefeuille met en avant l'importance de sélectionner des actifs avec des covariances faibles ou négatives. Cependant, en pratique, les investisseurs s'écartent souvent de cette théorie en raison de biais comportementaux.

La finance comportementale révèle que les investisseurs ne sont pas toujours rationnels et sont influencés par des biais comme le cadrage étroit, où les décisions d'investissement sont prises sans considérer l'impact global sur le portefeuille. Cela mène souvent à des stratégies de diversification sous-optimales, telles que la Naïve Buying Diversification (NBD), où les investisseurs répartissent leurs investissements de manière égale entre les actifs achetés, sans tenir compte des actifs déjà présents dans leur portefeuille.

Ce mémoire vise à reproduire l'étude de Gathergood et al. (2023) sur le phénomène de NBD en se concentrant sur les investisseurs belges. En comparant les résultats belges à ceux obtenus au Royaume-Uni, je cherche à comprendre comment les investisseurs belges adoptent des stratégies de diversification naïves et l'impact de ces stratégies sur leurs portefeuilles.

Le mémoire est structuré comme suit : une introduction, une revue de la littérature sur la diversification et les biais comportementaux, la formulation des hypothèses de recherche, une description de la base de données des investisseurs belges, les tests des hypothèses analysant la prévalence de la NBD et son impact, une conclusion résumant les résultats et proposant des pistes pour des recherches futures.

I. Revue de littérature

La diversification est un élément clé au sein des décisions d'investissement, car elle offre un moyen essentiel de gérer le risque et de maximiser le rendement. C'est en 1952 que Harry Markowitz a introduit la théorie moderne du portefeuille dans son article révolutionnaire « Portfolio selection » (Markowitz, 1952). Dans cet ouvrage fondateur, Markowitz définit la diversification comme étant la sélection d'un portefeuille d'actifs qui ont des covariances faibles ou négatives. Il explique que la diversification ne dépend pas seulement du nombre de titres détenus, mais plutôt de la sélection de titres qui ont des covariances faibles ou négatives. Cela permet de réduire le risque global du portefeuille tout en maximisant les rendements attendus

La théorie moderne du portefeuille, telle que développée par Harry Markowitz en 1952, offre une structure pour élaborer et choisir des portefeuilles en tenant compte des rendements anticipés des investissements et de la tolérance au risque de l'investisseur (Fabozzi et al., 2002).

La finance traditionnelle qui suppose que les individus se comportent de manière rationnelle et que les investisseurs traitent les nouvelles informations rapidement et correctement s'oppose à la finance comportementale, qui indique que les individus ne sont pas des êtres parfaitement rationnels, qu'ils souffrent d'une série de biais comportementaux et que le problème de ces biais, c'est qu'ils génèrent des comportements systématiques chez les individus qui vont les amener à prendre des décisions irrationnelles. De plus, afin de trouver une solution avec le moins d'efforts et le plus rapidement possible, les individus ont tendance à s'appuyer sur des heuristiques ou encore sur des processus mentaux, afin de simplifier les informations qu'ils traitent (Tversky & Kahneman, 1974).

Les investisseurs dévient donc des recommandations proposées par Markowitz car ces derniers « prennent leurs décisions d'investissement sous l'influence d'une combinaison de biais comportementaux [...] Sous l'influence de certains de ces biais ou de leur combinaison, les investisseurs individuels prennent souvent des décisions d'investissement irrationnelles. Par conséquent, les investisseurs individuels, dans l'ensemble, obtiennent des rendements médiocres à long terme » (Jain et al., 2015).

Parmi ces biais et heuristiques, que l'on peut par exemple retrouver dans les articles « A survey of behavioral finance » (Barberis & Thaler, 2002), ainsi que dans l'article « Judgment under Uncertainty: Heuristics and Biases » (Tversky & Kahneman, 1974), le plus pertinent pour mon étude est le suivant :

- **Le cadrage étroit** ou encore **Narrow framing**, est un biais cognitif qui résulte d'une évaluation intuitive de la situation, où l'individu adopte souvent le cadre le plus facilement disponible, généralement étroit. Cela se traduit par la tendance à choisir des investissements de manière isolée plutôt que de considérer l'impact global sur le portefeuille. Des facteurs tels que l'influence de l'endroit de résidence ou des biais comme la familiarité ou la localité peuvent conduire à une détention préférentielle d'actions locales ou bien connues géographiquement et linguistiquement. Cette préférence peut entraîner une sous-diversification, exposant ainsi les investisseurs à une volatilité accrue et à un effet de disposition, où ils sont plus enclins à vendre des actions gagnantes par rapport aux actions perdantes (Jain et al., 2015). Un aspect peu étudié du cadrage étroit est la valeur de synchroniser les achats récents d'un actif avec la détention actuelle de cet actif par l'investisseur, soulignant ainsi l'importance de se rappeler un niveau de détention cible spécifique (Gathergood et al., 2023).

Lorsqu'on regarde les décisions d'investissement, il devient évident que les biais comportementaux ont un impact significatif et s'écartent souvent des recommandations de la théorie moderne du portefeuille de Markowitz. Des biais tels que l'effet de disposition, le cadrage étroit, le biais de disponibilité, et d'autres, peuvent conduire à des décisions irrationnelles et compromettre la mise en œuvre de la diversification optimale. Dans ce contexte, des alternatives simplifiées et intuitives telles que la Naïve Buying Diversification (NBD) et la Naïve Portfolio Diversification (NPD) voient le jour.

La diversification naïve, autrement connue sous le nom de règle du $1/N$, est une approche qui consiste à investir une somme équitable dans chaque actif disponible parmi le total de N actifs distincts. Cette stratégie d'investissement ne tient pas compte des données historiques et attribue des pondérations de portefeuille invariantes dans le temps (Hsu et al., 2018). Cette approche intuitive a l'avantage de sa simplicité, ce qui permet aux investisseurs de créer des portefeuilles sans se soucier des corrélations entre les actifs.

Plus précisément, la NBD se caractérise par une approche d'investissement où les décisions des investisseurs sont prises sans considérer les corrélations entre les actifs individuels. Cette pratique découle de la propension des investisseurs à diversifier simplement les actifs acquis au fil du temps, sans égard à la composition globale de leur portefeuille existant. Ils appliquent souvent des heuristiques simplifiées, telles que la règle $1/N$, attribuant un poids d'achat égal à

chaque actif dans un nouveau portefeuille. Cette stratégie de diversification simple repose sur des décisions intuitives et influencées par des biais comportementaux, tels qu'un excès de confiance ou une préférence pour des solutions plus simples. À la différence de la diversification optimale, qui prend en compte les corrélations entre les actifs pour minimiser le risque global du portefeuille, la NBD se concentre sur l'allocation des investissements au moment de l'achat, sans considération pour la structure globale du portefeuille.

Selon l'article « Naïve Buying Diversification and Narrow Framing by Individual Investors » de John Gathergood (Gathergood et al., 2023), deux hypothèses expliquent l'utilisation de la NBD par les investisseurs.

La première, l'hypothèse de la sélection de titres, propose que les investisseurs adoptent la NBD en privilégiant des actions attrayantes pour des motifs indépendants de la diversification, plutôt qu'en ayant pour objectif explicite la diversification de leur portefeuille. Dans cette approche, l'allocation NBD repose sur une heuristique simple de pondération égale, sans motivation sous-jacente visant à délibérément accroître le nombre d'actions détenues dans le portefeuille en vue de réduire le risque global. Au sein de cette hypothèse, les actions ayant des profils de risque et de rendements similaires auront tendance à être retenues par les individus, en particulier celles ayant récemment généré des rendements élevés ou celles ayant retenu l'attention des médias. Ce dernier point est soutenu par Brad Barber et Terrance Odéan dans leur article intitulé « All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors », qui démontrent que « les investisseurs individuels sont des acheteurs nets d'actions attractives, par exemple, des actions faisant l'objet de l'actualité, des actions connaissant un volume de transactions anormalement élevé, et des actions affichant des rendements exceptionnels en une journée. L'achat motivé par l'attention découle de la difficulté qu'ont les investisseurs à rechercher parmi les milliers d'actions qu'ils pourraient potentiellement acheter » (Barber & Odean, 2008).

La seconde hypothèse, le motif de diversification, suggère que les investisseurs font un choix actif de se diversifier en utilisant la NBD. Dans ce cas, l'investisseur, qui aurait peut-être acheté seulement quelques actions, choisit délibérément de se diversifier en ajoutant des actions supplémentaires à son panier d'achat. Contrairement à l'hypothèse de la sélection de titres, les investisseurs ayant un motif de diversification cherchent à exploiter les avantages en termes de risque liés à la présence d'un plus grand nombre d'actions dans le portefeuille.

La Naïve Portfolio diversification quant à elle, repose sur une répartition égale des investissements sur tous les actifs, évitant ainsi les calculs de corrélation complexes. Cette approche se caractérise par une pondération égale des titres dans le portefeuille d'un investisseur. Les décisions d'allocation d'actifs sont simplifiées sur la base de la règle intuitive $1/N$, où chaque actif reçoit une part équitable. Cette stratégie, influencée par des biais comportementaux, offre une alternative accessible à la diversification optimale.

L'objectif de ce mémoire est de reproduire l'étude réalisée par Gathergood et al. sur le phénomène de "Naïve Buying Diversification" (NBD) chez les investisseurs individuels, mais en se concentrant sur un échantillon d'investisseurs belges. Cette recherche vise à comprendre comment les investisseurs belges adoptent des stratégies similaires ou différentes par rapport à leurs homologues d'autres régions.

a) Résumé de l'article :

L'article de Gathergood et al. intitulé « Naïve buying diversification and Narrow framing by individual investors » (Gathergood et al., 2023), examine la manière dont les investisseurs individuels abordent l'investissement multi-actions, en utilisant souvent des heuristiques simples et imparfaites pour guider leurs décisions. Contrairement aux recommandations de la théorie du portefeuille qui préconisent une diversification optimale des actions (Markowitz, 1952), les investisseurs individuels s'écartent souvent de ces dernières. Cet écart peut souvent être attribué à un phénomène appelé « cadrage étroit », dans lequel les investisseurs évaluent les décisions de manière isolée sans tenir compte des autres actifs disponibles ou du portefeuille dans son ensemble.

Cet article examine spécifiquement une heuristique appelée Naive Buying Diversification (NBD), dans laquelle les investisseurs répartissent souvent leurs investissements de manière égale entre plusieurs actions achetées en une seule journée. Les investisseurs utilisent cette heuristique principalement pour de nouveaux achats plutôt que pour diversifier l'ensemble de leur portefeuille. Cette tendance NBD est désavantageuse car elle ne permet pas une diversification optimale du portefeuille.

La recherche montre que si de nombreux investisseurs adoptent la NBD, rares sont ceux qui adoptent une diversification naïve de portefeuille (NPD), où les actions sont également pondérées dans l'ensemble du portefeuille. Cette divergence entre NBD et NPD suggère que

les investisseurs se concentrent sur les achats individuels plutôt que sur la composition globale du portefeuille.

Les résultats montrent que la NBD entraîne généralement des rendements inférieurs et met en évidence les défis auxquels les investisseurs particuliers sont confrontés lorsqu'ils prennent des décisions financières.

Enfin, cet article met en évidence le comportement sous-optimal des investisseurs individuels et leur tendance à adopter des stratégies de diversification simples qui peuvent nuire à la performance globale du portefeuille. Ces résultats soulignent l'importance pour les investisseurs de reconsidérer leurs approches d'investissement et de rechercher des stratégies plus efficaces pour obtenir une diversification optimale et maximiser les rendements de leur portefeuille.

b) Questions posées par les auteurs :

L'étude réalisée par Gathergood et al. explore comment les investisseurs individuels prennent des décisions d'investissement en répartissant leurs fonds entre plusieurs actions achetées le même jour. Les auteurs se demandent si les investisseurs essaient d'équilibrer leurs portefeuilles globaux ou simplement les valeurs des achats de plusieurs actifs effectués le même jour.

L'utilisation d'heuristiques simplistes peut conduire à une prise de risque excessive et affecter la stabilité des marchés financiers. De plus, comprendre le comportement d'investissement individuel peut soutenir les efforts visant à améliorer la littératie financière et à promouvoir des décisions d'investissement plus éclairées.

Pour répondre à ces questions, les auteurs ont analysé les choix de trading des investisseurs individuels fournis par Barclays Stockbroking au Royaume-Uni. Ils ont examiné les achats d'actions regroupés sur la même journée pour déterminer si les investisseurs cherchaient à équilibrer l'ensemble de leurs portefeuilles ou simplement le montant des achats de plusieurs actifs effectués le même jour. Les auteurs ont également étudié les heuristiques utilisées par les investisseurs lorsqu'ils prennent des décisions financières, en mettant particulièrement l'accent sur l'utilisation de la règle 1/N.

c) Données à leur disposition :

L'étude utilise les données de Barclays Wealth, l'un des principaux courtiers britanniques, contenant l'historique des transactions de 182 569 comptes d'avril 2012 à juin 2016. Après avoir

éliminé les comptes sans activité pendant la période de données, l'échantillon de base se compose de 118 169 comptes ayant effectué au moins une transaction d'achat dans cette période de 4 ans. Les données comprennent entre autre « les numéros d'identification des actions (numéros de la Stock Exchange Daily Official List (SEDOL), une liste d'identificateurs de titres utilisés au Royaume-Uni et en Irlande à des fins de compensation), les dates des transactions, les types de transactions (par exemple, achat, vente), les quantités de transactions et les prix des transactions » (Gathergood et al., 2023), ce qui a permis d'analyser le comportement d'investissement des clients de Barclays Wealth. Ces données ont fourni une base solide pour examiner comment les investisseurs individuels répartissent leurs fonds entre plusieurs actions achetées le même jour, ainsi que pour étudier l'utilisation des heuristiques telles que la règle 1/N dans le processus de prise de décision des investisseurs.

d) Echantillon :

Dans cette étude, plusieurs échantillons des données du Barclays Wealth ont été utilisés pour analyser le comportement d'investissement des clients du courtier. Dans un premier temps, un échantillon de tous les comptes a été constitué pour observer l'ensemble des achats d'actions effectués pendant la période étudiée. Des sous-échantillons ont ensuite été établis en fonction de l'intervalle entre les achats d'actions. L'analyse s'est concentrée sur un échantillon de jours d'achat de plusieurs actions, définis comme des jours où l'investisseur achète deux actions ou plus en une seule journée. De plus, un sous-échantillon comprenant tous les nouveaux comptes ouverts pendant la période de données a été constitué, permettant d'examiner les comportements d'investissement spécifiques à ces nouveaux comptes et de les comparer aux comptes existants. Ces échantillons ont permis aux auteurs d'étudier en détail comment les investisseurs individuels répartissent leurs fonds entre plusieurs actions achetées le même jour, ainsi que de comparer les modèles d'investissement entre les comptes existants et les nouveaux comptes.

e) Hypothèses

Dans cette section, j'examine les hypothèses formulées et testées par Gathergood et al. dans leur étude sur la diversification naïve des achats par les investisseurs individuels. Les auteurs ont tout d'abord centré leur analyse sur le comportement des investisseurs lors des achats de plusieurs actions, en particulier en ce qui concerne l'allocation des fonds entre ces actions. Leur hypothèse est basée sur l'idée que les investisseurs peuvent adopter une approche simpliste, appelée diversification naïve des achats (NBD), en allouant des fonds de manière équitable entre les actions achetées. Ainsi, ils ont formulé l'hypothèse suivante :

- Les investisseurs adoptent la diversification naïve des achats (NBD) lors des achats de plusieurs actions, en allouant généralement des fonds de manière équitable entre les actions.

Les auteurs se sont ensuite concentrés sur la distinction entre la diversification naïve des achats (NBD) et la diversification naïve du portefeuille (NPD), en analysant comment les investisseurs allouent leurs fonds lors des jours d'achat multiples. Leur hypothèse est basée sur l'idée que les investisseurs peuvent adopter l'une ou l'autre de ces approches pour équilibrer leur portefeuille. Ainsi, ils ont formulé l'hypothèse suivante :

- Les investisseurs peuvent opter pour la diversification naïve des achats (NBD) ou la diversification naïve du portefeuille (NPD) lors des jours d'achat multiples, en fonction de leur approche d'allocation des fonds.

Ils ont également analysé deux hypothèses concurrentes concernant l'utilisation de l'heuristique NBD (Diversification Naïve des Achats), qu'ils désignent sous les termes de "stock-picking NBD" et "diversification-motivated NBD".

- Hypothèse du stock-picking NBD : les investisseurs choisissent plusieurs actions attractives, souvent en raison de bonnes performances récentes ou de nouvelles positives, et investissent un montant égal dans chaque action en utilisant une heuristique naïve de pondération égale, sans considérer les avantages de diversification du risque liés à l'inclusion de plusieurs actions dans le portefeuille.
- Hypothèse de la diversification motivée par NBD : les investisseurs, qui pourraient n'avoir acheté que quelques actions motivés par l'avantage de réduction du risque de la diversification, achètent des actions supplémentaires. Ce faisant, ils perçoivent des avantages de diversification, bien qu'utilisant des pondérations égales.
- Sous la première hypothèse, la NBD est le résultat d'une attention limitée à la diversification ; sous la seconde hypothèse, la NBD est motivée par la diversification et constitue une amélioration par rapport à un contre-factuel dans lequel l'investisseur aurait pris une décision d'achat moins diversifiée.

Ensuite, ils ont également évalué l'impact de l'utilisation de l'heuristique NBD (Diversification Naïve des Achats) sur la performance du portefeuille, en comparaison avec une stratégie NPD (Diversification Naïve du Portefeuille). Ils explorent également la possibilité que des niveaux élevés de risque idiosyncratique affectent la propension des investisseurs à utiliser la NBD.

- Impact de la NBD sur la Performance du Portefeuille : les auteurs se demandent si l'utilisation fréquente de la NBD par les investisseurs nuit à la performance du portefeuille. Ils comparent les rendements des investisseurs NBD avec ceux des investisseurs non-NBD pour évaluer la performance du portefeuille.
- Relation entre le Risque Idiosyncratique et l'Utilisation de la NBD : ils examinent si des niveaux élevés de risque idiosyncratique sont associés à une moindre propension des investisseurs à utiliser la NBD, ce qui pourrait indiquer une sensibilité à la faiblesse de la NBD en présence d'un risque idiosyncratique élevé.

Enfin, les auteurs examinent la possibilité que l'utilisation de la stratégie NBD soit motivée par la simplification du problème de décision. Ils analysent comment les investisseurs choisissent le montant total à investir sur une journée donnée et le nombre d'actions achetées (N), en recherchant des indications d'une préférence pour la simplicité.

f) Méthodologie et tests réalisés :

Les auteurs de l'étude sur le comportement d'investissement des clients de Barclays Wealth ont adopté une approche méthodologique rigoureuse. Ils ont analysé les données transactionnelles de 182 569 comptes sur une période allant d'avril 2012 à juin 2016, en se concentrant sur les achats multiples d'actions effectués en une seule journée. L'étude a examiné la répartition des fonds investis entre ces actions, en utilisant l'heuristique 1/N pour évaluer si les investisseurs répartissaient équitablement leurs investissements. Des modèles de régression, y compris un modèle linéaire, ont été utilisés pour évaluer l'allocation des fonds lors des achats de deux actions conformément à la NBD.

Les variables de contrôle telles que le sexe, l'âge, la taille et les caractéristiques du portefeuille ont été incluses pour tenir compte d'autres facteurs influençant les décisions d'investissement. De plus, des tests de similarité des actions basés sur les rendements ont été réalisés pour évaluer la répartition des investissements entre des actions similaires.

Les auteurs ont examiné de près la diversification naïve des achats, y compris son utilisation dans d'autres types d'investissements comme les fonds communs de placement, ainsi que d'autres types d'allocations 1/N sur différentes périodes. Ils ont également exploré si les investisseurs appliquaient l'heuristique 1/N à l'ensemble de leur portefeuille ou uniquement aux nouveaux achats, et pourquoi ils utilisaient cette heuristique, cherchant à simplifier leur processus de prise de décision ou à atteindre une forme de diversification.

De plus, l'impact de l'utilisation de l'heuristique NBD sur la performance du portefeuille a été examiné, de même que son utilisation lors de la vente d'actions et le choix conjoint du montant d'investissement et de l'heuristique N.

En résumé, cette approche méthodologique exhaustive a permis aux auteurs de comprendre les comportements d'investissement des clients de Barclays Wealth, en mettant en lumière leurs stratégies de répartition des fonds et les heuristiques utilisées dans leurs décisions d'investissement.

g) Résultats et conclusion :

Les auteurs ont observé que les investisseurs ont tendance à appliquer l'heuristique NBD lorsqu'ils achètent plusieurs actions en une seule journée ou sur plusieurs semaines ou mois. La propension à utiliser l'heuristique NBD diminue avec l'expérience de l'investisseur et les enjeux financiers, mais même avec une expérience élevée et des enjeux importants, plus de 20% des investissements sont effectués avec NBD. Les investisseurs semblent être attirés par la simplicité offerte par la stratégie NBD.

Les auteurs constatent également que les investisseurs utilisent principalement l'heuristique NBD pour leurs nouveaux achats, plutôt que pour l'ensemble de leur portefeuille. Les investisseurs ont tendance à équilibrer leur portefeuille en achetant des actions qui sont sous-représentées dans leur portefeuille.

L'étude a également montré que les investisseurs qui n'utilisent jamais la NBD obtiennent des rendements annuels environ un tiers plus élevés que ceux qui l'adoptent. De plus, les investisseurs ont tendance à souvent appliquer l'heuristique 1/N lors de la vente d'actions, en répartissant les ventes de manière égale plutôt qu'en considérant les différences de risque ou de rendement.

Enfin, ils ont étudié les raisons pour lesquelles les investisseurs utilisaient l'heuristique NBD et se sont rendus compte qu'ils l'utilisaient pour simplifier le processus de prise de décision et pour atteindre un certain niveau de diversification.

II. Hypothèses

L'objectif de ce mémoire est de reproduire l'analyse effectuée par Gathergood et al. (Gathergood et al., 2023), sur un échantillon d'investisseurs belges afin de comparer les résultats et les conclusions obtenus. Dans leur article, Gathergood et al. mettent en lumière les tendances et les conséquences de l'utilisation par les investisseurs de la diversification naïve à l'achat (NBD). Malgré une expérience et des enjeux financiers élevés, les investisseurs ont largement recours à la NBD, avec plus de 20 % de leurs investissements effectués selon cette stratégie. Cependant, cette approche présente des risques et des inconvénients significatifs.

Tout d'abord, l'utilisation généralisée de la NBD est souvent accompagnée d'une négligence de la diversification du portefeuille, ce qui peut nuire aux rendements des investisseurs. En moyenne, ceux qui adoptent la NBD affichent des performances inférieures à ceux qui préfèrent d'autres stratégies plus diversifiées. Ensuite, la NBD peut introduire des distorsions dans les pondérations du portefeuille, accroissant le risque global, notamment lorsque le risque idiosyncratique est élevé. Malgré ces risques, les investisseurs ne semblent pas réduire leur utilisation de la NBD pour atténuer ces dangers potentiels. En résumé, l'article souligne que l'utilisation répandue de la diversification naïve à l'achat peut entraîner des conséquences néfastes pour les investisseurs, compromettant la diversification de leurs portefeuilles et augmentant leur vulnérabilité aux risques financiers.

Certaines études, dont celle de DeMiguel & al. (DeMiguel et al., 2009), ont mis en évidence l'efficacité d'une approche de diversification naïve simple, connue sous le nom de règle 1/N, dans la construction de portefeuilles, consistant à allouer également les fonds à toutes les classes d'actifs disponibles, sans tenir compte des prévisions de rendement ou de volatilité. D'autres études, telles que celles de Kritzman (Kritzman et al., 2010), Kirby et Ostdiek (Kirby & Ostdiek, 2012), ainsi que de Tu & Zou (Tu & Zhou, 2011), remettent en question la performance de la méthode 1/N par rapport aux portefeuilles optimisés.¹

En outre, comme le montre l'article de Gathergood et al., les investisseurs qui achètent plusieurs actions le même jour ont tendance à utiliser la règle 1/N et à allouer le même poids à chaque action. L'article intitulé «Irrational diversification : an examination of individual portfolio choice» (Baltussen & Post, 2011) confirme cela, en expliquant que « Les sujets ont tendance à

¹ Annexe 1 : Résumé de la discussion dans la littérature des avantages et inconvénients de la diversification naïve du portefeuille

se concentrer sur la distribution marginale² d'un actif, tout en ignorant largement ses avantages en matière de diversification ».

Ainsi, en prenant en compte les constatations de Gathergood et al. et celles présentes dans la littérature existante, voici les hypothèses que je vais tester :

Hypothèse 1 : Comportement des investisseurs lors des achats de plusieurs actions

- H0 : Lors des jours d'achat de plusieurs actions, les investisseurs ont tendance à utiliser l'heuristique $1/N$ et donc à dépenser environ $1/N$ de leur argent dans chaque action.
- H1 : L'allocation des investisseurs lors des jours d'achat de plusieurs actions ne suit pas une répartition équitable des fonds entre les actions, déviant significativement de l'heuristique $1/N$.

Hypothèse 2 : Prévalence de la NBD ou NPD lors des jours d'achat multiples

- H0 : Lors des jours d'achat multiples, les investisseurs utilisent principalement la diversification naïve des achats (NBD) pour allouer leurs fonds entre les actions.
- H1 : Lors des jours d'achat multiples, les investisseurs utilisent principalement la diversification naïve du portefeuille (NPD) pour équilibrer leur portefeuille en fonction du nombre d'actions détenues.

Hypothèse 3 : Impact de la Diversification Naïve des Achats sur la Performance du Portefeuille

- H0 : Il n'y a pas de différence significative entre les rendements annuels des investisseurs utilisant la NBD et ceux ne l'utilisant pas.
- H1 : Il existe une différence significative entre les rendements annuels des investisseurs utilisant la NBD et ceux ne l'utilisant pas.

Hypothèse 4 : Impact de Diverses Variables sur l'Utilisation de la Naive Buying Diversification (NBD)

- H0 : Les variables démographiques et financières n'ont pas d'impact significatif sur la probabilité d'utiliser la stratégie de diversification naïve à l'achat (NBD).
- H1 : Les variables démographiques et financières ont un impact significatif sur la probabilité d'utiliser la stratégie de diversification naïve à l'achat (NBD).

²Distribution marginale : attribuer des poids égaux à chaque actif sans évaluer leur rendement attendu, leur risque ou leur corrélation avec d'autres actifs

III. Données

Concernant la base de données que je vais utiliser, j'ai reçu de ma promotrice une base de données d'investisseurs belges, à la suite de la signature d'un accord de confidentialité. Cette base de données se divise en 6 fichiers :

1. **Ptf Returns:** Contient des informations sur les rendements des portefeuilles. Les colonnes incluent **time**, **user** (Numéro de l'investisseur), **g_return** (Gross return avant les coûts de transaction), et **n_return** (Net return après les coûts de transaction).
2. **Sample Investors Clean:** Ce fichier contient de nombreuses informations détaillées sur les investisseurs. Les colonnes incluent:
 - **User** (Numéro de l'investisseur)
 - **DateBirth** (Année de naissance)
 - **Sex** (Sexe de l'investisseur)
 - **Languageld** (Langage de préférence)
 - **AP_Q3** (Plus haut degré d'éducation): 0 (Aucun), 1 (Secondaire), 2 (Université ou équivalent)
 - **SP_Q0** (Niveau de vie actuel): 1 (Retraite), 2 (Préparation à la retraite), 3 (Jeune famille), 4 (Un couple sans enfant), 5 (Célibataire avec peu d'obligations financières), 6 (Famille mature)
 - **SP_Q1** (Rendement attendu des investissements): 1 (Un rendement sans aucun risque de perte en capital), 2 (Le taux d'inflation a augmenté de 2 à 4 % par an), 3 (Le taux d'inflation a augmenté de 5 à 7 % par an), 4 (Le taux d'inflation a augmenté de 8 à 12 % par an), 5 (Le taux d'inflation a augmenté de plus de 12 % par an)
 - **SP_Q2** (Réaction à une baisse de 20% de l'investissement en 6 mois): 1 (Un désastre total !), 2 (Vous vendrez l'investissement avec perte), 3 (Vous êtes anxieux mais ne bougerez pas), 4 (C'était un risque calculé), 5 (Vous profitez de la situation du marché)
 - **SP_Q3** (Niveau de connaissance des marchés financiers): 1 (Très peu de connaissances), 2 (Non familier mais intéressé), 3 (Suffisamment d'expérience pour diversifier), 4 (Bonne connaissance des marchés financiers), 5 (Investisseur expérimenté)

- **SP_Q4** (Objectifs d'investissement): 1 (Obtenir un revenu stable), 2 (Objectifs spécifiques dans les 5 ans), 3 (Préparation à la retraite), 4 (Somme importante à investir), 5 (Investir de manière équilibrée sur le long terme), 6 (Horizon d'investissement supérieur à 3 ans)
 - **SP_Q5** (Durée pendant laquelle l'investisseur peut se passer du montant investi): 1 (Moins de 2 ans), 2 (Entre 2 et 3 ans), 3 (Entre 3 et 5 ans), 4 (Entre 5 et 7 ans), 5 (Plus de 7 ans)
 - **SP_Q6** (Propriétaire immobilier): 0 (Non), 5 (Oui)
 - **SP_Q7** (Revenu net annuel total): 0 (Moins de 20 000 €), 1 (20 000 € à 40 000 €), 2 (40 000 € à 75 000 €), 3 (75 000 € à 150 000 €), 4 (150 000 € et plus)
 - **SP_Q8** (Pourcentage du revenu net épargné): 0 (0 % à 10 %), 1 (10 % à 20 %), 2 (20 % à 30 %), 3 (30 % à 40 %), 4 (Plus de 40 %)
 - **SP_Q9** (Montant investi dans les instruments financiers): 0 (Moins de 20 000 €), 1 (20 000 € à 50 000 €), 2 (50 000 € à 250 000 €), 3 (250 000 € à 1 000 000 €), 4 (Plus de 1 000 000 €)
 - **SP_Q10** (Niveau de connaissance et d'expérience des risques): 0 (Aucune connaissance), 1 (Connaissance moyenne), 2 (Bonne connaissance)
3. **Stat Instruments:** Contient des informations statistiques sur les investisseurs. Les colonnes incluent **user** (Numéro de l'investisseur), **N_W** (nombre de transactions de warrants), **N_B** (nombre de transactions d'obligations), **N_F** (nombre de transactions de fonds), **N_O** (nombre de transactions d'options).
 4. **Stock Holdings:** Détails des positions détenues par les investisseurs. Les colonnes incluent **positions** (Nombre de positions dans une action), **ISIN** (Code ISIN de l'action), **Time**, **tnof** (Description de l'action), **MV_pos** (Valeur de marché de la position), et **user** (Numéro de l'investisseur).
 5. **Stock Ptf:** Informations sur les portefeuilles des investisseurs. Les colonnes incluent **time**, **MV_ptf** (Valeur de marché du portefeuille), **NB_ISIN** (Nombre d'ISIN différents), et **User** (Numéro de l'investisseur).
 6. **Stock Trades:** Détails des transactions d'actions effectuées par les investisseurs. Les colonnes incluent **Jour** (Jour de la transaction), **ISIN** (Code ISIN de l'action), **User** (Numéro de l'investisseur), **TransType** (Type de transaction (buy ou sell)), **ExecTime**

(Moment de l'exécution de la transaction), **TradeCurrency** (Devise de la transaction), **Cost** (Coût de la transaction), **StopPrice** (Prix limite), **tnof** (Description de l'action), **ExecPrice** (Prix de l'exécution), **ExecQty** (Quantité de l'exécution), **ExecValue** (Valeur de l'exécution), **ExecPrice_EUR** (Prix de l'exécution en euros), **ExecValue_EUR** (Valeur de l'exécution en euros), et **Cost_EUR** (Coût de l'exécution en euros).

IV. Tests des hypothèses

Hypothèse 1 :

L'Hypothèse 1 examine si, lors des jours d'achat de plusieurs actions, les investisseurs suivent l'heuristique $1/N$, c'est-à-dire s'ils allouent environ $1/N$ de leurs fonds à chaque action.

- **H0 :** Lors des jours d'achat de plusieurs actions, les investisseurs ont tendance à utiliser l'heuristique $1/N$ et donc à dépenser environ $1/N$ de leur argent dans chaque action.
- **H1 :** L'allocation des investisseurs lors des jours d'achat de plusieurs actions ne suit pas une répartition équitable des fonds entre les actions, déviant significativement de l'heuristique $1/N$.

Pour tester l'hypothèse selon laquelle les investisseurs adoptent une stratégie de diversification naïve des achats (NBD), j'ai effectué une analyse en utilisant les données de transactions d'actions.

Préparation des données

La première étape a consisté à charger et préparer les données. J'ai filtré les transactions pour inclure uniquement les jours où un investisseur a acheté plus d'une action. Cette sélection permet de se concentrer sur les comportements de diversification des investisseurs. En parallèle, une analyse de la fréquence des jours d'achats uniques versus multiples révèle que les jours d'achats uniques sont légèrement plus fréquents (2387 jours uniques) par rapport aux jours d'achats multiples (2381 jours multiples).

Pour chaque jour où plusieurs actions ont été achetées, j'ai effectué les calculs suivants :

1. **Proportion de l'investissement total alloué à chaque action :** Cette proportion est calculée en divisant la valeur d'exécution de chaque action (ExecValue_EUR) par la valeur totale investie ce jour-là (total_value). Ceci est stocké dans « share_of_investment ».
2. **Allocation égale théorique ($1/N$) :** Pour chaque jour d'achat multiple, j'ai également calculé une allocation égale théorique. Cette allocation est définie comme $1/N$, où N est le nombre total d'actions achetées ce jour-là. Ceci est stocké dans « equal_share ». L'idée est que si un investisseur suit une stratégie NBD, il répartirait ses investissements également entre les différentes actions achetées ce jour-là. Par exemple, s'il achète 3

actions, on s'attend à ce que chaque action reçoive environ 33.33% (1/3) de l'investissement total.

Création des résumés de tableau

Pour analyser la proximité de l'allocation réelle à l'allocation égale théorique, j'ai créé des tableaux résumant la proportion de jours d'achat où les allocations sont proches de 1/N pour différentes tolérances ($\pm 2\%$, $\pm 5\%$, $\pm 10\%$). Ces tolérances ont été choisies pour examiner comment la précision de l'allocation égalitaire varie. Une tolérance plus large permet de voir si les investisseurs suivent encore la stratégie NBD de manière moins stricte. Cette méthode est également utilisée par Gathergood et al., permettant ainsi une comparaison directe des résultats.

Les tableaux obtenus pour les différentes tolérances sont les suivants :

Tolérance $\pm 2\%$:

N	1/N Buying (%)	Buy Days
2	18.93116	2374
3	12.50662	2106
4	19.46751	1565
5	20.68075	1026

Tableau 1 : Hypothèse 1; tolérance $\pm 2\%$

Tolérance $\pm 5\%$:

N	1/N Buying (%)	Buy Days
2	26.04652	2374
3	25.64500	2106
4	28.19926	1565
5	30.17606	1026

Tableau 2 : Hypothèse 1; tolérance $\pm 5\%$

Tolérance $\pm 10\%$:

N	1/N Buying (%)	Buy Days
2	40.92585	2374
3	40.56518	2106
4	48.56857	1565
5	48.60329	1026

Tableau 3 : Hypothèse 1; tolérance $\pm 10\%$

Définition et Calcul du "1/N Buying (%)"

Le "1/N Buying (%)" représente le pourcentage de jours d'achat multiples où les investisseurs répartissent leur investissement total de manière égale parmi les actions achetées ce jour-là, avec une tolérance de $\pm 2\%$, $\pm 5\%$ ou $\pm 10\%$.

Par exemple, pour $N=2$ et une tolérance $\pm 2\%$; si un investisseur achète deux actions un jour donné, "1/N Buying (%)" indique le pourcentage de ces jours où l'investissement est réparti presque également (entre 49% et 51% pour chaque action). Donc, si ce pourcentage est de 18.93%, cela signifie que dans 18.93% des jours où deux actions sont achetées, les investisseurs ont divisé leur investissement de manière égale dans la tolérance de $\pm 2\%$.

Pour une tolérance de $\pm 10\%$, "1/N Buying (%)" indique le pourcentage de jours où l'investissement est réparti presque également entre 45% et 55% pour chaque action. Donc, si ce pourcentage est de 40.93%, cela signifie que dans 40.93% des jours où deux actions sont achetées, les investisseurs ont divisé leur investissement de manière égale dans la tolérance de $\pm 10\%$.

Visualisation des résultats

Les graphiques ci-dessous montrent la distribution des proportions d'investissement pour différents nombres d'actions achetées. Les histogrammes montrent la proportion des jours d'achat et la ligne rouge pointillée représente l'allocation égale théorique 1/N. Les pics autour de cette ligne indiquent une tendance à répartir les investissements de manière égale entre les actions achetées.

De manière plus précise, la "Proportion de Buy Days" dans les graphiques représente le pourcentage de jours d'achat multiples où une certaine proportion de l'investissement est observée pour une action spécifique. Par exemple, pour $N=2$, une proportion de 7.5036% à 0.5 (50%) indique que dans 7.5036% des jours où deux actions sont achetées, les investisseurs répartissent leur investissement à 50%-50%.

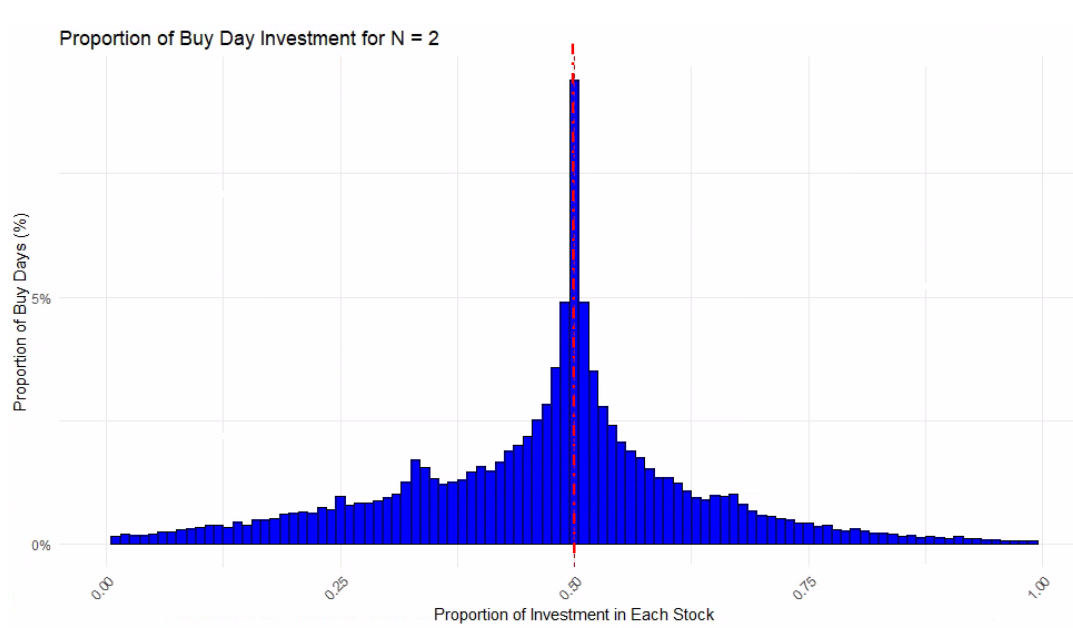


Figure 1 : Proportion des jours d'achats où 2 actions sont achetées

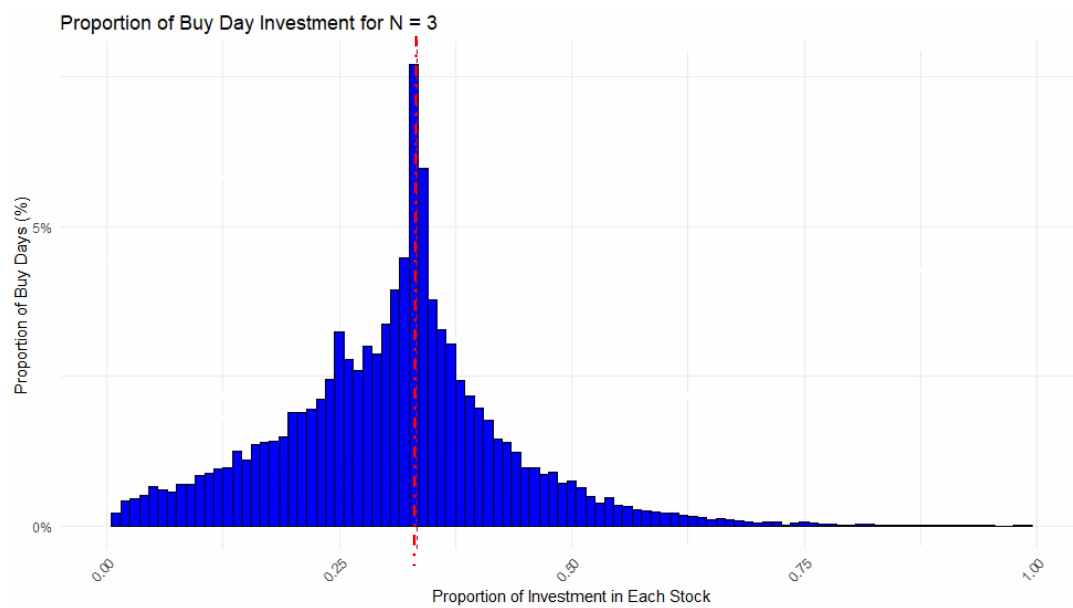


Figure 2 : Proportion des jours d'achats où 3 actions sont achetées

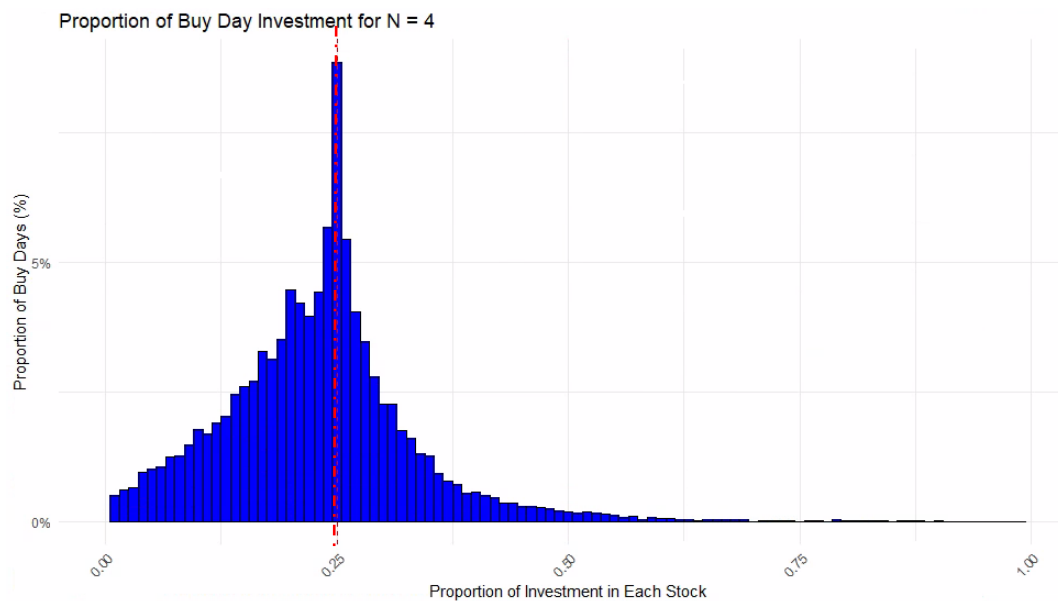


Figure 3 : Proportion des jours d'achats où 4 actions sont achetées

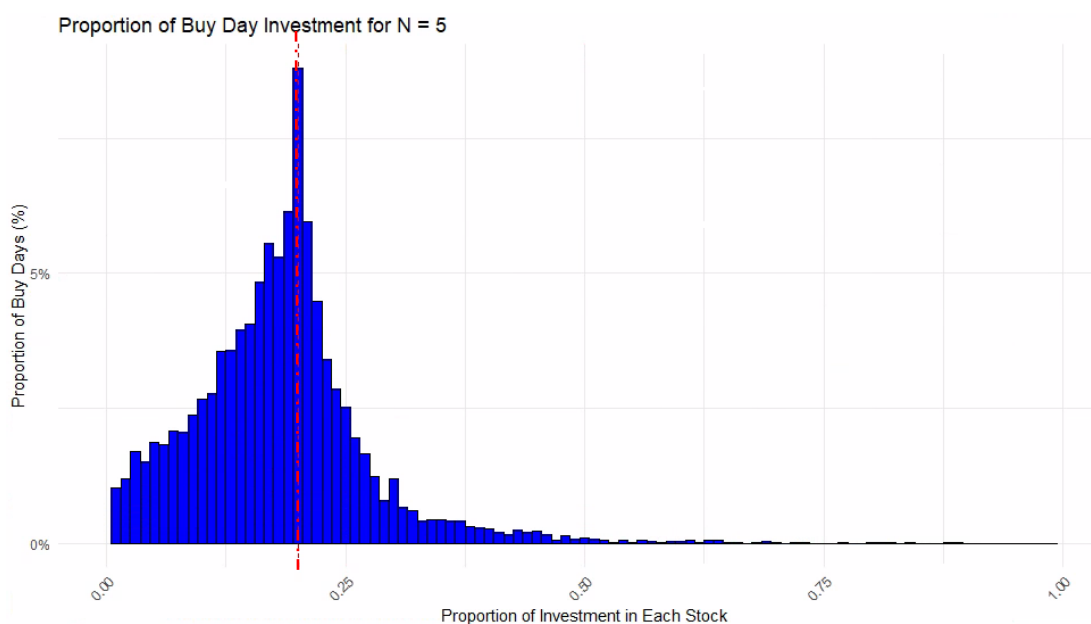


Figure 4 : Proportion des jours d'achats où 5 actions sont achetées

Ces graphiques confirment les valeurs obtenues dans les tableaux pour les différentes tolérances. Les pics significatifs autour des fractions égales ($1/N$) et les proportions élevées de jours d'achat correspondant à cette stratégie renforcent la robustesse de mes conclusions.

Comparaison avec Gathergood et al.

En comparant mes résultats avec ceux de Gathergood et al. (2023), j'observe des tendances convergentes. Les différents résultats obtenus par Gathergood et al. pour les différentes tolérances sont disponibles dans les annexes 2 à 4 de ce mémoire.

Mes résultats montrent des tendances similaires à celles de Gathergood et al. Bien que mes proportions de 1/N Buying soient légèrement différentes, les tendances générales restent les mêmes. En particulier, pour des tolérances de $\pm 5\%$, les pourcentages obtenus sont très proches (par exemple, 25.645% pour $N=3$ dans nos résultats contre 23.3% chez Gathergood et al.). Les variations peuvent être dues à des différences dans les échantillons de données ou à des périodes d'analyse différentes.

Conclusion

L'hypothèse selon laquelle les investisseurs adoptent une stratégie de diversification naïve des achats (NBD) est confirmée par les analyses effectuées. Les résultats montrent que les investisseurs tendent à allouer leurs investissements de manière égale entre les actions achetées, néanmoins cette tendance varie avec le niveau de tolérance. Lorsque celle-ci est faible, une minorité d'investisseurs utilisent la NBD. En revanche, une tolérance plus élevée entraîne une utilisation plus fréquente de la NBD.

Les pics significatifs autour des fractions égales ($1/N$) et les proportions élevées de jours d'achat correspondant à cette stratégie sont cohérents avec les résultats de Gathergood et al., renforçant la robustesse de mes conclusions. Les légères différences observées par rapport aux résultats de Gathergood et al. soulignent l'importance de prendre en compte le contexte des données et les variations potentielles dans les comportements d'investissement.

Le code R utilisé pour tester cette hypothèse est disponible à l'annexe 12.

Hypothèse 2 :

L'Hypothèse 2 vise à déterminer si, lors des jours d'achat multiples, les investisseurs utilisent principalement la diversification naïve des achats (NBD) pour allouer leurs fonds entre les actions ou s'ils utilisent la diversification naïve des portefeuilles (NPD) pour équilibrer leur portefeuille en fonction du nombre d'actions détenues.

- **H0** : Lors des jours d'achat multiples, les investisseurs utilisent principalement la diversification naïve des achats (NBD) pour allouer leurs fonds entre les actions.

- **H1** : Lors des jours d'achat multiples, les investisseurs utilisent principalement la diversification naïve du portefeuille (NPD) pour équilibrer leur portefeuille en fonction du nombre d'actions détenues

Données Utilisées

Pour tester cette hypothèse, j'ai utilisé les données de transactions d'actions de multiples investisseurs. Ces données incluent des informations sur les transactions d'achat et de vente (*stock_trades*), ainsi que la composition des portefeuilles à différents moments (*stock_holdings*). Pour cette analyse, j'ai limité l'échantillon à 1000 investisseurs choisis de manière aléatoire en raison de limitations techniques : mon ordinateur personnel connecté à la salle virtuelle de l'université ne permet pas d'exécuter ce code sur l'ensemble des investisseurs en un temps de calcul raisonnable, et des interruptions réseau compliquent le calcul.

Résultats :

Avec un nombre de jours d'achats multiples de 2188 jours, les résultats pour les différentes tolérances sont les suivants :

Tolérance 2% :

1/N Existing Positions	1/N Buying	1/N Resulting Positions	Proportion of Buy Days (%)
Non	Non	Non	75.84749
Non	Oui	Non	24.15251
Oui	Oui	Oui	0.00000
Non	Oui	Oui	0.00000
Oui	Non	Oui	0.00000
Non	Non	Oui	0.00000
Oui	Oui	Non	0.00000
Oui	Non	Non	0.00000

Tableau 4 : Hypothèse 2; tolérance $\pm 2\%$

Explication d'une ligne :

- 1/N Existing Positions : Non, 1/N Buying : Non, 1/N Resulting Positions : Non
 - Cela signifie qu'au début de la journée, les portefeuilles des investisseurs ne suivent pas la diversification naïve des portefeuilles (NPD).
 - Pendant la journée, les investisseurs n'utilisent pas la diversification naïve des achats (NBD).

- À la fin de la journée, les portefeuilles ne sont pas NPD.
- Proportion of Buy Days : 75.84% : 75.84% des jours d'achat multiples observés montrent que les investisseurs ne commencent pas avec un portefeuille NPD, n'utilisent pas NBD pour leurs achats, et ne terminent pas la journée avec un portefeuille NPD.

Analyse globale :

- Total des portefeuilles qui ne sont pas NPD en fin de journée : 100% : Pour une tolérance de 2%, aucun des portefeuilles ne termine en NPD.
- Pourcentage de jours avec des achats NBD : 24.15% : Pour une tolérance de 2%, 24.15% des jours d'achat multiples impliquent des transactions suivant la stratégie NBD.

Tolérance 5% :

1/N Existing Positions	1/N Buying	1/N Resulting Positions	Proportion of Buy Days (%)
Non	Oui	Non	35.895425
Non	Non	Non	48.861622
Oui	Oui	Oui	7.203755
Non	Oui	Oui	0.000000
Oui	Non	Oui	8.039197
Non	Non	Oui	0.000000
Oui	Oui	Non	0.000000
Oui	Non	Non	0.000000

Tableau 5 : Hypothèse 2; tolérance $\pm 5\%$

Explication d'une ligne :

- 1/N Existing Positions : Oui, 1/N Buying : Oui, 1/N Resulting Positions : Oui
 - Cela signifie qu'au début de la journée, les portefeuilles des investisseurs suivent la diversification naïve des portefeuilles (NPD).
 - Pendant la journée, les investisseurs utilisent la diversification naïve des achats (NBD).
 - À la fin de la journée, les portefeuilles sont NPD.

- Proportion of Buy Days : 7.20% : 7.20% des jours d'achat multiples montrent que les investisseurs commencent avec un portefeuille NPD, utilisent NBD pour leurs achats, et terminent la journée avec un portefeuille NPD.

Analyse globale :

- Total des portefeuilles qui ne sont pas NPD en fin de journée : 84.75% : Pour une tolérance de 5%, 84.75% des portefeuilles ne sont pas NPD en fin de journée.
- Pourcentage de jours avec des achats NBD : 43.09% : Pour une tolérance de 5%, 43.09% des jours d'achat multiples impliquent des transactions suivant la stratégie NBD.

Tolérance 10% :

1/N Existing Positions	1/N Buying	1/N Resulting Positions	Proportion of Buy Days (%)
Non	Oui	Non	47.2204567
Non	Non	Non	28.3878273
Oui	Oui	Oui	15.7516043
Non	Oui	Oui	0.0011190
Oui	Non	Oui	8.6325314
Non	Non	Oui	0.0064612
Oui	Oui	Non	0.0000000
Oui	Non	Non	0.0000000

Tableau 6 : Hypothèse 2; tolérance $\pm 10\%$

Explication d'une ligne :

- 1/N Existing Positions : Oui, 1/N Buying : Non, 1/N Resulting Positions : Oui
 - Cela signifie qu'au début de la journée, les portefeuilles des investisseurs suivent la diversification naïve des portefeuilles (NPD).
 - Pendant la journée, les investisseurs n'utilisent pas la diversification naïve des achats (NBD).
 - À la fin de la journée, les portefeuilles sont NPD.
 - Proportion of Buy Days : 8.63% : 8.63% des jours d'achat multiples montrent que les investisseurs commencent avec un portefeuille NPD, n'utilisent pas NBD pour leurs achats, mais terminent la journée avec un portefeuille NPD.

Analyse globale :

- Total des portefeuilles qui ne sont pas NPD en fin de journée : 75.60% : Pour une tolérance de 10%, 75.60% des portefeuilles ne sont pas NPD en fin de journée.
- Pourcentage de jours avec des achats NBD : 62.97% : Pour une tolérance de 10%, 62.97% des jours d'achat multiples impliquent des transactions suivant la stratégie NBD.

Comparaison avec Gathergood (2023)

Dans l'article de Gathergood (2023), les résultats indiquent que les investisseurs suivent rarement une stratégie de NPD lors des jours d'achat multiples. Les proportions de jours d'achat où les portefeuilles sont NPD en fin de journée sont faibles, similaires à mes résultats. Les tableaux obtenus par Gathergood et al. sont disponibles dans les annexes 5 à 7 de ce mémoire.

Par exemple, dans le tableau de Gathergood pour une tolérance de 2%, la proportion de jours d'achat où le portefeuille est NPD en fin de journée est également très faible (0.7%). De même que pour une tolérance de 5% où ce pourcentage est égal à 1.9% ; et pour une tolérance de 10% où ce pourcentage est égal à 3.4%.

Concernant la proportion de jours d'achats NBD, ces derniers sont de 29.7%, 35.6% et 42.3% respectivement pour des tolérances de 2%, 5% et 10%.

Conclusion

Sur la base de mes résultats et de la comparaison avec l'étude de Gathergood & al. (2023), je peux conclure que les investisseurs n'utilisent principalement pas la diversification naïve des portefeuilles (NPD) lors des jours d'achat multiples. La proportion de jours où les portefeuilles suivent une stratégie de NPD est très faible pour toutes les tolérances testées. Ainsi, je rejette l'hypothèse H1 et accepte l'hypothèse H0 : Les investisseurs utilisent majoritairement une stratégie de diversification naïve des achats (NBD) lors des jours d'achat multiples.

Le code R utilisé pour tester cette hypothèse est disponible à l'annexe 13.

Hypothèse 3 :

Pour tester l'hypothèse selon laquelle les investisseurs utilisant la diversification naïve à l'achat (NBD) obtiennent des rendements annuels différents par rapport aux autres investisseurs, j'ai formulé les hypothèses suivantes :

- **H0** : Il n'y a pas de différence significative entre les rendements annuels des investisseurs utilisant la NBD et ceux ne l'utilisant pas.
- **H1** : Il existe une différence significative entre les rendements annuels des investisseurs utilisant la NBD et ceux ne l'utilisant pas.

Méthodologie

1. Identification des Jours NBD :

- **Calcul de la part d'investissement** : Pour chaque transaction d'achat, j'ai calculé la part de l'investissement pour chaque action achetée. Si la différence absolue entre cette part et l'allocation égale ($1/N$) est inférieure à 2 %, je considère que l'investisseur utilise la NBD ce jour-là. Concrètement, pour chaque transaction, j'ai utilisé la formule suivante : $\text{part_d'investissement} = (\text{Valeur_d'Exécution_EUR} / \text{Valeur_Totale})$; où Valeur_d'Execution_EUR est la valeur de chaque transaction d'achat et Valeur_Totale est la somme des valeurs de toutes les transactions d'achat pour ce jour-là. Si la différence entre cette part et $1/N$ (où N est le nombre total d'actions achetées) est inférieure à 2 %, alors cette transaction est considérée comme utilisant la NBD.

2. Classification des Investisseurs :

- **50% ou Plus NBD** : Investisseurs ayant utilisé la NBD dans 50% ou plus de leurs jours d'achat multiples.
- **Au Moins une Fois NBD** : Investisseurs ayant utilisé la NBD au moins une fois mais moins de 50% du temps.
- **Jamais NBD** : Investisseurs n'ayant jamais utilisé la NBD.

Pour chaque investisseur, j'ai comptabilisé le nombre de jours où ils ont utilisé la NBD et calculé le ratio de ces jours par rapport au total de leurs jours d'achat multiples. Cela m'a permis de les classer dans l'un des trois groupes mentionnés ci-dessus.

3. Calcul des Rendements Géométriques :

- **Pourquoi la Moyenne Géométrique ?** : La moyenne géométrique est plus appropriée pour mesurer la performance réalisée des investisseurs sur des périodes variables car elle tient compte de l'effet de la capitalisation. Elle donne

une mesure plus précise des performances d'un portefeuille sur le long terme, surtout lorsque les rendements varient significativement d'une période à l'autre.

- **Application** : J'ai calculé les rendements géométriques moyens pour chaque investisseur en utilisant leurs rendements bruts mensuels. La formule pour la moyenne géométrique des rendements est la suivante :

$$\text{Moyenne Géométrique} = \left(\prod (1 + r_i) \right)^{1/n} - 1$$

où r_i est le rendement pour chaque période et n est le nombre de périodes. Ces rendements sont ensuite annualisés pour faciliter la comparaison avec les autres études et groupes d'investisseurs.

4. Intégration de la période de détention :

- Pour chaque investisseur, la période de détention est définie comme la période entre la première et la dernière date où des rendements sont observés, excluant les NA (données non disponibles). La moyenne géométrique des rendements est calculée sur cette période, reflétant ainsi la performance réelle sur la durée de l'investissement.

Résultats et interprétation

Tableau des Statistiques Descriptives :

Groupe	Rendement Géométrique Annuel Moy. (%)	Écart- Type Annuel	Période de Détention Moyenne (jours)	Nombre d'Investisseurs	Nombre Total d'Achats
50% or More NBD	-6.06	6.73	1890.46	771	4487
At Least One NBD	-4.75	9.75	2120.57	1628	37425
Never NBD	-9.09	19.37	1750.95	1091	2402

Tableau 7 : Tableau des statistiques descriptives

Explication des Colonnes :

- **Rendement Géométrique Annuel Moy. (%)** : La moyenne des rendements géométriques pour chaque groupe d'investisseurs, annualisée pour la comparaison.

- **Écart-Type Annuel** : Indique la variabilité des rendements géométriques annuels autour de la moyenne.
- **Période de Détention Moyenne (jours)** : Le nombre moyen de jours pendant lesquels les investisseurs ont détenu leurs portefeuilles.
- **Nombre d'Investisseurs** : Le nombre total d'investisseurs dans chaque groupe.
- **Nombre Total d'Achats** : Le nombre total d'achats réalisés par les investisseurs de chaque groupe.

Analyse du tableau :

- **Rendements géométriques moyens** :
 - Le groupe "At Least One NBD" a le rendement géométrique annuel moyen le moins négatif à -4.75%, suivi par "50% or More NBD" à -6.06%.
 - Le groupe "Never NBD" a le rendement le plus négatif à -9.09%, indiquant que les investisseurs qui n'utilisent jamais la NBD ont généralement des performances inférieures.
- **Écart-Type Annuel** :
 - Le groupe "Never NBD" a l'écart-type le plus élevé à 19.37, suggérant une plus grande variabilité et incertitude dans les rendements.
 - L'écart-type des groupes utilisant la NBD est inférieur, avec "50% or More NBD" à 6.73 et "At Least One NBD" à 9.75, indiquant des rendements plus prévisibles.
- **Période de détention** :
 - Le groupe "At Least One NBD" tend à avoir une période de détention plus longue (2120.57 jours), ce qui peut refléter une stratégie d'investissement plus stable.

Graphique des rendements annuels moyens :

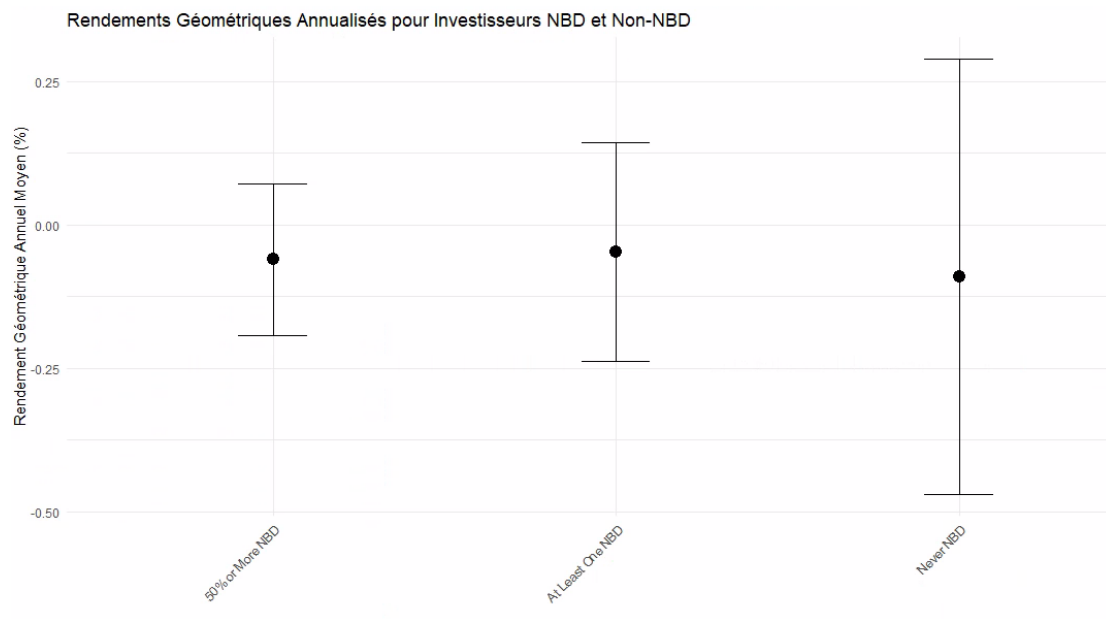


Figure 5 : Graphique des rendements annuels moyens

Le graphique ci-dessus illustre les rendements géométriques annuels moyens pour chaque groupe d'investisseurs avec des intervalles de confiance à 95%. Les points représentent les moyennes et les barres verticales indiquent les intervalles de confiance.

- Les investisseurs qui n'utilisent jamais la NBD ont des rendements géométriques annuels moyens plus faibles par rapport aux deux autres groupes.
- Bien que les différences de rendement semblent indiquer que les investisseurs "Never NBD" ont une performance inférieure, les intervalles de confiance qui se chevauchent suggèrent que ces différences ne sont pas statistiquement significatives.

Graphique des distributions de rendements :

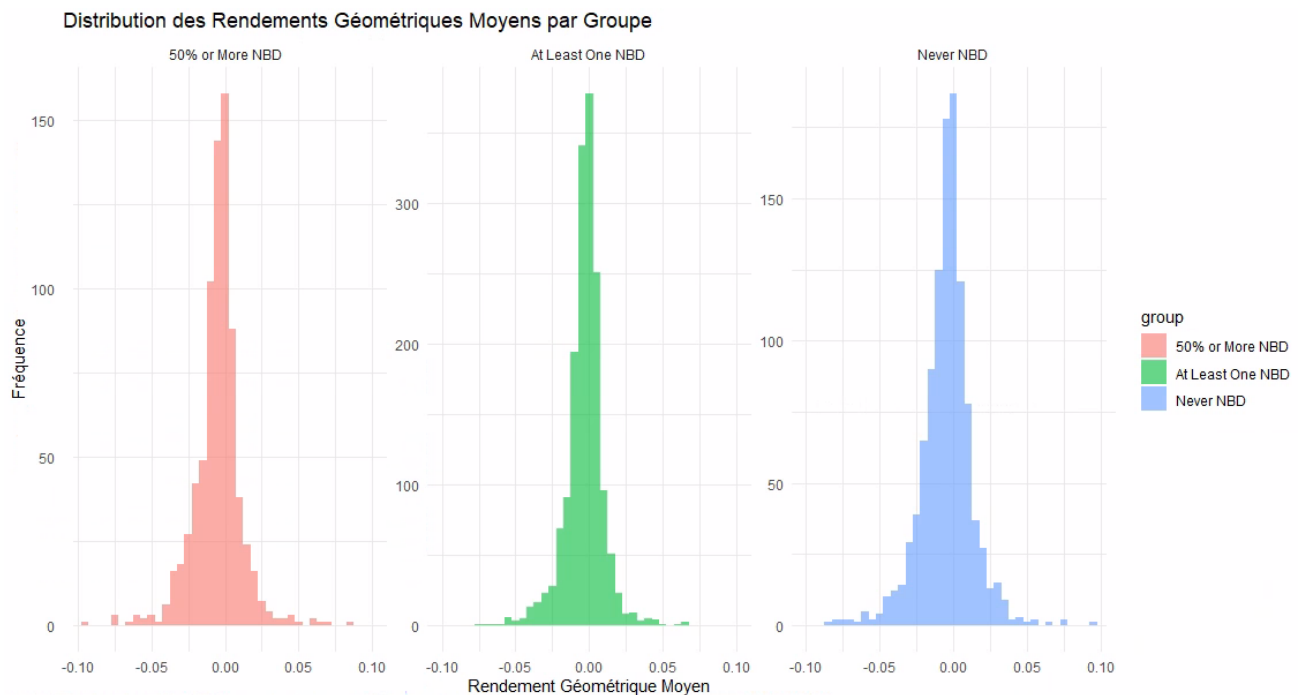


Figure 6 : Graphique des distributions de rendements

Le second graphique, intitulé "Distribution des Rendements Géométriques Moyens par Groupe," montre que le groupe **"50% or More NBD"** présente une distribution fortement concentrée autour de zéro, indiquant une gestion cohérente des portefeuilles. Le groupe **"At Least One NBD"** a une distribution similaire mais légèrement plus dispersée, tandis que le groupe **"Never NBD"** montre une distribution plus large et volatile, suggérant des performances plus imprévisibles.

Tests Statistiques :

Le test t de Welch est utilisé pour comparer les moyennes des rendements géométriques entre les groupes. Ce test est utilisé pour comparer les moyennes de deux échantillons avec des variances inégales et des tailles d'échantillons différentes. Il est approprié ici car les groupes ont des écarts-types et des tailles d'échantillons différents. Les p-values obtenues sont :

- Comparaison entre **"50% or More NBD"** et **"At Least One NBD"** : p-value = 0.81.
- Comparaison entre **"50% or More NBD"** et **"Never NBD"** : p-value = 0.50.
- Comparaison entre **"At Least One NBD"** et **"Never NBD"** : p-value = 0.14.

Les p-values sont toutes supérieures au seuil de 0.05, indiquant que les différences entre les groupes ne sont pas statistiquement significatives. Cela signifie que je ne peux pas rejeter l'hypothèse nulle (H_0) et conclure qu'il n'existe pas de différences significatives entre les rendements annuels moyens des différents groupes.

Comparaison avec les résultats de Gathergood

Dans l'article de Gathergood, les résultats sont présentés dans le graphique suivant :

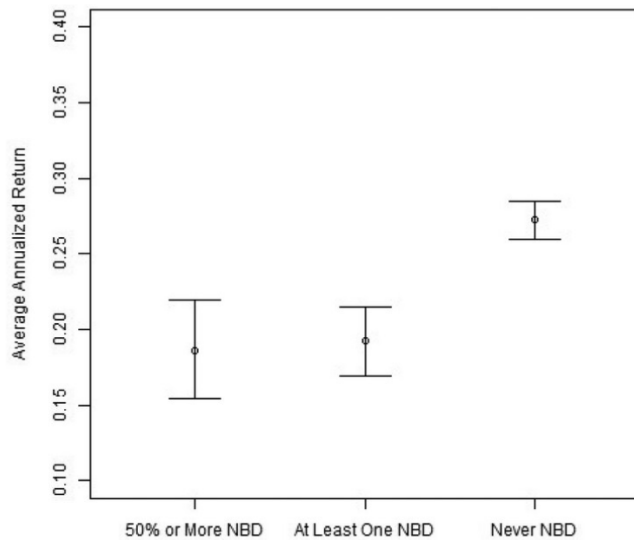


Figure 7 : Résultats de Gathergood & al. hypothèse 3

Gathergood & al. ont constaté dans leur étude que les investisseurs qui n'avaient jamais utilisé la NBD avaient, en moyenne, des rendements annualisés d'environ un tiers plus élevés par rapport à ceux utilisant la NBD 50% ou plus, et par rapport à ceux l'ayant utilisée au moins une fois. En revanche, mes résultats montrent une tendance opposée, bien que non significative, avec des rendements plus faibles pour les investisseurs "Never NBD".

Ces différences peuvent s'expliquer par plusieurs facteurs :

- **Période d'Étude** : Les périodes d'étude différentes peuvent avoir un impact significatif sur les résultats en raison des fluctuations du marché.
- **Échantillon** : La composition de l'échantillon d'investisseurs peut varier, influençant les résultats.
- **Contexte Économique** : Les conditions économiques durant la période d'étude peuvent également jouer un rôle important.

Conclusion

Sur la base des analyses effectuées, l'hypothèse nulle (H_0) n'est pas rejetée. Il n'y a pas de différence statistiquement significative entre les rendements annuels des investisseurs utilisant la NBD et ceux ne l'utilisant pas. Cela suggère que l'utilisation de la NBD n'a pas d'impact significatif sur les rendements des portefeuilles d'investissement, contrairement aux conclusions de Gathergood. Ces résultats indiquent que d'autres facteurs, tels que la sélection des actions, le timing du marché et d'autres stratégies d'investissement, peuvent jouer un rôle plus déterminant dans la performance des portefeuilles.

Le code R utilisé pour tester cette hypothèse est disponible à l'annexe 14.

Hypothèse 4 :

Pour tester l'hypothèse selon laquelle l'utilisation de la diversification naïve à l'achat (NBD) est influencée par diverses variables sociodémographiques et financières des investisseurs, j'ai formulé les hypothèses suivantes :

- **H0 :** Les variables démographiques et financières n'ont pas d'impact significatif sur la probabilité d'utiliser la stratégie de diversification naïve à l'achat (NBD).
- **H1 :** Les variables démographiques et financières ont un impact significatif sur la probabilité d'utiliser la stratégie de diversification naïve à l'achat (NBD).

Modèle de Régression Logistique

J'ai utilisé deux modèles de régression logistique pour examiner l'effet de diverses variables explicatives sur l'utilisation de la stratégie NBD. La régression logistique est appropriée pour les analyses où la variable dépendante est binaire (utilisation de la NBD ou non).

- **Modèle 1 :** L'utilisation de la NBD au moins une fois.
 - La variable dépendante (Y_i) est binaire et égale à 1 si l'investisseur utilise la NBD au moins une fois, et 0 sinon.
- **Modèle 2 :** L'utilisation de la NBD 50% ou plus du temps.
 - La variable dépendante (Y_i) est binaire et égale à 1 si l'investisseur utilise la NBD pour 50% ou plus de ses transactions, et 0 sinon.

Le modèle est spécifié comme suit :

$$\text{logit}(P(Y_i = 1)) = \beta_0 + \beta_1 \times \text{Age_Group}_i + \beta_2 \times \text{Sex_Male}_i + \beta_3 \times \text{SP_Q3}_i + \beta_4 \times \text{SP_Q7}_i + \epsilon_i$$

Explication des Termes

- **Y_i** : Variable binaire indiquant l'utilisation de la NBD par l'investisseur i
- **β₀** : Interception du modèle
- **β₁, β₂, ..., β₄** : Coefficients des variables explicatives
- **ε_i** : Terme d'erreur aléatoire pour capturer la variation inexpliquée par le modèle.
- **Age_Group** : Groupes d'âge des investisseurs (0-20, 21-40, 41-60, 61-80, 81-100).
- **Sex_Male** : Variable binaire (1 pour masculin, 0 pour féminin).
- **SP_Q3** : Niveau de connaissance des marchés financiers (1 pour Très peu de connaissances, 2 pour Non familier mais intéressé, 3 pour Suffisamment d'expérience pour diversifier, 4 pour Bonne connaissance des marchés financiers, 5 pour Investisseur expérimenté).
- **SP_Q7** : Revenu net annuel total (0 pour Moins de 20 000 €, 1 pour 20 000 € à moins de 40 000 €, 2 pour 40 000 € à moins de 75 000 €, 3 pour 75 000 € à moins de 150 000 €, 4 pour 150 000 € et plus).

Résultats de la Régression Logistique

Modèle 1 : NBD utilisé au moins une fois

Term	Estimate	Std. Error	Statistic	P-value	Conf. Low	Conf. High	Odds Ratio
Intercept	36.0	10335.	0.00349	0.997	-479.	3039.	4.40e+15
Age_Group21-40	-17.4	10225.	-0.00170	0.999	NA	1646.	2.86e-8
Age_Group41-60	-16.3	10225.	-0.00159	0.999	NA	1650.	8.63e-8
Age_Group61-80	-15.6	10225.	-0.00152	0.999	NA	1649.	1.72e-7
Age_Group81-100	-0.646	10259.	-0.000063	1.00	-229.	227.	5.24e-1
Sex_Male	0.544	0.547	0.994	0.320	-0.691	1.51	1.72e+0
SP_Q32	-16.0	1505.	-0.0106	0.992	-410.	62.1	1.16e-7
SP_Q33	-15.5	1505.	-0.0103	0.992	-410.	62.6	1.84e-7
SP_Q34	-15.2	1505.	-0.0101	0.992	-409.	62.9	2.53e-7
SP_Q35	-15.1	1505.	-0.0100	0.992	-409.	63.0	2.75e-7
SP_Q71	0.120	0.560	0.214	0.831	-1.03	1.19	1.13e+0
SP_Q72	0.0238	0.614	0.0387	0.969	-1.24	1.19	1.02e+0
SP_Q73	-0.144	0.718	-0.201	0.841	-1.55	1.33	8.66e-1
SP_Q74	-0.397	1.15	-0.345	0.730	-2.36	2.62	6.72e-1

Tableau 8 : Résultats de la régression logistique modèle 1

Modèle 2 : NBD utilisé 50% ou plus

Term	Estimate	Std. Error	Statistic	P-value	Conf. Low	Conf. High	Odds Ratio
Intercept	1.35	1.28	1.05	0.292	-1.09	4.48	3.84
Age_Group21-40	0.0386	1.28	0.0302	0.976	-3.09	2.48	1.04
Age_Group41-60	0.882	1.26	0.702	0.482	-2.22	3.28	2.42
Age_Group61-80	0.852	1.25	0.680	0.497	-2.25	3.25	2.35
Age_Group81-100	0.679	1.26	0.537	0.591	-2.44	3.10	1.97
Sex_Male	0.0912	0.183	0.499	0.618	-0.282	0.437	1.10
SP_Q32	0.435	0.331	1.31	0.189	-0.245	1.06	1.54
SP_Q33	0.561	0.319	1.76	0.0783	-0.0985	1.16	1.75
SP_Q34	0.770	0.318	2.42	0.0155	0.111	1.37	2.16
SP_Q35	0.337	0.347	0.972	0.331	-0.371	0.996	1.40
SP_Q71	-0.298	0.194	-1.53	0.125	-0.687	0.0751	0.743
SP_Q72	-0.239	0.204	-1.17	0.242	-0.648	0.154	0.788
SP_Q73	-0.797	0.221	-3.61	0.000301	-1.24	-0.369	0.451
SP_Q74	-1.29	0.282	-4.58	0.00000471	-1.84	-0.731	0.275

Tableau 9 : Résultats de la régression logistique modèle 2

Explication des Colonnes

- **Term** : Le nom de la variable.
- **Estimate** : Le coefficient estimé pour la variable, indiquant l'effet sur l'utilisation de la NBD. Un coefficient positif signifie que l'augmentation de la variable accroît la probabilité d'utiliser la NBD, tandis qu'un coefficient négatif indique le contraire.
- **Std. Error** : L'erreur standard de l'estimation du coefficient.
- **Statistic** : La statistique de test pour le coefficient, utilisée pour déterminer la significativité du coefficient.
- **P-value** : La valeur p pour le test de significativité du coefficient. Une valeur p inférieure à 0.05 indique que le coefficient est statistiquement significatif.
- **Conf. Low & Conf. High** : La borne inférieure et supérieure de l'intervalle de confiance à 95% pour l'estimation du coefficient.
- **Odds Ratio** : L'odds ratio exprime la probabilité qu'un événement se produise (ici, l'utilisation de la NBD) par rapport à la probabilité qu'il ne se produise pas, pour une unité de changement dans la variable explicative. Un odds ratio supérieur à 1 indique que l'augmentation de la variable accroît la probabilité d'utiliser la NBD, tandis qu'un odds ratio inférieur à 1 indique une diminution de cette probabilité.

Interprétation des Résultats

Voici une explication détaillée des résultats obtenus pour chaque variable de chaque modèle :

Modèle 1

- Aucun des groupes d'âge n'est statistiquement significatif.
- Le sexe masculin (Sex_Male) n'est pas statistiquement significatif.
- Les niveaux de connaissance des marchés financiers (SP_Q3) et les revenus (SP_Q7) ne montrent pas d'effet significatif globalement

Modèle 2

- **Age_Group** : Aucun des groupes d'âge n'est statistiquement significatif par rapport au groupe de référence (0-20 ans), indiquant que l'âge n'a pas un impact significatif sur l'utilisation de la NBD.

- **Sex_Male** : Le sexe (1 = homme) n'est pas statistiquement significatif, ce qui suggère que le sexe n'affecte pas de manière significative la probabilité d'utiliser la NBD.
- **SP_Q3 (Niveau de Connaissance des Marchés Financiers)** :
 - Seule la catégorie SP_Q34 ("Bonne connaissance des marchés financiers") est significative avec un p-value de 0.0155. L'odds ratio de 2.16 suggère que les investisseurs avec une bonne connaissance des marchés financiers sont significativement plus susceptibles d'utiliser la NBD pour 50% ou plus de leurs transactions, comparé à ceux ayant très peu de connaissances. Les autres catégories de connaissances ne montrent pas d'effet significatif.
 - Cependant, les investisseurs classés comme "investisseurs expérimentés" (SP_Q35) montrent une légère baisse dans l'utilisation de la NBD, ce qui peut suggérer une transition vers des stratégies d'investissement plus complexes ou, alternativement, être dû à une surconfiance lors de l'évaluation de leur niveau de connaissances des marchés financiers.
- **SP_Q7 (Revenu Net Annuel Total)** :
 - Les catégories SP_Q73 et SP_Q74 sont significatives avec des odds ratios de 0.451 et 0.275 respectivement. Cela indique que les investisseurs ayant un revenu net plus élevé sont moins susceptibles d'utiliser la NBD par rapport à ceux ayant un revenu net de moins de 20000€.

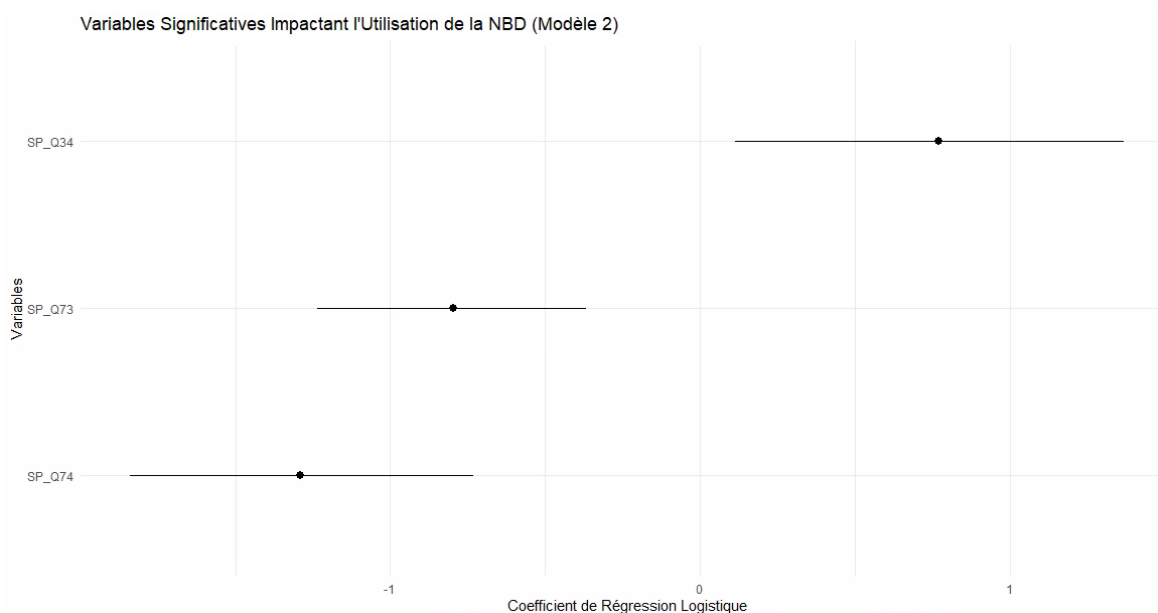


Figure 8 : Variables significatives impactant l'utilisation de la NBD modèle 2

Graphiques et Tableaux

Les tableaux suivants illustrent les proportions de NBD par groupe d'âge, sexe, niveau de connaissance des marchés financiers et revenu net annuel total pour le modèle 2. Les graphiques relatifs à ces variables sont disponibles dans les annexes 8 à 11 de ce mémoire :

SP_Q7 :

SP_Q7	Nombre de NBD (nbd_count)	Nombre total (total_count)	Pourcentage NBD (percentage_nbd)
2	1529	1637	93.4%
1	1690	1824	92.7%
0	653	705	92.6%
3	573	644	89.0%
4	126	154	81.8%

Tableau 10 : Proportion de NBD par revenu net annuel total

SP_Q3 :

SP_Q3	Nombre de NBD (nbd_count)	Nombre total (total_count)	Pourcentage NBD (percentage_nbd)
4	2003	2148	93.2%
3	1388	1507	92.1%
2	687	753	91.2%
1	116	130	89.2%
5	377	426	88.5%

Tableau 11 : Proportion de NBD par niveau de connaissance des marchés financiers

Sex :

Sexe masculin (Sex_Male)	Nombre de NBD (nbd_count)	Nombre total (total_count)	Pourcentage NBD (percentage_nbd)
1 (Homme)	4153	4508	92.1%
0 (Femme)	418	456	91.7%

Tableau 12 : Proportion de NBD par sexe

Age_Group :

Groupe d'âge (Age_Group)	Nombre de NBD (nbd_count)	Nombre total (total_count)	Pourcentage NBD (percentage_nbd)
41-60	2106	2274	92.6%
61-80	1856	2013	92.2%
81-100	405	442	91.6%
21-40	200	230	87.0%
0-20	2	3	66.7%

Tableau 13 : Proportion de NBD par groupe d'âge

Conclusion

L'analyse menée sur l'utilisation de la stratégie de diversification naïve à l'achat (NBD) a révélé plusieurs résultats intéressants concernant l'influence des variables sociodémographiques et financières. Malgré l'absence de significativité des variables dans le modèle 1, le modèle 2 a mis en évidence des effets significatifs pour certaines catégories de connaissance des marchés financiers et de revenu net annuel.

Les résultats montrent que les investisseurs ayant une bonne connaissance des marchés financiers sont plus enclins à utiliser la NBD fréquemment, ce qui souligne l'importance de la familiarité avec les investissements dans la prise de décisions financières. En revanche, les investisseurs à revenu élevé tendent à éviter cette stratégie, ce qui peut suggérer que des revenus plus élevés permettent l'accès à des conseils financiers plus sophistiqués ou à une plus grande diversification stratégique.

Ce modèle de régression logistique offre une compréhension approfondie des facteurs influençant les décisions d'investissement, notamment en ce qui concerne la diversification naïve. Ces informations sont précieuses pour les conseillers financiers et les gestionnaires de portefeuille dans la création de stratégies personnalisées pour différents profils d'investisseurs.

Le code R utilisé pour tester cette hypothèse est disponible à l'annexe 15.

V. Conclusion

Cette étude visait à examiner le comportement des investisseurs belges en matière de diversification naïve des achats (NBD) en répliquant l'analyse réalisée par Gathergood et al. (2023). Mon objectif était de déterminer dans quelle mesure les investisseurs belges adoptent des stratégies similaires ou différentes par rapport à leurs homologues britanniques.

Mes résultats confirment en grande partie les conclusions de Gathergood et al. J'ai trouvé que les investisseurs belges, tout comme leurs équivalents britanniques, ont tendance à utiliser l'heuristique de diversification naïve des achats (NBD) lorsqu'ils achètent plusieurs actions le même jour. Plus précisément, j'ai observé que :

1. Comportement des Investisseurs (Hypothèse 1) : Les investisseurs belges montrent une forte tendance à répartir équitablement leurs investissements entre les différentes actions achetées en une journée, suivant ainsi une stratégie de NBD. Cette tendance est particulièrement marquée lorsque l'on considère des tolérances plus larges, comme $\pm 5\%$ ou $\pm 10\%$, pour l'allocation égale théorique ($1/N$). Ces résultats sont cohérents avec les observations de Gathergood et al., qui ont également noté une utilisation répandue de la NBD chez les investisseurs britanniques.
2. Prévalence de la NBD par rapport à la NPD (Hypothèse 2) : Lors des jours d'achat multiples, les investisseurs belges utilisent majoritairement la NBD plutôt que la NPD. La proportion de jours où les portefeuilles suivent une stratégie de NPD en fin de journée est très faible pour toutes les tolérances testées, indiquant que les investisseurs se concentrent davantage sur la répartition égale lors des nouveaux achats plutôt que sur l'équilibrage global de leur portefeuille. Cette observation corrobore les conclusions de Gathergood et al., qui ont également trouvé que les investisseurs suivent rarement une stratégie NPD lors des jours d'achats multiples.
3. Impact sur la Performance du Portefeuille (Hypothèse 3) : Mes analyses n'ont pas révélé de différence statistiquement significative entre les rendements annuels des investisseurs utilisant la NBD et ceux ne l'utilisant pas. Cela suggère que l'utilisation de la NBD n'a pas d'impact significatif sur les rendements des portefeuilles d'investissement, contrairement aux conclusions de Gathergood et al., qui avaient trouvé une tendance à des rendements inférieurs pour les utilisateurs de la NBD.
4. Influence des Variables Démographiques et Financières (Hypothèse 4) : Malgré l'absence de significativité des variables du modèle 1, certaines variables du modèle 2,

notamment les catégories de niveau de connaissance des marchés financiers et de revenu net annuel total, influencent significativement l'utilisation de la NBD parmi les investisseurs belges. En particulier, les investisseurs ayant une meilleure connaissance des marchés financiers sont plus susceptibles d'utiliser la NBD tandis que ceux ayant des revenus élevés tendent à éviter cette stratégie. Ces résultats diffèrent des conclusions de Gathergood et al., qui ont observé que la propension à utiliser l'heuristique NBD diminue avec l'expérience de l'investisseur et les enjeux financiers. Cela peut suggérer des différences contextuelles ou culturelles entre les investisseurs belges et britanniques.

Limites et Perspectives de Recherche

Mes analyses présentent plusieurs limites, notamment l'utilisation de données historiques spécifiques à un échantillon d'investisseurs belges, ce qui peut limiter la généralisation des résultats à d'autres contextes géographiques ou économiques. De plus, la période d'étude et les conditions économiques spécifiques peuvent avoir influencé les comportements d'investissement observés.

Pour l'hypothèse 2, je n'ai pas pu optimiser le programme suffisamment pour traiter l'ensemble des données, ce qui m'a conduit à travailler avec un échantillon aléatoire de 1000 investisseurs. À l'avenir, utiliser un programme plus optimisé pourrait permettre de gagner du temps de calcul et de traiter l'intégralité des données de manière plus efficace.

Pour les recherches futures, il serait intéressant d'examiner l'impact de programmes d'éducation financière et d'explorer des mécanismes automatisés qui aident les investisseurs à maintenir une diversification optimale de leur portefeuille. Par exemple, des systèmes qui ajustent automatiquement la composition d'un portefeuille pour maintenir une allocation cible pourraient aider les investisseurs à éviter les biais comportementaux et à améliorer la performance de leurs investissements.

Conclusion finale :

En conclusion, cette étude montre que les investisseurs belges, à l'instar de leurs homologues britanniques, utilisent également la NBD, influencée par divers biais comportementaux. Bien que cette stratégie simplifie la prise de décision, elle ne garantit pas nécessairement une performance optimale du portefeuille. Les résultats de cette étude soulignent l'importance de développer des outils et des stratégies qui peuvent aider les investisseurs à prendre des décisions plus informées et à diversifier leurs portefeuilles de manière plus efficace.

Bibliographie

- Baltussen, G., & Post, G. T. (2011). Irrational Diversification: An Examination of Individual Portfolio Choice. *The Journal of Financial and Quantitative Analysis*, 46(5), 1463–1491.
- Barber, B. M., & Odean, T. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *The Review of Financial Studies*, 21(2), 785–818.
- Barberis, N., & Thaler, R. (2002). *A Survey of Behavioral Finance* (Working Paper 9222). National Bureau of Economic Research. <https://doi.org/10.3386/w9222>
- DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus Naïve Diversification: How Inefficient Is the 1/N Portfolio Strategy? *The Review of Financial Studies*, 22(5), 1915–1953.
- Fabozzi, F. J., Gupta, F., & Markowitz, H. M. (2002). The Legacy of Modern Portfolio Theory. *Institutional Investor*.
- Gathergood, J., Hirshleifer, D., Leake, D., Sakaguchi, H., & Stewart, N. (2023). Naïve Buying Diversification and Narrow Framing by Individual Investors. *The Journal of Finance*, 78(3), 1705–1741. <https://doi.org/10.1111/jofi.13222>
- Hsu, P.-H., Han, Q., Wu, W., & Cao, Z. (2018). Asset allocation strategies, data snooping, and the 1 / N rule. *Journal of Banking & Finance*, 97, 257–269. <https://doi.org/10.1016/j.jbankfin.2018.09.021>
- Jain, R., Jain, P., & Jain, C. (2015). *Behavioral Biases in the Decision Making of Individual Investors*. 14(3), 7–27.
- Kirby, C., & Ostdiek, B. (2012). It's All in the Timing: Simple Active Portfolio Strategies that Outperform Naïve Diversification. *The Journal of Financial and Quantitative Analysis*, 47(2), 437–467.
- Kliger, D., & Kudryavtsev, A. (2010). The Availability Heuristic and Investors' Reaction to Company-Specific Events. *Journal of Behavioral Finance*, 11(1), 50–65. <https://doi.org/10.1080/15427561003591116>
- Kritzman, M., Page, S., & Turkington, D. (2010). In Defense of Optimization: The Fallacy of 1/N. *Financial Analysts Journal*, 66(2), 31–39.
- Lassance, N., DeMiguel, V., & Vrms, F. (n.d.). *Optimal Portfolio Diversification via Independent Component Analysis*.
- Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77–91. <https://doi.org/10.2307/2975974>

- Odean, T. (1998). Are Investors Reluctant to Realize Their Losses? *The Journal of Finance*, 53(5), 1775–1798. <https://doi.org/10.1111/0022-1082.00072>
- Shefrin, H., & Statman, M. (1985). The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence. *The Journal of Finance*, 40(3), 777–790. <https://doi.org/10.2307/2327802>
- Tu, J., & Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 99(1), 204–215. <https://doi.org/10.1016/j.jfineco.2010.08.013>
- Tversky, A., & Kahneman, D. (1974). *Judgment under Uncertainty: Heuristics and Biases*. 185.

Annexes

I. Revue de littérature

Annexe 1 : Résumé de la discussion dans la littérature des avantages et inconvénients de la diversification naïve du portefeuille

Plusieurs articles scientifiques font référence au phénomène de Naïve diversification, comme par exemple l'article de DeMiguel et al. intitulé « Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy? » (DeMiguel et al., 2009). Dans cet article, les auteurs ont mené une étude comparative de 14 modèles d'allocation d'actifs en utilisant des données empiriques et des simulations. Les résultats ont montré que la règle de diversification naïve 1/N a souvent produit des résultats similaires ou même meilleurs que les modèles d'allocation d'actifs optimaux, en particulier lorsque les portefeuilles étaient diversifiés.

L'article souligne également que les poids de portefeuille basés sur les estimations de l'échantillon peuvent être excessivement positifs ou négatifs, s'éloignant de l'allocation optimale. Les "erreurs d'allocation" dues à l'utilisation des pondérations 1/N peuvent être moins importantes que celles causées par l'utilisation de pondérations provenant d'un modèle d'optimisation avec des paramètres estimés de manière incorrecte. Même les modèles spécifiquement conçus pour atténuer l'impact des erreurs d'estimation n'obtiennent qu'un succès modeste en raison de l'ampleur des erreurs d'estimation sur les pondérations.

Une autre raison pour laquelle la stratégie 1/N fonctionne bien dans les jeux de données examinés est son utilisation pour répartir la richesse entre des portefeuilles d'actions plutôt qu'entre des actions individuelles. Les portefeuilles diversifiés présentent une volatilité idiosyncratique plus faible que les actifs individuels, réduisant ainsi les pertes dues à une diversification simple par rapport à une diversification optimale lors de la répartition de la richesse entre des portefeuilles. Les simulations montrent que les politiques de diversification optimale l'emportent sur la règle 1/N uniquement dans le cas de niveaux très élevés de volatilité idiosyncratique. Un avantage supplémentaire de la règle 1/N est sa facilité d'application à un grand nombre d'actifs, contrairement aux modèles d'optimisation qui nécessitent généralement l'estimation de paramètres supplémentaires à mesure que le nombre d'actifs augmente (DeMiguel et al., 2009). Cependant, si certains actifs sont plus risqués que d'autres, un portefeuille équipondéré ne sera pas diversifié de manière optimale du point de vue du risque (Lassance et al., n.d.)

Sur base de la conclusion de l'article de DeMiguel et al. (2009), comme quoi l'optimisation du portefeuille n'apporte aucune valeur ajoutée, la valeur de l'optimisation de portefeuille a été défendue par de nombreuses études.

Par exemple l'article « In defense of Optimization ; The Fallacy of 1/N » (Kritzman et al., 2010) conteste l'idée que les portefeuilles équitables pondérés surpassent les portefeuilles optimisés, en affirmant qu'avec des données naïves, les portefeuilles optimisés tendent à surpasser ceux pondérés équitables. La soi-disant supériorité de l'approche 1/N ne vient pas de limitations de l'optimisation, mais de l'utilisation de courtes séquences d'échantillons pour estimer les rendements attendus, souvent conduisant à des attentes peu plausibles. En revanche, en se basant sur des échantillons à plus long terme ou sur des hypothèses naïves mais plausibles, les portefeuilles optimisés surpassent ceux pondérés équitables hors échantillon. Ils démontrent par exemple que la stratégie de diversification 1/N est surpassée par les stratégies de variance minimale et de variance moyenne, basées sur une estimation de la prime de risque à partir d'un échantillon de données long.

Kirby et Ostdiek remettent en question la méthodologie utilisée par DeMiguel et al. (2009) dans leur analyse, mettant en évidence un biais potentiel dans la comparaison des stratégies de portefeuille. Ils soulignent que l'approche de DeMiguel et al. vise un rendement attendu conditionnel nettement supérieur pour les portefeuilles MVE par rapport à celui de la stratégie 1/N, ce qui introduit un biais en faveur de la diversification naïve 1/N. Selon les auteurs, cette méthode augmente le risque d'estimation et provoque un turnover excessif, ce qui compromet la validité des conclusions de l'étude. Kirby et Ostdiek proposent deux nouvelles techniques de sélection de portefeuille pour résoudre ces problèmes méthodologiques ; le "volatility timing" (VT) et le "reward-to-risk timing" (RRT). La première, connue sous le nom de "volatility timing" (VT), est conçue pour atténuer les effets négatifs du turnover élevé sur les performances globales des portefeuilles tout en offrant des performances supérieures à la diversification naïve pour divers jeux de données. La deuxième technique, le "reward-to-risk timing" (RRT), vise à optimiser le rapport entre le rendement et le risque, ce qui permet également d'améliorer les performances des portefeuilles par rapport aux méthodes de diversification naïve. (Kirby & Ostdiek, 2012).

Enfin, l'article de Tu & Zhou (Tu & Zhou, 2011) suggère que la combinaison de la simplicité de la règle 1/N avec des stratégies sophistiquées basées sur la théorie peut considérablement améliorer les performances des portefeuilles, même avec des échantillons de petite taille, surperformant ainsi la diversification naïve.

II. Hypothèse 1

Annexe 2 : Table tolérance $\pm 2\%$

Panel A: ($\mathbb{E}P/N \times (1 \pm 0.02)$)				
N	$1/N$ Buying (%)	LL	UL	Buy Days
2	29.7	29.1	30.3	177,193
3	20.3	19.4	21.1	48,896
4	18.6	17.6	19.7	17,672
5	17.5	16.0	19.0	7,925
6+	15.2	13.1	17.3	9,899
All	26.3	25.6	26.9	261,585

Annexe 3 : Table tolérance $\pm 5\%$

Panel B: ($\mathbb{E}P/N \times (1 \pm 0.05)$)				
N	$1/N$ Buying (%)	LL	UL	Buy Days
2	36.5	36.0	37.1	177,193
3	23.3	22.4	24.2	48,896
4	20.9	19.7	22.1	17,672
5	20.1	18.6	21.7	7,925
6+	18.0	15.3	20.6	9,899
All	31.8	31.1	32.5	261,585

Annexe 4 : Table tolérance $\pm 10\%$

Panel C: ($\mathbb{E}P/N \times (1 \pm 0.1)$)				
N	$1/N$ Buying (%)	LL	UL	Buy Days
2	45.6	45.1	46.2	177,193
3	27.8	26.8	28.7	48,896
4	23.9	22.6	25.2	17,672
5	22.4	20.7	24.0	7,925
6+	20.0	17.2	22.8	9,899
All	39.1	38.4	39.9	261,585

III. Hypothèse 2

Annexe 5 : Table tolérance $\pm 2\%$

Panel A: ($EP/N \times (1 \pm 0.02)$)					
1 / N Existing Positions	1 / N Buying	1 / N Resulting Positions	Proportion of Buy Days	LL	UL
Yes	Yes	Yes	0.5	0.4	0.7
Yes	Yes	No	0.6	0.4	0.7
Yes	No	Yes	0.0	0.0	0.0
Yes	No	No	1.1	0.9	1.3
No	Yes	Yes	0.1	0.0	0.1
No	Yes	No	28.5	27.3	29.8
No	No	Yes	0.1	0.0	0.1
No	No	No	69.2	67.8	70.4

Annexe 6 : Table tolérance $\pm 5\%$

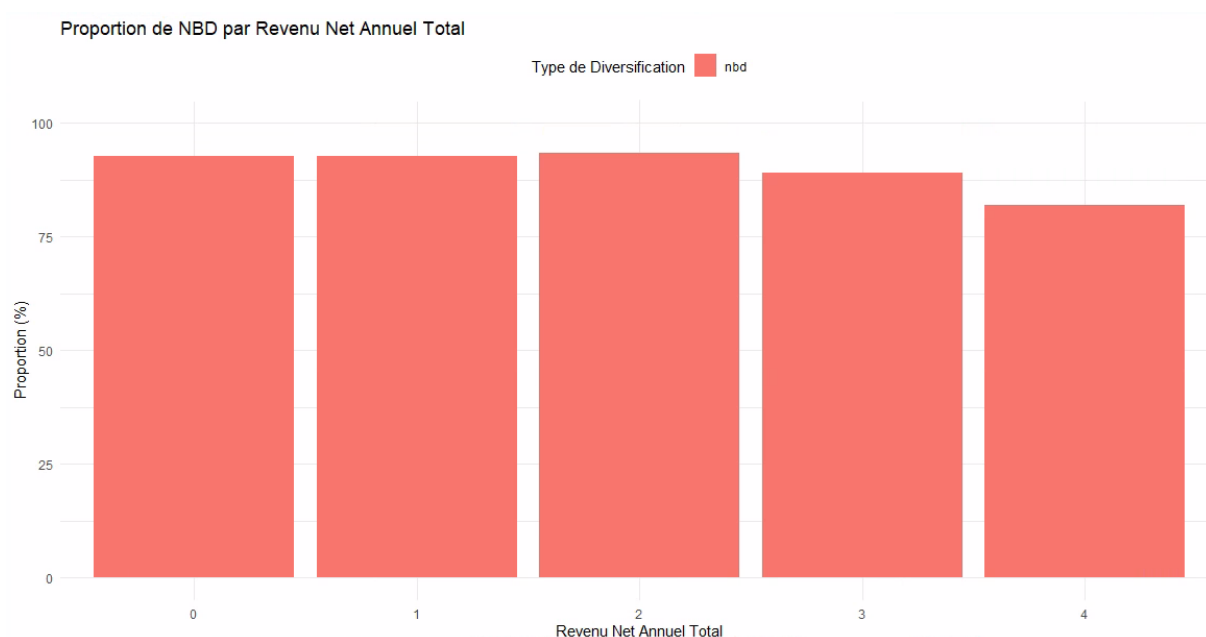
Panel B: ($EP/N \times (1 \pm 0.05)$)					
1 / N Existing Positions	1 / N Buying	1 / N Resulting Positions	Proportion of Buy Days	LL	UL
Yes	Yes	Yes	1.5	1.3	1.7
Yes	Yes	No	1.3	1.1	1.5
Yes	No	Yes	0.1	0.0	0.1
Yes	No	No	1.9	1.6	2.1
No	Yes	Yes	0.1	0.1	0.2
No	Yes	No	32.7	31.5	34.0
No	No	Yes	0.2	0.1	0.3
No	No	No	62.2	60.9	63.5

Annexe 7 : Table tolerance $\pm 10\%$

Panel C ($EP/N \times (1 \pm 0.1)$)					
1 / N Existing Positions	1 / N Buying	1 / N Resulting Positions	Proportion of Buy Days	LL	UL
Yes	Yes	Yes	2.7	2.4	3.1
Yes	Yes	No	2.2	2.0	2.5
Yes	No	Yes	0.2	0.1	0.2
Yes	No	No	2.7	2.4	3.0
No	Yes	Yes	0.2	0.2	0.3
No	Yes	No	37.2	36.0	38.3
No	No	Yes	0.3	0.2	0.4
No	No	No	54.5	53.3	55.8

IV. Hypothèse 4

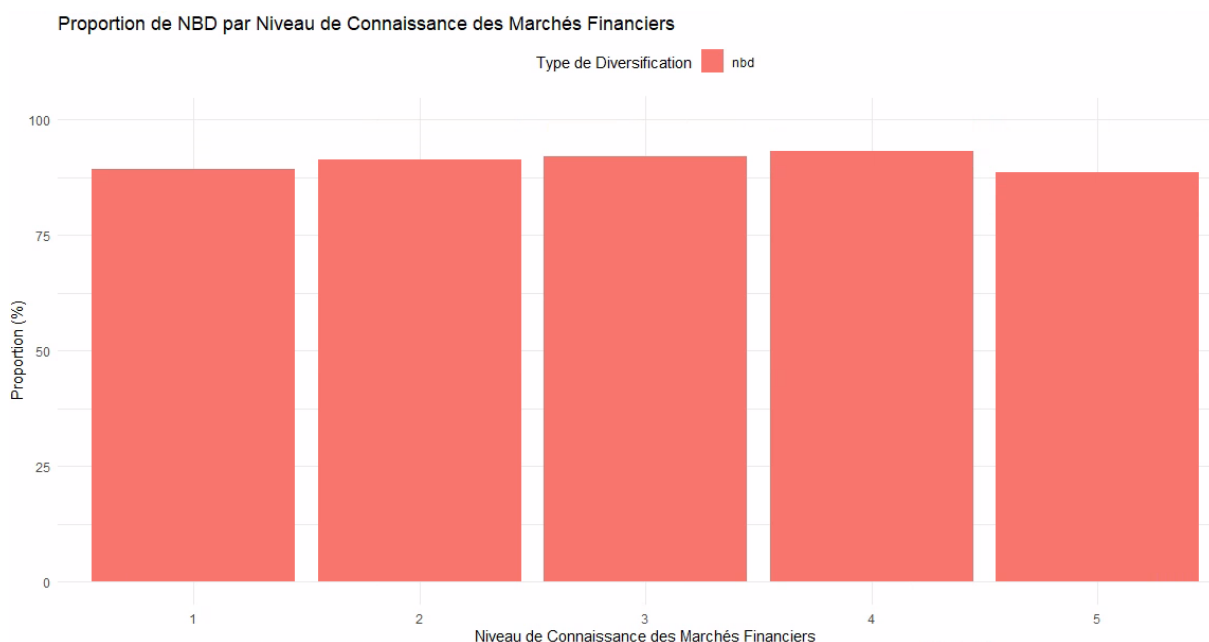
Annexe 8 : Table SP_Q7; Proportion de NBD par revenu net annuel total



Légende pour l'axe des x :

- 0 = Moins de 20000 €
- 1 = Entre 20000 € et moins de 40000 €
- 2 = Entre 40000 € et moins de 75000 €
- 3 = Entre 75000 € et moins de 150000 €
- 4 = 150000 € et plus

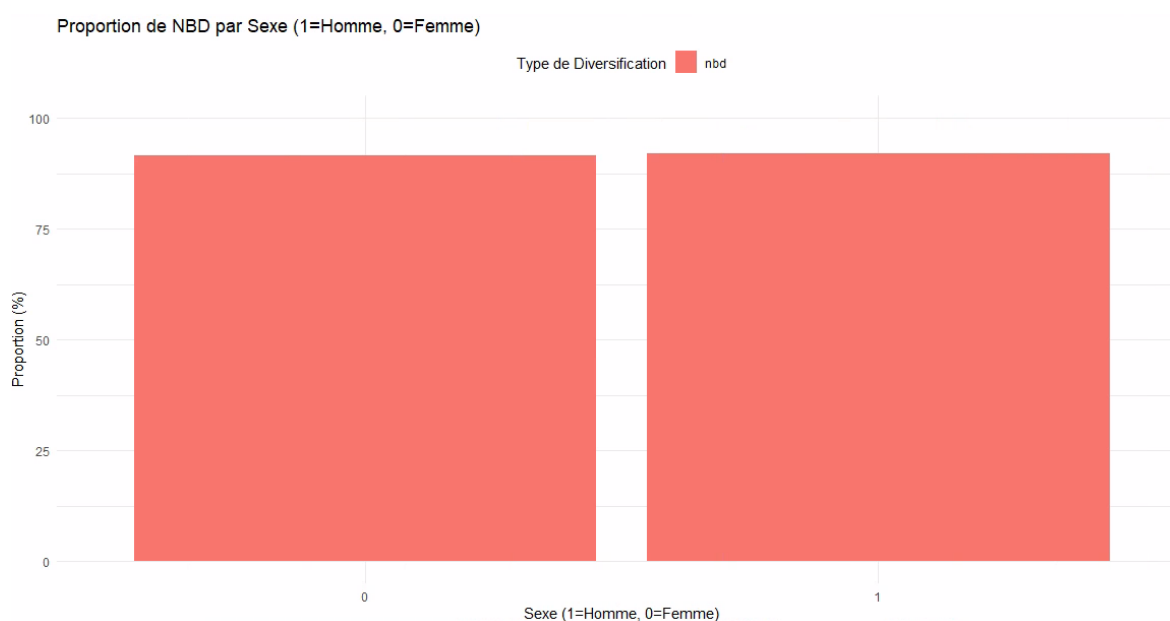
Annexe 9 : Table SP_Q3; Proportion de NBD par niveau de connaissance des marchés financiers



Légende pour l'axe des x :

- 1 = Très peu de connaissances et pas vraiment intéressé
- 2 = Non familier mais intéressé
- 3 = Suffisamment d'expérience pour reconnaître l'importance de la diversification des risques
- 4 = Bonne connaissance des marchés financiers et conscience que les marchés financiers peuvent fluctuer fortement, que les catégories de secteurs et d'actifs ont des caractéristiques différentes en matière de revenus, de croissance et de profil de risque
- 5 = Investisseur expérimenté qui maîtrise parfaitement tous les aspects des marchés financiers

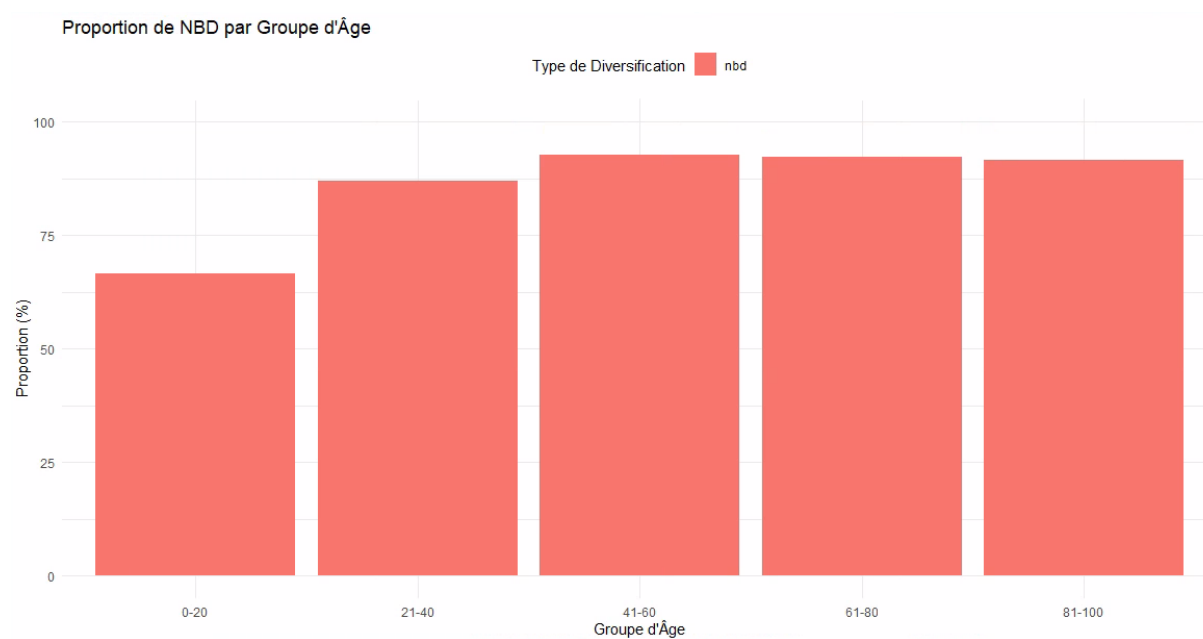
Annexe 10 : Table Proportion de NBD par Sexe



Légende pour l'axe des x :

- 0 = Femme
- 1 = Homme

Annexe 11 : Table Proportion de NBD par groupe d'âge



V. Codes utilisés & explications :

Annexe 12 : Code hypothèse 1

Installation et Chargement des Packages Nécessaires

Installer et charger les packages nécessaires

```
# install.packages(c("dplyr", "ggplot2", "knitr", "haven", "purrr", "tidyr"))
```

```
library(dplyr)
```

```
library(ggplot2)
```

```
library(knitr)
```

```
library(haven)
```

```
library(purrr)
```

```
library(tidyr)
```

Cette partie du code commence par la commande `install.packages`, qui est commentée, suggérant que les packages nécessaires doivent être installés si ce n'est pas déjà fait. Ensuite, les bibliothèques nécessaires sont chargées :

- `dplyr` et `tidyr` pour la manipulation des données.
- `ggplot2` pour la création de graphiques.
- `knitr` pour la création de tables formatées.
- `haven` pour lire les fichiers SAS.
- `purrr` pour appliquer des fonctions à des listes.

Chargement des Données

Charger les données

```
stock_trades <- read_sas("~/Document mémoire à modifier/Hypothèse  
1/stock_trades.sas7bdat", NULL)
```

Cette ligne charge les données de transactions boursières à partir d'un fichier SAS. Le chemin d'accès au fichier est spécifié, et la fonction `read_sas` de la bibliothèque `haven` est utilisée pour lire le fichier.

Calcul de la proportion de jours d'achats uniques versus multiples

```
# Filtrer les jours d'achat pour l'ensemble des investisseurs
```

```
purchase_data <- stock_trades %>%  
  
  filter(TransType == "B") %>%  
  
  group_by(User, Jour) %>%  
  
  mutate(num_stocks = n_distinct(ISIN),  
  
         total_value = sum(ExecValue_EUR)) %>%  
  
  ungroup()
```

```
# Séparer les jours d'achat en jours uniques et multiples
```

```
unique_days <- purchase_data %>%  
  
  filter(num_stocks == 1) %>%  
  
  summarise(unique_purchase_days = n_distinct(Jour))  
  
multiple_days <- purchase_data %>%  
  
  filter(num_stocks > 1) %>%  
  
  summarise(multiple_purchase_days = n_distinct(Jour))
```

```
# Créer un tableau comparatif des jours uniques et multiples
```

```
comparison_table <- data.frame(  
  
  Purchase_Type = c("Unique Purchase Days", "Multiple Purchase Days"),  
  
  Number_of_Days = c(unique_days$unique_purchase_days,  
multiple_days$multiple_purchase_days))
```

```
# Afficher le tableau comparatif
```

```
print(kable(comparison_table, format = "markdown", col.names = c("Type d'achat", "Nombre  
de Jours")))
```

Cette section de code est conçue pour calculer et afficher la proportion de jours où les investisseurs effectuent des achats uniques par rapport à ceux où ils effectuent des achats multiples.

1. Résumé des Jours d'Achats :

- **Groupement et Calcul :** Le code commence par regrouper les transactions par utilisateur (User) et par jour (Jour). Il calcule ensuite le nombre d'actions distinctes (num_stocks) achetées chaque jour.
- **Somme des Jours Uniques et Multiples :** À partir de ce calcul, le code détermine le nombre de jours où un seul titre a été acheté (Unique_Purchase_Days) et où plusieurs titres ont été achetés (Multiple_Purchase_Days).

2. Affichage des Résultats :

- **Construction du Résumé :** Les résultats sont synthétisés dans un tableau montrant la répartition entre les jours d'achats uniques et multiples.
- **Présentation des Données :** Le tableau final est affiché avec kable, fournissant une vue claire et formatée du nombre de jours pour chaque type d'achat.

Filtrage des Jours d'Achat Multiples

```
# Filtrer les jours d'achat multiples pour l'ensemble des investisseurs
```

```
nbd_data <- stock_trades %>%
```

```
  filter(TransType == "B") %>%
```

```
  group_by(User, Jour) %>%
```

```
  mutate(num_stocks = n_distinct(ISIN),
```

```
         total_value = sum(ExecValue_EUR)) %>%
```

```
  filter(num_stocks > 1) %>%
```

```
mutate(share_of_investment = ExecValue_EUR / total_value,
       equal_share = 1 / num_stocks) %>%
ungroup()
```

Ce bloc de code procède à plusieurs opérations :

1. Filtrage des Transactions d'Achat:

- `filter(TransType == "B")` filtre les transactions pour ne garder que les achats (B pour "Buy").

2. Groupement par Utilisateur et Jour:

- `group_by(User, Jour)` groupe les données par utilisateur et par jour.

3. Calcul du Nombre d'Actions et de la Valeur Totale:

- `mutate(num_stocks = n_distinct(ISIN), total_value = sum(ExecValue_EUR))` calcule le nombre d'actions distinctes achetées (`num_stocks`) et la valeur totale des achats (`total_value`) pour chaque utilisateur et chaque jour.

4. Filtrage des Jours avec Achats Multiples:

- `filter(num_stocks > 1)` ne garde que les jours où plus d'une action a été achetée.

5. Calcul des Parts d'Investissement:

- `mutate(share_of_investment = ExecValue_EUR / total_value, equal_share = 1 / num_stocks)` calcule la part de chaque investissement (`share_of_investment`) et la part égale théorique si l'investissement était réparti équitablement (`equal_share`).

6. Désagrégation des Données:

- `ungroup()` désagrège les données pour enlever les groupements précédents.

Fonction pour Créer des Graphiques et des Tableaux

Fonction pour créer les graphiques et les tables

```
create_plot_and_table <- function(data, n) {
  data_filtered <- data %>%
```

```

filter(num_stocks == n)

# Créer le tableau de fréquence

freq_table <- data_filtered %>%

  mutate(share_of_investment_bin = floor(share_of_investment * 100)) %>%

  group_by(share_of_investment_bin) %>%

  summarise(Frequency = n() / nrow(data_filtered) * 100, Buy_Days = n_distinct(Jour)) %>%

  ungroup() %>%

  complete(share_of_investment_bin = 0:100, fill = list(Frequency = 0, Buy_Days = 0))

# Afficher le tableau

print(kable(freq_table, format = "markdown", col.names = c("Percentage", "Frequency (%)",
"Buy Days")))

# Créer le graphique

plot <- ggplot(data_filtered, aes(x = share_of_investment)) +

  geom_histogram(aes(y = (..count..)/sum(..count..)*100), binwidth = 0.01, fill = "blue", color
= "black") +

  geom_vline(aes(xintercept = 1/n), color = "red", linetype = "dashed") +

  scale_y_continuous(labels = scales::percent_format(scale = 1), breaks = seq(0, 100, 5)) +

  scale_x_continuous(limits = c(0, 1)) +

  labs(title = paste0("Proportion of Buy Day Investment for N = ", n),

    x = "Proportion of Investment in Each Stock",

    y = "Proportion of Buy Days (%)") +

  theme_minimal() +

```

```

theme(

  strip.text = element_text(size = 12),

  axis.text.x = element_text(angle = 45, hjust = 1),

  panel.spacing = unit(2, "lines")

)

return(list(plot = plot, table = freq_table, total_buy_days = n_distinct(data_filtered$Jour)))
}

```

Cette fonction `create_plot_and_table` accomplit plusieurs tâches :

1. Filtrage des Données pour le Nombre d'Actions Donné (n):

- `data_filtered <- data %>% filter(num_stocks == n)` filtre les données pour les jours où exactement `n` actions ont été achetées.

2. Création d'un Tableau de Fréquence:

- `mutate(share_of_investment_bin = floor(share_of_investment * 100))` transforme les proportions d'investissement en pourcentages entiers.
- `group_by(share_of_investment_bin) %>% summarise(Frequency = n() / nrow(data_filtered) * 100, Buy_Days = n_distinct(Jour))` calcule la fréquence de chaque bin d'investissement.

1. un "bin" fait référence à une catégorie ou un intervalle spécifique pour les proportions d'investissement.
2. cette ligne de code signifie que je regroupe les données par ces intervalles (`share_of_investment_bin`), puis je calcule la fréquence de chaque intervalle (bin). La fréquence est le pourcentage du nombre total d'observations (`n() / nrow(data_filtered) * 100`) qui tombent dans chaque intervalle. Le nombre de jours d'achat distincts (`Buy_Days = n_distinct(Jour)`) est également calculé pour chaque bin.

- `complete(share_of_investment_bin = 0:100, fill = list(Frequency = 0, Buy_Days = 0))` complète les bins manquants avec des fréquences et des jours d'achat de zéro.

3. Affichage du Tableau:

- `print(kable(freq_table, format = "markdown", col.names = c("Percentage", "Frequency (%)", "Buy Days")))` affiche le tableau de fréquence formaté en Markdown.

1. Markdown est un langage de balisage léger utilisé pour formater du texte de manière simple.

4. Création d'un Graphique:

- `geom_histogram(aes(y = (..count..)/sum(..count..)*100), binwidth = 0.01, fill = "blue", color = "black")` crée un histogramme des parts d'investissement.
- `geom_vline(aes(xintercept = 1/n), color = "red", linetype = "dashed")` ajoute une ligne verticale indiquant la part égale théorique (1/n).
- `scale_y_continuous(labels = scales::percent_format(scale = 1), breaks = seq(0, 100, 5))` formate l'axe des ordonnées pour afficher des pourcentages.
- `scale_x_continuous(limits = c(0, 1))` fixe les limites de l'axe des abscisses de 0 à 1.
- `labs(title = paste0("Proportion of Buy Day Investment for N = ", n), x = "Proportion of Investment in Each Stock", y = "Proportion of Buy Days (%)")` ajoute des titres et labels aux axes.
- `theme_minimal() + theme(strip.text = element_text(size = 12), axis.text.x = element_text(angle = 45, hjust = 1), panel.spacing = unit(2, "lines"))` applique un thème minimaliste et améliore la lisibilité des textes de l'axe des abscisses.

5. Retourne les Résultats:

- `return(list(plot = plot, table = freq_table, total_buy_days = n_distinct(data_filtered$Jour)))` retourne une liste contenant le graphique, le tableau de fréquence et le nombre total de jours d'achat.

Calcul des Pourcentages de NBD pour Différentes Tolérances

```
# Fonction pour calculer les pourcentages de 1/N buying pour différentes tolérances

calculate_tolerance_table <- function(freq_table, n, tolerance, total_buy_days) {

  if (tolerance == 0.02) {

    lower_bound <- (1/n - 0.01) * 100

    upper_bound <- (1/n + 0.01) * 100

  } else if (tolerance == 0.05) {

    lower_bound <- (1/n - 0.025) * 100

    upper_bound <- (1/n + 0.025) * 100

  } else if (tolerance == 0.1) {

    lower_bound <- (1/n - 0.05) * 100

    upper_bound <- (1/n + 0.05) * 100

  }

  tolerance_table <- freq_table %>%

    filter(share_of_investment_bin >= lower_bound & share_of_investment_bin <=

upper_bound) %>%

    summarise(`1/N Buying (%)` = sum(Frequency))

  tolerance_table <- data.frame(

    N = n,

    `1/N Buying (%)` = tolerance_table$`1/N Buying (%)`,

    `Buy Days` = total_buy_days

  )

  return(tolerance_table)}


```

Cette fonction `calculate_tolerance_table` calcule les pourcentages de NBD pour différentes tolérances :

1. Détermination des Bornes de Tolérance:

- En fonction de la tolérance (0.02, 0.05, ou 0.1), les bornes inférieures et supérieures sont ajustées pour englober les proportions d'investissement autour de $1/n$.

2. Filtrage des Données selon la Tolérance:

- `filter(share_of_investment_bin >= lower_bound & share_of_investment_bin <= upper_bound)` filtre le tableau de fréquence pour inclure uniquement les parts d'investissement dans les bornes de tolérance.

3. Calcul des Pourcentages de NBD:

- `summarise(1/N Buying (%) = sum(Frequency))` calcule la somme des fréquences pour obtenir le pourcentage de NBD.

4. Création du Tableau de Tolérance:

- `tolerance_table <- data.frame(N = n, 1/N Buying (%) = tolerance_table$1/N Buying (%), Buy Days = total_buy_days)` crée un tableau de tolérance contenant le nombre d'actions, le pourcentage de NBD, et le nombre total de jours d'achat.

5. Retourne les Résultats:

- `return(tolerance_table)` retourne le tableau de tolérance.

Génération des Graphiques et des Tableaux pour N = 2 à N = 5

Générer et afficher les graphiques et les tables pour N = 2 à N = 5 pour l'ensemble des investisseurs

```
results <- map(2:5, ~create_plot_and_table(nbd_data, .x))
```

```
for (result in results) {
```

```
  print(result$plot)
```

```

    print(kable(result$table, format = "markdown", col.names = c("Percentage", "Frequency
(%)", "Buy Days")))
  }
}

```

Ce segment de code utilise la fonction `map` de la bibliothèque `purrr` pour appliquer la fonction `create_plot_and_table` à chaque valeur de `N` allant de 2 à 5 :

1. Application de la Fonction `create_plot_and_table`:

- `results <- map(2:5, ~create_plot_and_table(nbd_data, .x))` applique la fonction à chaque valeur de `N` et stocke les résultats dans la liste `results`.

2. Affichage des Graphiques et des Tableaux:

- `for (result in results) { ... }` parcourt les résultats et affiche chaque graphique et tableau de fréquence.

Calcul et Affichage des Tableaux de Tolérance

Calculer les tableaux de tolérance pour `N = 2` à `N = 5`

```
tolerances <- c(0.02, 0.05, 0.1)
```

```
tolerance_tables <- lapply(tolerances, function(tol) {
```

```
  tolerance_results <- map_df(2:5, function(n) {
```

```
    freq_table <- results[[n-1]]$table
```

```
    total_buy_days <- results[[n-1]]$total_buy_days
```

```
    calculate_tolerance_table(freq_table, n, tol, total_buy_days)
```

```
  })
```

```
  return(tolerance_results)
```

```
})
```

Afficher les tableaux de tolérance

```
names(tolerance_tables) <- paste0("Tolerance ±", tolerances * 100, "%")
```

```
for (i in 1:length(tolerance_tables)) {
```

```
cat("\nTable for ", names(tolerance_tables)[i], "\n")

print(kable(tolerance_tables[[i]], format = "markdown", col.names = c("N", "1/N Buying
(%)", "Buy Days"))))}
```

Ce segment de code calcule les tableaux de tolérance et les affiche :

1. Définition des Tolérances:

- `tolerances <- c(0.02, 0.05, 0.1)` définit les tolérances utilisées pour mesurer la NBD.

2. Calcul des Tableaux de Tolérance:

- `tolerance_tables <- lapply(tolerances, function(tol) { ... })` utilise `lapply` pour appliquer la fonction `calculate_tolerance_table` à chaque tolérance.

3. Application de la Fonction `calculate_tolerance_table`:

- `tolerance_results <- map_df(2:5, function(n) { ... })` applique la fonction à chaque valeur de `N` allant de 2 à 5.

4. Stockage et Affichage des Résultats:

- `names(tolerance_tables) <- paste0("Tolerance ±", tolerances * 100, "%")` nomme les tableaux de tolérance.
- `for(i in 1:length(tolerance_tables)) { ... }` parcourt les résultats et affiche chaque tableau de tolérance formaté en Markdown.

[Annexe 13 : Code hypothèse 2](#)

Charger les bibliothèques nécessaires

Les bibliothèques nécessaires pour manipuler les données et créer des tableaux sont chargées.

```
library(dplyr)
```

```
library(tidyr)
```

```
library(haven)
```

```
library(purrr)
```

```
library(knitr)
```

Charger les données

Les données de transactions (stock_trades) et les données de portefeuilles (stock_holdings) sont chargées à partir de fichiers SAS.

```
stock_trades <- read_sas("~/Document mémoire à modifier/Hypothèse  
2/stock_trades.sas7bdat", NULL)
```

```
stock_holdings <- read_sas("~/Document mémoire à modifier/Hypothèse  
2/stock_holdings.sas7bdat", NULL)
```

Conversion des valeurs numériques en dates

Les dates dans les fichiers de données sont converties en format de date approprié. La date de référence est définie comme le 1er mars 2006, correspondant à une valeur de 16861 jours.

```
reference_date <- as.Date("2006-03-01")
```

```
days_difference <- 16861
```

```
origin_date <- reference_date - days_difference
```

```
stock_holdings <- stock_holdings %>%
```

```
  mutate(Time = as.Date(Time, origin = origin_date))
```

```
stock_trades <- stock_trades %>%
```

```
  mutate(Jour = as.Date(Jour, origin = origin_date))
```

Cette conversion est nécessaire car lors de l'importation du fichier stock_holdings dans R, le format de la colonne Time ne s'importe pas correctement. En convertissant ces nombres en dates, je facilite l'analyse des transactions et des portefeuilles.

Sélection d'un échantillon d'investisseurs

Un échantillon de 1000 investisseurs est sélectionné aléatoirement pour l'analyse. Cette étape est importante pour réduire la taille des données à un niveau gérable tout en conservant une diversité suffisante pour des conclusions robustes.

```
investisseurs_ids <- unique(stock_trades$User)

set.seed(123) # Pour garantir la reproductibilité de l'échantillonnage

sampled_investors <- sample(investisseurs_ids, 1000)
```

L'utilisation de `set.seed(123)` garantit que le même échantillon est sélectionné chaque fois que le code est exécuté, assurant ainsi la reproductibilité des résultats.

Définir les tolérances

Les tolérances de $\pm 2\%$, $\pm 5\%$, et $\pm 10\%$ sont définies pour vérifier la diversification $1/N$. Ces tolérances permettent de capturer différentes marges d'erreur possibles dans les décisions d'investissement, en suivant l'approche de Gathergood et al.

```
tolerances <- c(0.02, 0.05, 0.10)
```

Fonction pour vérifier la diversification $1/N$

La fonction `is_npd` vérifie si un portefeuille suit une stratégie de diversification $1/N$ en fonction de la tolérance spécifiée.

```
is_npd <- function(portfolio, tolerance) {

  n <- length(portfolio)

  target <- 1 / n

  deviations <- abs(portfolio / sum(portfolio) - target)

  all(deviations <= tolerance)

}
```

Cette fonction calcule la déviation de chaque position par rapport à une allocation égale et vérifie si toutes les déviations sont inférieures ou égales à la tolérance spécifiée. Si c'est le cas, le portefeuille suit la stratégie de diversification $1/N$.

Fonction pour calculer les proportions des jours d'achat

La fonction `calculate_proportions` a pour objectif de déterminer si les portefeuilles des investisseurs suivent la stratégie de diversification 1/N (NPD) avant, pendant et après les jours d'achats multiples. Elle calcule les proportions de jours d'achat correspondant à chaque combinaison d'états de portefeuille.

```
calculate_proportions <- function(trades_data, holdings_data, tolerance) {
```

Génération de toutes les combinaisons possibles

La première étape consiste à générer toutes les combinaisons possibles des états du portefeuille pour les positions initiales, les achats et les positions finales. Chaque état peut être "Oui" (le portefeuille suit la stratégie 1/N) ou "Non" (le portefeuille ne suit pas la stratégie 1/N).

```
all_combinations <- expand.grid(  
  initial_in_1_over_n = c("Oui", "Non"),  
  buy_in_1_over_n = c("Oui", "Non"),  
  final_in_1_over_n = c("Oui", "Non")  
)
```

Cette combinaison exhaustive permet de s'assurer que toutes les configurations possibles sont prises en compte lors de l'analyse des jours d'achats multiples.

Filtrage des jours avec plusieurs transactions d'achat

On filtre les jours où les investisseurs ont effectué plusieurs transactions d'achat sur différentes actions (ISIN). Cela permet de se concentrer uniquement sur les jours où la diversification est pertinente.

```
multiple_buy_days <- trades_data %>%  
  group_by(Jour) %>%  
  filter(n_distinct(ISIN) > 1) %>%  
  ungroup()
```

Si aucun jour d'achat multiple n'est trouvé, la fonction retourne NULL.

```
if (nrow(multiple_buy_days) == 0) {  
  return(NULL)  
}
```

Initialiser une liste pour stocker les résultats

Une liste vide est initialisée pour stocker les résultats de chaque jour d'achat multiple.

```
results_list <- list()
```

Boucle sur chaque jour d'achat multiple

Pour chaque jour d'achat multiple, l'état du portefeuille initial, l'état des achats et l'état final du portefeuille sont calculés et vérifiés.

```
for (day in unique(multiple_buy_days$Jour)) {  
  daily_trades <- multiple_buy_days %>%  
    filter(Jour == day) %>%  
    group_by(ISIN) %>%  
    summarise(ExecValue_EUR = sum(ExecValue_EUR, na.rm = TRUE)) %>%  
    ungroup()
```

Calcul de l'état du portefeuille initial

L'état du portefeuille avant les achats est déterminé en filtrant les données des portefeuilles pour les transactions antérieures à la date d'achat.

```
start_portfolio <- holdings_data %>%  
  filter(as.Date(Time) < as.Date(day)) %>%  
  group_by(isin) %>%  
  summarise(MV_pos = sum(MV_pos, na.rm = TRUE)) %>%  
  pull(MV_pos)
```

L'état initial est vérifié pour savoir s'il suit la stratégie 1/N avec la tolérance spécifiée.

```
initial_npd <- ifelse(is_npd(start_portfolio, tolerance), "Oui", "Non")
```

Calcul de l'état des achats

Les valeurs des achats effectués le jour en question sont extraites et vérifiées pour savoir si elles suivent la stratégie 1/N.

```
buy_values <- daily_trades$ExecValue_EUR
```

```
buy_nbd <- ifelse(is_npd(buy_values, tolerance), "Oui", "Non")
```

Calcul de l'état final du portefeuille

L'état final du portefeuille après les achats est déterminé en ajoutant les nouvelles valeurs des achats aux positions initiales et en vérifiant à nouveau si elles suivent la stratégie 1/N.

```
end_portfolio <- c(start_portfolio, buy_values)
```

```
final_npd <- ifelse(is_npd(end_portfolio, tolerance), "Oui", "Non")
```

Stocker les résultats

Les résultats pour chaque jour sont stockés dans une liste, incluant la date et les états du portefeuille avant, pendant et après les achats.

```
results_list[[as.character(day)]] <- data.frame(  
  Jour = day,  
  initial_in_1_over_n = initial_npd,  
  buy_in_1_over_n = buy_nbd,  
  final_in_1_over_n = final_npd  
)
```

Combinaison des résultats

Les résultats de tous les jours d'achat multiples sont combinés en un seul data

```
results_df <- do.call(rbind, results_list)
```

Calcul des proportions

Les proportions de chaque combinaison d'états sont calculées. On compte le nombre d'occurrences de chaque combinaison d'états et on calcule la proportion de jours d'achat correspondants.

```
proportions <- results_df %>%  
  
  group_by(initial_in_1_over_n, buy_in_1_over_n, final_in_1_over_n) %>%  
  
  summarise(count = n(), .groups = "drop") %>%  
  
  right_join(all_combinations, by = c("initial_in_1_over_n", "buy_in_1_over_n",  
"final_in_1_over_n")) %>%  
  
  replace_na(list(count = 0)) %>%  
  
  mutate(proportion_of_buy_days = count / sum(count) * 100) %>%  
  
  arrange(desc(proportion_of_buy_days)) %>%  
  
  select(  
  
    `1/N Existing Positions` = initial_in_1_over_n,  
  
    `1/N Buying` = buy_in_1_over_n,  
  
    `1/N Resulting Positions` = final_in_1_over_n,  
  
    `Proportion of Buy Days (%)` = proportion_of_buy_days  
  
  )
```

Retourner les résultats

La fonction retourne les proportions calculées et les jours d'achats multiples analysés.

```
list(proportions = proportions, multiple_buy_days = unique(results_df$Jour))
```

Fonction pour combiner les résultats de plusieurs investisseurs

La fonction `combine_multiple_results` combine les résultats de plusieurs investisseurs en une seule table.

```
combine_multiple_results <- function(results_list) {  
  
  combined <- results_list[[1]]
```

```

for (i in 2:length(results_list)) {

  combined <- combined %>%

    full_join(results_list[[i]], by = c("1/N Existing Positions", "1/N Buying", "1/N Resulting
Positions"), suffix = c("", paste0(".y", i))) %>%

    mutate(`Proportion of Buy Days (%)` = coalesce(`Proportion of Buy Days (%)`, 0) +
coalesce(!sym(paste0("Proportion of Buy Days (%)y", i)), 0)) %>%

    select(-matches(paste0("Proportion of Buy Days (%)y", i)))}

combined <- combined %>%

  mutate(`Proportion of Buy Days (%)` = `Proportion of Buy Days (%)` / length(results_list))

return(combined)
}

```

Cette fonction prend les résultats individuels des investisseurs et les combine en une seule table en additionnant les proportions de jours d'achat pour chaque combinaison d'états et en divisant par le nombre d'investisseurs pour obtenir une moyenne.

Calculer les proportions pour différentes tolérances pour chaque investisseur

Les proportions sont calculées pour chaque investisseur échantillonné et chaque tolérance.

```

results_investors <- lapply(sampled_investors, function(id) {

  trades_data <- stock_trades %>% filter(User == id & TransType == "B")

  if (nrow(trades_data) == 0) {

    return(NULL)

  }

  res <- map(tolerances, ~ calculate_proportions(trades_data, stock_holdings, .x))

  if (all(sapply(res, is.null))) {

    return(NULL)

  }
}

```

```

    return(res)
  })

```

```

results_investors <- results_investors[!sapply(results_investors, is.null)]

```

Cette étape filtre les transactions pour chaque investisseur, calcule les proportions de jours d'achat suivant les tolérances spécifiées, et stocke les résultats dans une liste.

Combiner les résultats pour tous les investisseurs

Les résultats de tous les investisseurs sont combinés pour chaque tolérance.

```

combined_results <- list()

for (i in 1:length(tolerances)) {

  non_null_results <- lapply(results_investors, function(res) res[[i]]$proportions)

  non_null_results <- non_null_results[!sapply(non_null_results, is.null)]

  combined_results[[i]] <- combine_multiple_results(non_null_results)

}

```

```

names(combined_results) <- paste0("Tolérance: ", tolerances * 100, "%")

for (i in seq_along(combined_results)) {

  cat("\nTableau pour la ", names(combined_results)[i], "\n")

  print(knitr::kable(combined_results[[i]][, 1:4], format = "markdown"))

}

```

```

combined_buy_days <- unique(unlist(lapply(results_investors, function(res)
res[[i]]$multiple_buy_days))))

```

```

cat("\nNombre de jours d'achats multiples pour l'échantillon d'investisseurs :",
length(combined_buy_days), "\n")

```

```

cat("Dates des jours d'achats multiples : ", paste(combined_buy_days, collapse = ", "), "\n")

```

Cette dernière étape :

1. **Combine les résultats non nuls** pour chaque tolérance.
2. **Affiche les résultats** sous forme de tableau pour chaque tolérance.
3. **Calcule et affiche le nombre total de jours d'achat multiples** pour l'échantillon d'investisseurs, donnant un aperçu de la fréquence et de la répartition des jours d'achat multiples.

Annexe 14 : Code hypothèse 3

Chargement des packages nécessaires

Je commence par charger les bibliothèques requises pour la manipulation des données, les calculs et la visualisation.

Charger les packages nécessaires

```
library(dplyr)
```

```
library(ggplot2)
```

```
library(knitr)
```

```
library(haven)
```

```
library(lubridate)
```

- dplyr : Pour la manipulation des données.
- ggplot2 : Pour la visualisation des données.
- knitr : Pour la création de tableaux.
- haven : Pour lire les fichiers SAS.
- purrr : Pour les opérations fonctionnelles.
- lubridate : Pour faciliter la manipulation des dates.

Chargement des données

Je charge les données de transactions (stock_trades) et les rendements des portefeuilles (ptf_returns) à partir de fichiers SAS.

Charger les données

```
stock_trades <- read_sas("~/Document mémoire à modifier/Hypothèse 3/stock_trades.sas7bdat", NULL)
```

```
ptf_returns <- read_sas("~/Document mémoire à modifier/Hypothèse 3/ptf_returns.sas7bdat", NULL)
```

- stock_trades : Contient les transactions d'achat et de vente des investisseurs.
- ptf_returns : Contient les rendements des portefeuilles des investisseurs.

Conversion des Jours en Dates Correctes

Les jours sont convertis en dates correctes en ajustant l'origine des dates.

Convertir les jours en dates correctes

```
reference_date <- as.Date("2003-01-01")
```

```
days_difference <- 15706
```

```
origin_date <- reference_date - days_difference
```

```
ptf_returns <- ptf_returns %>%
```

```
mutate(Time = as.Date(Time, origin = origin_date))
```

- reference_date : La date de référence utilisée pour calculer l'origine des dates.
- days_difference : La différence en jours utilisée pour ajuster l'origine des dates.
- mutate(Time = as.Date(Time, origin = origin_date)) : Convertit la colonne Time en dates appropriées en fonction de l'origine calculée.

Filtrage des Jours d'Achat Multiples

Je filtre les transactions pour inclure uniquement les jours où un investisseur a acheté plus d'une action, et calcule les proportions d'investissement pour chaque action.

Filtrer les jours d'achat multiples

```
nbd_npd_data <- stock_trades %>%
```

```
filter(TransType == "B") %>%
```

```
group_by(User, Jour) %>%
```

```
mutate(num_stocks = n_distinct(ISIN),

       total_value = sum(ExecValue_EUR)) %>%

filter(num_stocks > 1) %>%

mutate(share_of_investment = ExecValue_EUR / total_value,

       equal_share = 1 / num_stocks) %>%

ungroup()
```

- filter(TransType == "B") : Sélectionne uniquement les transactions d'achat.
- group_by(User, Jour) : Regroupe les transactions par utilisateur et par jour.
- mutate(num_stocks = n_distinct(ISIN), total_value = sum(ExecValue_EUR)) : Calcule le nombre d'actions distinctes achetées et la valeur totale des transactions pour chaque jour.
- filter(num_stocks > 1) : Filtre les jours où plus d'une action a été achetée.
- mutate(share_of_investment = ExecValue_EUR / total_value, equal_share = 1 / num_stocks) : Calcule la proportion d'investissement pour chaque action et la part égale théorique (1/N).

Calcul des Proportions pour NBD

Je détermine si les investissements respectent la règle 1/N (NBD) en comparant la proportion d'investissement avec une tolérance de 2%.

Calcul des proportions pour NBD

```
nbd_performance <- nbd_npd_data %>%

mutate(NBD = abs(share_of_investment - equal_share) < 0.02) %>%

group_by(User) %>%

summarize(

  NBD_days = sum(NBD),

  total_days_multiple = n(),

  total_days = n_distinct(Jour),

  NBD_ratio = NBD_days / total_days_multiple
```

```

) %>%

left_join(ptf_returns %>%

  group_by(User) %>%

  summarize(

    first_date = min(Time[!is.na(g_return)], na.rm = TRUE),

    last_date = max(Time[!is.na(g_return)], na.rm = TRUE),

    holding_period = as.numeric(difftime(last_date, first_date, units = "days")),

    geo_mean_g_return = (prod(1 + g_return[!is.na(g_return)], na.rm =
TRUE))^(1/sum(!is.na(g_return))) - 1,

    geo_mean_n_return = (prod(1 + n_return[!is.na(n_return)], na.rm =
TRUE))^(1/sum(!is.na(n_return))) - 1

  ),

  by = "User")

```

- `mutate(NBD = abs(share_of_investment - equal_share) < 0.02)` : Vérifie si la proportion d'investissement respecte la règle 1/N avec une tolérance de 2%.
- `group_by(User)` : Regroupe les données par utilisateur pour calculer des statistiques individuelles.
- `summarize(NBD_days = sum(NBD), total_days_multiple = n(), total_days = n_distinct(Jour), NBD_ratio = NBD_days / total_days_multiple)` : Calcule le nombre de jours NBD, le nombre total de jours d'achats multiples, et le ratio NBD pour chaque utilisateur.
- `left_join(ptf_returns %>% ...)` : Joint les rendements des portefeuilles aux données NBD par utilisateur, calculant la période de détention et les rendements géométriques moyens.

Classification des Investisseurs

Les investisseurs sont classés en trois groupes selon leur utilisation de la NBD.

```
# Classifier les investisseurs
```

```
nbdd_performance <- nbdd_performance %>%
```

```
mutate(
  group = case_when(
    NBD_ratio >= 0.5 ~ "50% or More NBD",
    NBD_ratio > 0 & NBD_ratio < 0.5 ~ "At Least One NBD",
    NBD_ratio == 0 ~ "Never NBD"
  )
)
```

`mutate(group = case_when(...))` : Crée une colonne `group` pour classer les investisseurs :

- 50% or More NBD : Investisseurs utilisant la NBD dans au moins 50% de leurs jours d'achat multiples.
- At Least One NBD : Investisseurs utilisant la NBD au moins une fois mais moins de 50% du temps.
- Never NBD : Investisseurs n'ayant jamais utilisé la NBD.

Calcul des Rendements Géométriques Moyens sur la Période de Détention

Je calcule les rendements géométriques annuels moyens pour chaque groupe d'investisseurs.

Calculer les rendements géométriques moyens sur la période de détention

```
annualized_returns <- nbd_performance %>%
  filter(!is.na(geo_mean_g_return)) %>%
  group_by(group) %>%
  summarize(
    mean_geo_return = mean(geo_mean_g_return, na.rm = TRUE) * 12,
    sd_geo_return = sd(geo_mean_g_return, na.rm = TRUE) * sqrt(12),
    avg_holding_period = mean(holding_period, na.rm = TRUE),
    count = n(),
    total_purchases = sum(total_days, na.rm = TRUE))
```

- `filter(!is.na(geo_mean_g_return))` : Filtre les valeurs manquantes pour les rendements géométriques moyens.
- `group_by(group)` : Regroupe les données par groupe d'investisseurs.
- `summarize(mean_geo_return = mean(geo_mean_g_return, na.rm = TRUE) * 12, sd_geo_return = sd(geo_mean_g_return, na.rm = TRUE) * sqrt(12), avg_holding_period = mean(holding_period, na.rm = TRUE), count = n(), total_purchases = sum(total_days, na.rm = TRUE))` : Calcule les rendements annuels moyens géométriques, l'écart-type annuel, la période de détention moyenne, le nombre d'investisseurs et le nombre total d'achats pour chaque groupe.

Affichage des Résultats

Les résultats sont présentés sous forme de tableau et de graphique.

Afficher les résumés sous forme de tableau

```
print(kable(annualized_returns, format = "markdown", col.names = c("Groupe", "Rendement
Géométrique Annuel Moy. (%)", "Écart-Type Annuel", "Période de Détention Moyenne
(jours)", "Nombre d'Investisseurs", "Nombre Total d'Achats"))))
```

Créer un graphique des rendements géométriques annuels moyens

```
ggplot(annualized_returns, aes(x = group, y = mean_geo_return)) +
  geom_point(size = 4) +
  geom_errorbar(aes(ymin = mean_geo_return - 1.96 * sd_geo_return,
                    ymax = mean_geo_return + 1.96 * sd_geo_return), width = 0.2) +
  labs(title = "Rendements Géométriques Annualisés pour Investisseurs NBD et Non-NBD",
        x = "",
        y = "Rendement Géométrique Annuel Moyen (%)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

- `print(kable(...))` : Affiche les résultats sous forme de tableau formaté en Markdown.

- `ggplot(annualized_returns, aes(x = group, y = mean_geo_return)) + ...` : Crée un graphique montrant les rendements annuels moyens géométriques pour chaque groupe avec des barres d'erreur représentant les intervalles de confiance à 95%.

Test t de Welch pour Comparer les Moyennes

J'effectue un test t de Welch pour comparer les moyennes des rendements géométriques entre les groupes.

Test t de Welch pour comparer les moyennes

```
welch_test <- pairwise.t.test(nbd_performance$geo_mean_g_return, nbd_performance$group,
p.adjust.method = "bonferroni", pool.sd = FALSE)
```

```
print(welch_test)
```

- `pairwise.t.test(...)` : Effectue un test t de Welch pour comparer les moyennes entre les groupes, en ajustant les p-values avec la méthode de Bonferroni.

Histogrammes des Distributions de Rendement pour Chaque Groupe

Je crée des histogrammes pour visualiser la distribution des rendements dans chaque groupe.

Histogrammes des distributions de rendement pour chaque groupe

```
ggplot(nbd_performance %>% filter(!is.na(geo_mean_g_return)), aes(x = geo_mean_g_return,
fill = group)) +
```

```
geom_histogram(alpha = 0.6, position = "identity", binwidth = 0.005) +
```

```
facet_wrap(~ group, scales = "free") +
```

```
labs(title = "Distribution des Rendements Géométriques Moyens par Groupe",
```

```
x = "Rendement Géométrique Moyen",
```

```
y = "Fréquence") +
```

```
scale_x_continuous(limits = c(-0.1, 0.1)) +
```

```
theme_minimal()
```

- `ggplot(nbd_performance %>% filter(!is.na(geo_mean_g_return)), aes(x = geo_mean_g_return, fill = group))` : Crée un graphique en utilisant les rendements moyens géométriques bruts, filtrés pour exclure les valeurs manquantes.

- `geom_histogram(alpha = 0.6, position = "identity", binwidth = 0.005)` : Ajoute des histogrammes avec une transparence de 0.6 et une largeur de bin de 0.005.
- `facet_wrap(~ group, scales = "free")` : Crée des facettes pour chaque groupe, permettant des échelles indépendantes pour chaque histogramme.
- `labs(...)` : Ajoute des étiquettes pour le titre, l'axe x, et l'axe y.
- `scale_x_continuous(limits = c(-0.1, 0.1))` : Fixe les limites de l'axe des x pour uniformiser les échelles.

Annexe 15 : Code hypothèse 4

Installation et chargement des packages nécessaires

Les bibliothèques `dplyr`, `ggplot2`, `broom`, `tidyr`, `haven` et `knitr` sont utilisées pour la manipulation de données, la visualisation, et l'analyse statistique.

Installer et charger les packages nécessaires

install.packages(c("dplyr", "ggplot2", "broom", "tidyr", "haven"))

library(dplyr)

library(ggplot2)

library(broom)

library(tidyr)

library(haven)

library(knitr)

Chargement des données

Les fichiers de données `sample_investors_clean` et `stock_trades` sont chargés à partir des fichiers SAS.

Charger les données

```
sample_investors_clean <- read_sas("~/Document mémoire à modifier/Hypothèse  
4/sample_investors_clean.sas7bdat", NULL)
```

```
stock_trades <- read_sas("~/Document mémoire à modifier/Hypothèse  
4/stock_trades.sas7bdat", NULL)
```

Calcul de l'âge des investisseurs

L'âge des investisseurs est calculé à partir de leur année de naissance.

```
# Calculer l'âge des investisseurs
```

```
current_year <- as.numeric(format(Sys.Date(), "%Y"))
```

```
sample_investors_clean <- sample_investors_clean %>%
```

```
  mutate(Age = current_year - DateBirth)
```

Filtrer les âges irréalistes

Les âges supérieurs à 100 ans sont exclus pour garantir la cohérence des données.

```
# Filtrer les âges pour ne pas inclure les âges irréalistes
```

```
sample_investors_clean <- sample_investors_clean %>%
```

```
  filter(Age <= 100)
```

Catégoriser les groupes d'âge

Les âges sont divisés en groupes pour l'analyse.

```
# Catégoriser les groupes d'âge
```

```
sample_investors_clean <- sample_investors_clean %>%
```

```
  mutate(Age_Group = cut(Age, breaks = c(0, 20, 40, 60, 80, 100),
```

```
    labels = c("0-20", "21-40", "41-60", "61-80", "81-100"),
```

```
    right = FALSE))
```

- `cut(Age, breaks = c(0, 20, 40, 60, 80, 100), labels = c("0-20", "21-40", "41-60", "61-80", "81-100"))` : Divise l'âge en catégories définies par des intervalles.

Affichage de la Répartition par Groupe d'Âge

Un tableau est créé pour montrer le nombre d'investisseurs dans chaque catégorie d'âge.

Afficher un tableau du nombre d'investisseurs dans chaque catégorie d'âge

```
age_distribution <- sample_investors_clean %>%
```

```
  group_by(Age_Group) %>%
```

```
  summarise(Count = n())
```

```
cat("Répartition des investisseurs par groupe d'âge:\n")
```

```
print(kable(age_distribution, format = "markdown", col.names = c("Groupe d'Âge", "Nombre  
d'Investisseurs")))
```

- `group_by(Age_Group) %>% summarise(Count = n())` : Compte le nombre d'investisseurs dans chaque groupe d'âge.
- `kable(...)` : Formate les résultats en tableau Markdown.

Conversion de la Variable Sexe en Binaire

La variable sexe est convertie en binaire pour simplifier l'analyse statistique.

Convertir la variable sexe en binaire (1 pour homme, 0 pour femme)

```
sample_investors_clean <- sample_investors_clean %>%
```

```
  filter(Sex != "U") %>% # Supprimer les cas où le sexe est inconnu
```

```
  mutate(Sex_Male = ifelse(Sex == "M", 1, 0))
```

- `filter(Sex != "U")` : Exclut les enregistrements avec un sexe inconnu.
- `mutate(Sex_Male = ifelse(Sex == "M", 1, 0))` : Convertit le sexe en binaire (1 pour les hommes, 0 pour les femmes).

Calcul de la Proportion de Jours NBD pour Chaque Investisseur

Je calcule la proportion de jours où chaque investisseur utilise la NBD.

Calculer la proportion de jours NBD pour chaque investisseur

```
merged_data <- stock_trades %>%
```

```
  filter(TransType == "B") %>%
```

```

group_by(User, Jour) %>%
mutate(num_stocks = n_distinct(ISIN),
       total_value = sum(ExecValue_EUR),
       share_of_investment = ExecValue_EUR / total_value,
       equal_share = 1 / num_stocks,
       NBD = abs(share_of_investment - equal_share) < 0.02) %>%
ungroup() %>%
group_by(User) %>%
summarise(NBD_Proportion = mean(NBD, na.rm = TRUE)) %>%
left_join(sample_investors_clean, by = "User")

```

- `filter(TransType == "B")` : Sélectionne uniquement les transactions d'achat.
- `group_by(User, Jour)` : Regroupe les données par utilisateur et par jour d'achat.
- `mutate(...)` : Calcule plusieurs variables clés pour l'analyse :
 - `num_stocks` : Nombre d'actions distinctes achetées.
 - `total_value` : Valeur totale des achats.
 - `share_of_investment` : Part de l'investissement total pour chaque action.
 - `equal_share` : Part égale théorique (1/N).
 - `NBD` : Indicateur binaire indiquant si l'investissement respecte la stratégie NBD.
- `summarise(NBD_Proportion = mean(NBD, na.rm = TRUE))` : Calcule la proportion de jours où la NBD est utilisée pour chaque investisseur.

Modèle 1 : Utilisation de la NBD au Moins une Fois

J'utilise un modèle de régression logistique pour tester l'hypothèse que certaines variables influencent l'utilisation de la NBD au moins une fois.

Modèle 1 : Utilisation de la NBD au moins une fois

```

merged_data <- merged_data %>%
mutate(NBD_AtLeastOnce = ifelse(NBD_Proportion > 0, 1, 0))

```

```
logit_model1 <- glm(NBD_AtLeastOnce ~ Age_Group + Sex_Male + SP_Q3 + SP_Q7,
  family = binomial, data = merged_data)

# Afficher les résultats du Modèle 1 avec les Odds Ratios

tidy_results1 <- tidy(logit_model1, conf.int = TRUE, conf.level = 0.95) %>%
  mutate(Odds_Ratio = exp(estimate))

cat("Résultats du Modèle 1: NBD utilisé au moins une fois\n")

print(tidy_results1 %>% select(term, estimate, std.error, statistic, p.value, conf.low, conf.high,
  Odds_Ratio), n=Inf)
```

- `mutate(NBD_AtLeastOnce = ifelse(NBD_Proportion > 0, 1, 0))` : Crée une variable binaire indiquant si la NBD a été utilisée au moins une fois.
- `glm(NBD_AtLeastOnce ~ ...)` : Modèle de régression logistique pour tester l'impact des variables explicatives sur l'utilisation de la NBD.
- `tidy(...)` : Extrait et formate les résultats du modèle, calculant les odds ratios.

Modèle 2 : NBD Utilisé 50% ou Plus

Un deuxième modèle de régression logistique est utilisé pour analyser l'utilisation de la NBD dans 50% ou plus des transactions.

```
# Modèle 2: NBD utilisé 50% ou plus

merged_data <- merged_data %>%
  mutate(NBD_Majority = ifelse(NBD_Proportion >= 0.5, 1, 0))

logit_model2 <- glm(NBD_Majority ~ Age_Group + Sex_Male + SP_Q3 + SP_Q7,
  family = binomial, data = merged_data)

# Afficher les résultats du Modèle 2 avec les Odds Ratios
```

```
tidy_results2 <- tidy(logit_model2, conf.int = TRUE, conf.level = 0.95) %>%
```

```
mutate(Odds_Ratio = exp(estimate))
```

```
cat("Résultats du Modèle 2: NBD utilisé 50% ou plus\n")
```

```
print(tidy_results2 %>% select(term, estimate, std.error, statistic, p.value, conf.low, conf.high,
Odds_Ratio), n=Inf)
```

- `mutate(NBD_Majority = ifelse(NBD_Proportion >= 0.5, 1, 0))` : Indique si la NBD a été utilisée dans la majorité des transactions.
- `glm(NBD_Majority ~ ...)` : Modèle de régression logistique pour analyser les facteurs influençant une utilisation majoritaire de la NBD.

Filtrage des Variables Significatives

Je filtre les résultats de la régression pour extraire les variables significatives.

```
# Filtrer les résultats pour obtenir les variables significatives pour chaque modèle
```

```
significant_results1 <- tidy_results1 %>%
```

```
filter(p.value < 0.05)
```

```
significant_results2 <- tidy_results2 %>%
```

```
filter(p.value < 0.05)
```

- `filter(p.value < 0.05)` : Sélectionne uniquement les variables significatives, avec une p-value inférieure à 0.05.

Graphique des Coefficients des Variables Significatives

Je crée des graphiques pour visualiser les coefficients des variables significatives.

```
# Créer un graphique des coefficients avec les variables significatives pour le Modèle 1
```

```
ggplot(significant_results1, aes(x = reorder(term, estimate), y = estimate, ymin = conf.low,
ymax = conf.high)) +
```

```
geom_pointrange() +
```

```
coord_flip() +
```

```
labs(title = "Variables Significatives Impactant l'Utilisation de la NBD (Modèle 1)",
      x = "Variables",
      y = "Coefficient de Régression Logistique") +
theme_minimal()
```

Créer un graphique des coefficients avec les variables significatives pour le Modèle 2

```
ggplot(significant_results2, aes(x = reorder(term, estimate), y = estimate, ymin = conf.low,
                                ymax = conf.high)) +
  geom_pointrange() +
  coord_flip() +
  labs(title = "Variables Significatives Impactant l'Utilisation de la NBD (Modèle 2)",
        x = "Variables",
        y = "Coefficient de Régression Logistique") +
theme_minimal()
```

- `geom_pointrange()` : Crée des graphiques avec des intervalles de confiance pour chaque coefficient.
- `coord_flip()` : Inverse les axes pour une meilleure lisibilité.

Graphiques pour les Variables Explicatives

Des graphiques en barres empilées sont créés pour montrer la proportion de NBD en fonction de différentes variables explicatives.

Créer des graphiques pour les variables explicatives

```
variables <- list("SP_Q7" = "Revenu Net Annuel Total",
                  "SP_Q3" = "Niveau de Connaissance des Marchés Financiers",
                  "Sex_Male" = "Sexe (1=Homme, 0=Femme)",
```

```
"Age_Group" = "Groupe d'Âge")
```

```
create_stacked_bar_plot <- function(data, variable, var_name, x_label, binary_var) {  
  data %>%  
    filter(!is.na(!sym(variable))) %>%  
    group_by(!sym(variable)) %>%  
    summarise(  
      nbd_count = sum(!sym(binary_var), na.rm = TRUE),  
      total_count = n()  
    ) %>%  
    mutate(percentage_nbd = nbd_count / total_count * 100) %>%  
    ggplot(aes(x = factor(!sym(variable)), y = percentage_nbd, fill = "nbd")) +  
    geom_bar(stat = "identity") +  
    labs(title = paste("Proportion de NBD par", var_name),  
         x = x_label,  
         y = "Proportion (%)",  
         fill = "Type de Diversification") +  
    theme_minimal() +  
    theme(legend.position = "top") +  
    scale_y_continuous(limits = c(0, 100), breaks = c(0, 25, 50, 75, 100))  
}
```

```
# Modèles pour chaque type
```

```
plots1 <- lapply(names(variables), function(var) {
```

```

    create_stacked_bar_plot(merged_data, var, variables[[var]], variables[[var]],
"NBD_AtLeastOnce")

})

```

```

plots2 <- lapply(names(variables), function(var) {

    create_stacked_bar_plot(merged_data, var, variables[[var]], variables[[var]],
"NBD_Majority")

})

```

Afficher les graphiques pour le Modèle 1

```
cat("\nGraphiques pour le Modèle 1\n")
```

```

for (plot in plots1) {

    print(plot)

}

```

Afficher les graphiques pour le Modèle 2

```
cat("\nGraphiques pour le Modèle 2\n")
```

```

for (plot in plots2) {

    print(plot)

}

```

- `create_stacked_bar_plot` : Fonction pour créer des graphiques en barres empilées pour chaque variable.
- `geom_bar(stat = "identity")` : Crée un graphique en barres pour représenter les proportions de NBD.

Création de Tableaux Résumant les Données

Des tableaux sont créés pour montrer le nombre de personnes engagées en NBD en fonction des variables explicatives.

Créer des tableaux pour montrer le nombre de personnes engagées en NBD en fonction des variables pour chaque modèle

```
create_summary_table <- function(data, variable, var_name, binary_var) {
```

```
  data %>%
```

```
    filter(!is.na(!sym(variable))) %>%
```

```
    group_by(!sym(variable)) %>%
```

```
    summarise(
```

```
      nbd_count = sum(!sym(binary_var), na.rm = TRUE),
```

```
      total_count = n()
```

```
    ) %>%
```

```
    mutate(percentage_nbd = nbd_count / total_count * 100) %>%
```

```
    arrange(desc(percentage_nbd))
```

```
}
```

```
summary_tables1 <- lapply(names(variables), function(var) {
```

```
  create_summary_table(merged_data, var, variables[[var]], "NBD_AtLeastOnce")
```

```
})
```

```
summary_tables2 <- lapply(names(variables), function(var) {
```

```
  create_summary_table(merged_data, var, variables[[var]], "NBD_Majority")
```

```
})
```

```
# Afficher les tableaux pour le Modèle 1
```

```
cat("\nTableaux pour le Modèle 1\n")
```

```
for (i in 1:length(summary_tables1)) {  
  cat("\nTableau pour", names(variables)[i], ":\n")  
  print(summary_tables1[[i]])  
}
```

```
# Afficher les tableaux pour le Modèle 2
```

```
cat("\nTableaux pour le Modèle 2\n")
```

```
for (i in 1:length(summary_tables2)) {  
  cat("\nTableau pour", names(variables)[i], ":\n")  
  print(summary_tables2[[i]])  
}
```

- `create_summary_table` : Fonction pour créer des tableaux résumant la proportion de NBD en fonction de chaque variable explicative.
- `arrange(desc(percentage_nbd))` : Trie les résultats par ordre décroissant de la proportion de NBD.

