



HAL
open science

Liquidity Risk in the Universe of Open-End Funds

Ran Sun

► **To cite this version:**

Ran Sun. Liquidity Risk in the Universe of Open-End Funds. Business administration. Université Paris sciences et lettres, 2018. English. NNT : 2018PSLED017 . tel-01891210

HAL Id: tel-01891210

<https://theses.hal.science/tel-01891210>

Submitted on 9 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'Université Paris-Dauphine

Liquidity Risk in the Universe of Open-End Funds

Ecole doctorale n°543

Spécialité Sciences de gestion

COMPOSITION DU JURY :

Professeur. Gaëlle LE FOL
Université Paris-Dauphine
Directrice de thèse

Soutenue par **Ran SUN**
le **29 août 2018**

Dirigée par **Professeur Gaëlle LE FOL**

Professeur. Christelle LECOURT
Université Aix-Marseille
Rapporteure

Professeur. Carole GRESSE
Université Paris-Dauphine
Membre du jury

Professeur. Christophe PERIGNON
HEC Paris
Membre du jury



© Copyright by L'Université Paris-Dauphine

L'Université Paris-Dauphine n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses; ces opinions doivent être considérées comme propres à leur auteur.

Remerciements

L'élaboration de cette thèse n'aurait pas été possible sans la collaboration et le soutien de nombreuses personnes. A travers ces quelques mots, je tiens ainsi à leur adresser ma profonde gratitude.

Mes remerciements vont d'abord à ma directrice de thèse, Professeur Gaëlle LE FOL. Je tiens à lui exprimer toute ma gratitude pour ses conseils, ses critiques, ses encouragements ainsi que toutes les heures qu'elle m'a consacrées. Je la remercie pour la confiance qu'elle m'a témoignée, en me donnant l'opportunité de réaliser cette thèse sous sa bienveillante supervision.

Je tiens également à exprimer ma gratitude au Professeur Serge DAROLLES qui est le co-encadrant de mes travaux de recherche. Son écoute, ses conseils précieux et son enthousiasme contagieux m'a accompagné tout long de cette thèse. Je garderai de beaux souvenirs de mes deux encadrants, non seulement pour la qualité de leurs encadrements scientifiques, mais aussi pour leurs qualités humaines. Je ne les remercierais jamais assez.

Je remercie professeur Christelle LECOURT de l'Université Aix-Marseille et professeur Patrick ROGER de l'Université Strasbourg d'avoir accepté d'être les rapporteurs de cette thèse. Je remercie également professeur Carole GRESSE de l'Université Paris-Dauphine, professeur Christophe PERIGNON de l'HEC Paris, d'avoir accepté de faire partie des membres de jury de cette thèse.

Je tiens ensuite à remercier mon co-auteur (chapitre 2) Yang. En tant que chercheur plus senior, il a partagé avec moi généreusement ses connaissances et ses expériences. Celle-ci m'ont beaucoup aidé dans le cadre de ma recherche.

Je tiens également à remercier tous les partenaires industriels du projet MGPF (Modélisation et Gestion du Passif du Fonds): Groupama AM, OFI AM, CMCIC AM, Scaled Risk, A2 Consulting, Scénario Consulting et BPSS. La collaboration avec eux m'a appris l'aspect pratique de ce sujet de recherche et cela a beaucoup nourrit ma thèse. Cela a surtout été un grand plaisir de travailler avec ces personnes chaleureuses. Je remercie Yann DE SAINT-MELEUC, Marie-Claude MUGNIER, Patrick MAURAUX, Monique

BRUNO, Bruno COQUILLE, Pascal MARNAY, Baptiste DIDOIS, Willam MARGOLINE, Jean-Robert HERVY, Thierry DUCAMP, Bertrand THILLAY-DOLEDEC, Anas LAAMIRI et tous les autres participants du projet MGPF.

J'adresse également mes remerciements à l'ensemble de membres du laboratoire DRM-Finance et tous les enseignants du Master 104. Je tiens par ailleurs à exprimer ma gratitude au Professeur Carole GRESSE et Juan RAPOSO, pour leurs précieux conseils et remarques lors de ma pré-soutenance de thèse, qui m'ont permis d'améliorer considérablement ce manuscrit. Je dois un grand merci à Sébastien pour toutes son aide sur ce volet informatique. Je remercie également Madame Françoise CARBON pour son aide relative à l'ensemble des considérations administratives.

L'élaboration de cette thèse aurait été beaucoup plus difficile sans l'accompagnement quotidien de tous les collègues-doctorants de la salle P402. Je remercie Julien, Tingwei, Adrien, Arthur, Louis, Mario, Yasmine, Camille, Thomas D, Anmar, Béatrice, Thomas M ainsi que tous les autres doctorants de DRM-Finance.

Je voudrais exprimer un grand remerciement à Madame Marion Boisseau-Sierra, pour son amitié, son aide, ses encouragements et surtout, son parrainage.

Je ne pourrais pas oublier de remercier tous les enseignants-chercheurs de FFBC de l'Université de Lille 2 qui m'ont grandement aidé à m'installer quand je suis arrivé en France la première fois. Surtout, je voudrais exprimer ma gratitude au Professeur Michel LEVASSEUR et au Professeur Joël LESOIN qui m'ont orienté en Finance au tout début de mes études en France.

Je remercie également tous mes amis, dont le soutien et les encouragements ont été une grande source de force et de motivation durant cette thèse. Je remercie Qun et Yi, doctorants en mathématique-physique, qui ont élargi ma vision de la recherche. Je remercie Zishang, Zeyu, Yu, Xiang, Yuchen et Pablo qui sont comme les frères de sang pour moi. Je remercie Xiao, qui m'a beaucoup aidé quand j'ai suivi les cours à l'ESSEC. Je remercie Huanhuan, Siqu, Yan, Ying, Edmond, Emilien, Vicky, Ronan, Yichen, Luc, Lucie, Maxime, Mamy, Hanxiao, Yubing, Wenxu, Yann, Hugo, Paul QUINSON, monsieur et madame BARBARIN, ainsi tous les autres amis qui m'ont accompagné pendant mes études en France.

Enfin, je remercie ma famille pour son amour et son soutien pendant ces longues années d'études. Aucun langage ne saurait exprimer tout mon amour à l'égard de mes parents. Je ne pourrais que rassembler toutes mes émotions dans un unique mais sincère mot: Merci.

Résumé

Cette thèse étudie le comportement des investisseurs des fonds mutuels ouverts et ses implications au risque de liquidité. Ces travaux de recherche ont pour objectif d'aider les gérants de fonds à éviter le scénario de "fund run" où ils perdent leurs clients de manière soudaine. La première étape de cette étude est de collecter une nouvelle base de données qui enregistre les "micro-transactions" des investisseurs. Cela nous permet d'analyser leurs comportements au niveau individuel et d'effectuer trois articles de recherche autour de ce sujet. Dans le premier article, nous développons un modèle de comptage auto-excitant qui capture des faits stylisés des séries des flux du fonds. De là, nous montrons un risque lié au passif du fonds qui est différent de celui lié à l'actif déjà documenté par la littérature précédente. Nous identifions également une contagion des chocs de liquidité entre les différents clients dans un même fonds. Dans le chapitre suivant, nous étudions les horizons d'investissement des clients individuels. Ces horizons sont fortement liés aux caractéristiques des investisseurs et aux conditions économiques. Nous montrons également que les gérants de fonds subissent un risque de sortie pré-maturée relatif au raccourcissement des horizons d'investissement de ses clients. Nous observons ensuite une hétérogénéité entre les investisseurs: ceux de long-terme comportent différemment que ceux de court-terme. Enfin, dans le dernier chapitre, nous nous intéressons aux activités de rééquilibrage. Nous trouvons que de nombreux investisseurs détiennent un portefeuille contenant plusieurs fonds et le rééquilibrent afin de garder la même allocation d'actifs.

Mots clés: liquidité de financement, fonds ouverts, panique financière, micro-transactions

Abstract

This thesis studies the behaviour of investors in open-end mutual funds and its implications to the liquidity risk. We seek to help the fund managers to avoid the "fund run" scenarios where they loss their clients in a sudden way. We begin our research by collecting a unique data set which records the micro-transactions of fund investors. It allows us to monitor investors' behaviour at the individual level and to accomplish three research articles around this topic. In the first article, we develop a self-exciting counting process to model the stylized facts of fund flows. Therefrom, we highlight a novel risk linked to the fund liability which is different than the asset-related risk documented by the previous literature. We also identify a liquidity contagion among different investors in a same fund. In the next chapter, we study the dispersion in the investing horizons of individual fund clients. These horizons are strongly determined by investors' characteristics and economic conditions. We show that the fund managers suffer a pre-mature redemption risk, i.e. clients shorten their investing horizons and redeem pre-maturely. Especially, we observe a heterogeneity among investors: long-term ones bring a higher pre-mature redemption risk. In the last chapter, we are interested in the rebalance behaviour. We find that numerous investors hold a multi-funds portfolio and rebalance it to keep the target asset allocation.

Key words: funding liquidity, open-end funds, fund run, micro-transactions

À mes parents.

Table des matières

Remerciements	iii
Résumé	iv
Abstract	v
List of Figures	x
List of Tables	xi
Introduction Générale	1
1 Background on Mutual Fund Liability Risk	21
1.0.1 Introduction	22
1.1 Literature review	22
1.2 Transaction-based fund flow database	28
1.2.1 Data-base collection	28
1.2.2 Comparison to previous Fund flow datasets	30
1.2.3 Non-normality of fund flow arrivals	32
1.2.4 Heterogeneity among Fund Investors	35
1.2.5 Conclusion	37
2 A Self-Exciting Model for Mutual Fund Flows:	
Investor Behaviour and Liability Risk	39
2.1 The Data	44
2.1.1 Analysing investors' behaviour	44
2.1.2 Merging fund managers' database	45
2.1.3 Final database characteristics	48
2.2 Modelling investors' behaviour	54

2.2.1	Subscription/redemption counts estimation	54
2.2.2	Capturing over-dispersion	54
2.2.3	Capturing autocorrelation	55
2.2.4	General bivariate specification	56
2.3	Empirical results	58
2.3.1	Are all liquid funds equally risky?	58
2.3.2	Are money market fund predictable?	63
2.3.3	Consequences of clients heterogeneity?	65
2.4	Controlling for other risk factors	73
2.5	Conclusion and Discussion	76
3	Survival Analysis Of Mutual Fund Investors:	
	Evidence From Micro-Transaction Data	77
3.1	Data	83
3.1.1	Summary statistics	83
3.1.2	Preliminary survival analysis	88
3.2	Empirical method	91
3.2.1	Cox proportional hazard rate model	91
3.2.2	Presentation of covariates	93
3.3	Hypotheses on investor's behaviour	96
3.4	Why investors redeem their fund shares?	99
3.4.1	Sector effects on surviving times	103
3.4.2	Time-varying rationality of fund clients	105
3.4.3	Other control variables	107
3.5	Which investors bring the higher "pre-mature" redemption risk	109
3.6	Robustness checks	113
4	Rebalance Activity Of Individual Mutual Fund Investors	119
4.1	Data description & summary statistics	123
4.1.1	Data-base description	123
4.1.2	Summary statistics	124
4.2	Empirical approach	129
4.2.1	Individual trade model	129
4.2.2	Different rebalance methods	132
4.3	First evidence on rebalance activity	133

4.4	How investors switch among funds	139
4.5	Discussion and conclusion	145
	Conclusion générale	147
	Annexe générale	150
	Bibliographie générale	166

Liste des figures

1.1	Times series statistics of fund flows	33
1.2	Auto-correlation of fund flows	34
1.3	Simulation of the "Toy" Poisson model	34
2.1	Inaccurate fits of traditional iid model	51
2.2	Fund flow times series	52
2.3	Distribution of fund flow numbers	52
2.4	Distribution of Model's Parameters: Equity and Fixed Income Funds (an extended sample)	62
2.5	The Liability Risk	65
2.6	Flow Contagion between Investors	68
4.1	Distribution of trade number of individual fund investors	125
4.2	Trade choice rule of ordered response model	131
4.3	Les souscriptions cumulées dans les fonds "fixed income" (2007-2016)	150
4.4	Évolution de l'encours du fonds LTCM (1994.3- 1998.9)	151
4.5	Différentes classes d'actifs gérées par des fonds mutuels ouverts	152

Liste des tableaux

1.1	Raw Data Example	29
1.2	Full fund sample description	30
1.3	Presentation of activity sectors	36
1.4	Weight of each investor sector in the full sample	37
2.1	The fund sample (12 selected funds)	49
2.2	Summary statistics for fund flows	50
2.3	Summary statistics on a demonstrative example:	53
2.4	Count model estimation for 12 selected funds	59
2.5	Models' estimation (for the selected money market fund)	64
2.6	Improvements in model's quality (for the selected money market fund)	64
2.7	Disaggregated sector model	70
2.8	Model's improvements in disaggregation	72
2.9	Factor Flow Model	75
3.1	Summary statistics for the fund sample (one company)	85
3.2	Summary statistics for micro-transactions	87
3.3	Preliminary analysis of surviving times	90
3.4	Covariates' presentation (for the survival analysis)	94
3.5	Cox proportional hazard rate model: full sample estimation	100
3.6	Cox proportional hazard rate model: sub-sample tests comparison	111
3.7	Survival analysis for asset class sub-samples	114
3.8	Long- and short-term investors' comparison with alternative criteria	116
4.1	Investor holding situation	127
4.2	Summary statistics for individual trade intensities	128
4.3	Basic Trade Model (for rebalance analysis)	135
4.4	Fund Switch Model (for rebalance analysis)	140

4.5 Fund Switch Model in quarterly sub-samples (for rebalance analysis) 144

Introduction Générale

Cette thèse s'intéresse à la gestion de la liquidité au passif des fonds ouverts, à partir de la modélisation du comportement des investisseurs. Avant de présenter les aspects techniques à travers des articles de recherche, il nous semble nécessaire dans un premier temps d'introduire certaines notions préliminaires. Dans cette introduction, nous allons d'abord expliquer comment un fonds ouvert est exposé au risque de "run". Dans le secteur bancaire, la littérature utilise le terme "run", panique bancaire en français, pour représenter le scénario où les clients d'une banque craignent qu'elle devienne insolvable et retirent leurs dépôts de manière massive et brutale (voir, e.g., [Zhiguo and Manela 2016](#)). Nous présentons ensuite quelques scénarios de "runs" les plus connus comme celui du fonds "Long-Term Capital Management". Puis, nous comparons les structures opérationnelles des différents types des fonds comme les fonds ouverts, les fonds fermés, les fonds d'arbitrage et les fonds cotés en bourse. Cela nous permet d'identifier les caractéristiques des fonds susceptibles de favoriser ce risque. Les nouvelles régulations européennes, mises en place au 21 janvier 2019, demandent à chaque gestionnaire de fonds de mettre en place des démarches permettant de gérer ce risque, qui passent par une meilleure compréhension du comportement des investisseurs. Nous revenons sur deux modèles théoriques qui expliquent ces comportements. Puis, afin de fournir les outils opérationnels, nous présenterons les articles de recherche qui constituent le corps de cette thèse.

Contexte économique

Les fonds mutuels sont des outils d'investissement très populaires parmi les investisseurs. Selon le rapport de l'Investment Company Institute (**ICI**) sur le marché de gestion d'actifs, les fonds mutuels gèrent environ 40,4 trilliards de dollars d'actifs financiers dans le monde entier (voir, [ICI 2017](#)). Pour les marchés les plus importants, les encours totaux des fonds mutuels sont de : 18,9 trilliards de dollars aux États-Unis, 1,5 trilliard de dollars au Royaume-Unis, 1,9 trilliard de dollars en France, 1,9 trilliard de dollars en Allemagne, 1,3

trilliard de dollars en Chine et 1,1 trilliard de dollars au Brésil. Une raison expliquant cette popularité est que ces fonds adoptent la structure ouverte, i.e. les investisseurs peuvent souscrire et racheter librement leurs parts du fonds, leur permettent de fournir la liquidité à leurs clients. Les clients qui ont des contraintes de liquidité peuvent ainsi s'exposer indirectement à des actifs moins liquides via un investissement dans un fonds mutuel liquide. La tendance de la dernière décennie confirme ce point. De janvier 2007 à décembre 2016, les fonds obligataires¹ aux États-Unis ont connu une croissance remarquable. Au début de cette décennie, les souscriptions cumulées dans cette catégorie de fonds ne représentaient que quelques milliards. Ce montant augmente ensuite rapidement à 500 milliards à la mi-2009, dépasse 1000 milliards en 2010 par atteindre 2000 milliards à la fin 2016 (Voir Annexe A). Comparé aux actions et aux actifs monétaires, les titres obligataires sont des actifs relativement illiquides. Ainsi, nous pouvons conjecturer qu'après la crise de 2008, de plus en plus de clients choisissent de souscrire des fonds mutuels afin d'investir dans les actifs moins liquides.

La structure ouverte, ou "open-end" en anglais, est un avantage pour les investisseurs. Elle leur laisse la possibilité de sortir quand ils ont besoin. Cependant, elle engendre le risque que les clients sortent de manière massive et que le fonds perd alors brutalement ses encours. Inspiré par le "bank run", la littérature nomme ce scénario analogue le "fund run" dans le cadre des fonds (voir, [Schmidt et al. 2016](#)). Un des épisodes les plus connus de "fund run" est la faillite du fonds "Long-Term Capital Management L.P" (LTCM).

LTCM² est un fonds d'arbitrage créé en 1991 par John Meriwether, ancien directeur des transactions "fixed income" de la banque "Salomon Brothers". Il réunit de nombreux experts du marché "fixed income", tels que Eric Rosenfeld, Dick Leahy, James McEntee, Robert Shustak. Deux célèbres professeurs de finance, Myron Scholes et Robert Merton, participaient également à ce projet. Le fonds LTCM était spécialisé dans les activités d'arbitrage sur les marchés de taux ("fixed income"), grâce à des stratégies diversifiées telles que: (1) l'arbitrage des obligations souveraines de long-terme, à 30 ans par exemple; (2) l'arbitrage de la prime de liquidité, i.e. acheter les dettes illiquides et vendre les dettes similaires mais plus liquides; (3) les stratégies spéculant sur la volatilité des taux d'intérêt; (4) les transactions de produits dérivés de taux, etc. Au cours de l'année de sa création, LTCM a atteint une rentabilité supérieure à 40%.

Toutefois, dès le début de la crise asiatique en 1997, le fonds LTCM a commencé à

¹Nous utilisons le terme anglais "fixed income" dans la suite de cette introduction.

²L'histoire, les chiffres et les événements rapportés dans ces paragraphes proviennent de [Lowenstein \(2000\)](#).

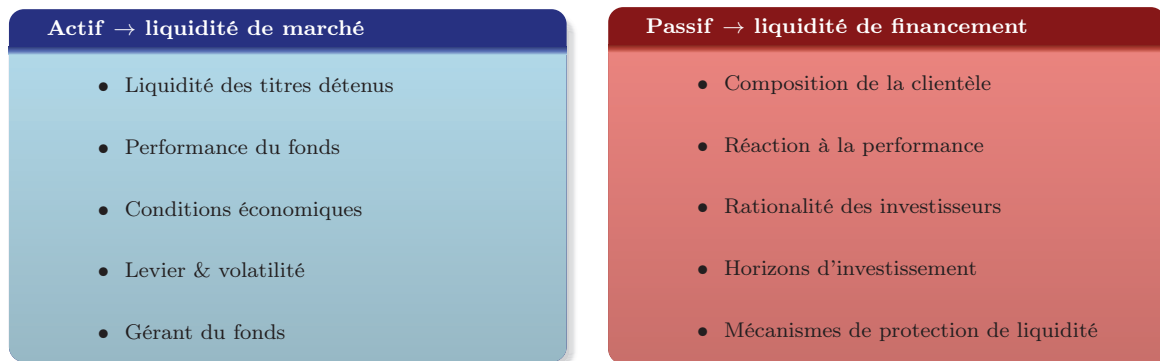
accumuler des pertes. En mai et juin 1998, le fonds a perdu 6,42% puis 10,14%, ce qui correspond à une moins-value de 461 millions dollars pour ses investisseurs. En août de cette même année, le défaut de la Russie génère une panique sur le marché de la dette. Les investisseurs vendent massivement les dettes européennes et se tournent vers la dette américaine. LTCM a perdu 1,85 milliards de dollars suite à ce choc macroéconomique. Présentant un ratio de levier important, le fonds s'est retrouvé sous la pression d'appels de marge. De plus, l'échec de l'arbitrage de Royal Dutch Shell a généré une perte de 286 millions dollars. Ces mauvaises nouvelles ont poussé les clients à procéder à des rachats massifs. Pendant les trois premières semaines de septembre 2008, les actifs sous gestion sont passés de 2.3 milliards à 400 millions dollars et le levier a atteint à 250:1. Nous présentons l'évolution de l'encours de LTCM dans l'annexe B.

La difficulté de LTCM a suscité la crainte des régulateurs, qui ont eu peur que la faillite de ce fonds impacte les marchés et cause une crise systémique. Le 23 septembre 1998, plusieurs grandes institutions financières ont injecté 3,77 milliards dollars à LTCM en échange de 90% de capitaux propres. En plus, la Fed a également renfloué LTCM à hauteur de 3,62 milliards dollars. Grâce à ces soutiens, LTCM a pu continuer ses activités encore un an et a pu généré une performance de 10%. Cela a permis à la plupart des investisseurs de récupérer au moins leurs investissements initiaux.

Nous pouvons résumer les principales raisons de cet échec. Premièrement, LTCM a sur-investi dans des dettes de long-terme. Ces actifs sont peu échangés sur le marché, donc ils sont relativement illiquides par rapport aux autres actifs comme les actions ou les titres monétaires. De plus, le fonds a souffert de mauvaises performances pendant plusieurs mois et les investisseurs ont commencé à douter de la capacité des gérants à générer une performance satisfaisante. Troisièmement, les chocs macroéconomiques, comme la crise asiatique et russe, ont été défavorables à LTCM. Puis, ce fonds s'est engagé dans des stratégies nécessitant un effet de levier important et la détention de nombreux titres volatiles. Enfin, confortés sans doute par leur réputation, les gérants ont souffert d'une confiance trop importante dans leurs capacités.

Ces explications sont-elles suffisantes pour comprendre les raisons de ce "fund run"? Brunnermeier et Pedersen considèrent à cette question dans leur recherche et proposent d'analyser le niveau de liquidité à deux côtés du bilan d'une banque, ou d'autres intermédiaires financiers (voir, [Brunnermeier and Pedersen 2009](#)). Le premier côté, mesuré par la liquidité des titres détenus par le fonds, est lié à la liquidité du marché. Les raisons mentionnées au paragraphe précédent se trouvent de ce côté du bilan. Autrement dit, c'est la liquidité du côté actif du fonds. Le deuxième élément est la liquidité de financement

que l'on peut également appeler liquidité du côté passif du fonds. Dans l'histoire racontée par [Lowenstein \(2000\)](#), on trouve rarement d'information sur ce dernier et le côté passif de la liquidité est ignoré. Cependant, nous pouvons imaginer comment la structure du passif peut influencer le niveau de risque de liquidité du fonds. En premier lieu, la composition des clients est-elle diversifiée? Si le fonds LTCM n'est détenu que par quelques investisseurs, une sortie de l'un entre eux peut impacter le fonds entier. Deuxièmement, les clients sont-ils sensibles à la performance du fonds ou aux conditions macroéconomiques? Ensuite, les décisions de sorties de fonds sont-elles rationnelles? En moyenne, combien de temps les clients restent-ils dans le fonds? Leur investissement doit-il être considéré comme court ou long terme? Enfin, existe-t-il des mécanismes de protection de liquidité, comme le seuil de rachats ou la période d'inaccessibilité, qui empêchent les sorties massives³? L'ensemble de ces éléments ont des impacts cruciaux sur les risques de "bank run". Le mini-bilan ci-dessous résume les facteurs qui influencent les deux types de liquidité:



Selon [Brunnermeier and Pedersen \(2009\)](#), ces deux côtés de liquidité sont aussi importants l'un que l'autre. Cependant, comme mentionné précédemment, le côté passif est plus ou moins oublié. La première motivation de nos travaux est d'étudier la liquidité du côté passif afin de combler ce vide. Par ailleurs, les "fund runs" sont des événements qui sont loin d'être isolés et les conséquences sont possiblement systémiques, leur étude semble aujourd'hui de première importance. En 2006, l'Allemagne a subi un scénario de "fund run" au niveau du marché. Le secteur des fonds immobiliers a souffert d'une fuite importante des investisseurs. Par conséquent, ces fonds ont perdu 40% de leurs encours pendant cette période. Ainsi, les gérants ont été obligés de liquider une partie de leurs actifs. Cela a causé ultérieurement une chute considérable des prix immobiliers en Allemagne. Le risque de "fund run" existe non seulement parmi les fonds illiquides, mais aussi dans les fonds qui investissent dans des actifs supposés liquides. En 2008, les fonds monétaires aux États-Unis ont subi 200 milliards de rachats cumulés pendant une seule semaine, du 15 au 22 septembre (voir, [Schmidt et al. \(2016\)](#)). Surtout, [Chernenko and Sunderam \(2014\)](#) indiquent que

³Voir les mécanismes qui protègent la liquidité des fonds d'arbitrage dans [Blackrock \(2016\)](#)

cette crise des fonds monétaires a eut des impacts sévères sur le marché des prêts aux entreprises. Les entreprises ont en effet subi un problème de financement de court-terme. Ainsi, comprendre le risque de "run" est devenu un besoin fondamental du secteur de la gestion d'actifs.

Les autorités réglementaires ont réagi face à cette préoccupation. Plusieurs réglementations, qui tentent d'atténuer et prévenir les rachats massifs, ont été progressivement mises en place. Un nouveau règlement du parlement européen et du conseil de l'Union Européenne du 17 juillet 2017 propose plusieurs mesures qui visent à mieux connaître le passif d'un fonds OPVCM et de renforcer sa liquidité de financement. Nous résumons et expliquons certains articles de cette nouvelle réglementation à ci-dessous⁴

L'article 21 exige de chaque société de gestion à mettre en place des politiques permettant de se familiariser avec sa base d'investisseurs. Cette exigence possède trois composants. Premièrement, une société de gestion doit connaître ses clients au niveau individuel. Elle a besoin de bien déterminer l'identité de chaque investisseur et d'enregistrer également son profil. Deuxièmement, la société de gestion doit mesurer l'importance de chaque investisseur dans son fonds. Par exemple, un gérant pourrait calculer le ratio d'emprise de chaque investisseur en divisant le montant de sa part par la taille du fonds. Une attention particulière doit être accordée aux grands investisseurs représentant un ratio d'emprise élevé. Troisièmement, la société de gestion doit anticiper les éventuels rachats importants dans le futur.

L'article 28 demande chaque société de gestion d'adopter des processus des simulations des rachats futurs. La méthode de simulation doit prendre en compte les différentes hypothèses économiques. Surtout, le gestionnaire d'un fonds doit avoir la capacité de simuler les conséquences de chocs de liquidité extrêmes.

L'article 37 demande chaque société de gestion de mettre en ?uvre et d'appliquer des procédures afin de prévoir les rachats collectifs. Cette collectivité a deux dimensions: (1) les rachats peuvent être simultanés: de nombreux investisseurs réagissent à un même facteur économique et sortent du fonds simultanément. L'identification des facteurs explicatifs des rachats est essentiel pour comprendre cette situation. (2) Les rachats peuvent être corrélés. Une sortie d'un grand client peut susciter des sorties d'autres clients. Pour cette situation, nous devons étudier les réactions des investisseurs aux rachats des autres. Par ailleurs,

⁴Voir: Journal officiel de l'Union Européenne

cet article exige que les sociétés de gestion prennent en considération les facteurs "clients": le type de client, le nombre de parts détenues par un client dans le fonds, l'historique de souscriptions et rachats d'un client.

Ces articles ne concernent que les OPVCMs monétaires. Pourtant, le risque de "runs" existe dans tous les types de fonds ouverts. Les gérants doivent ainsi réagir contre ce risque provenant du passif. Ce règlement impose aux sociétés de gestion la connaissance de leurs clients et de leurs comportements. Comparé à la directive européenne UCITS IV datée de 2009, qui s'intéresse plutôt au risque de liquidité du côté actif, cette nouvelle réglementation montre l'importance du côté passif. Avant de commencer à étudier ces questions concernant le risque de liquidité, il est nécessaire de comprendre les différentes structures du passif des fonds d'investissement et c'est l'objet de la prochaine section.

Différentes structures des fonds d'investissement

Un gérant de fonds d'investissement doit choisir les conditions de liquidité qui précisent la façon dont les investisseurs peuvent souscrire et racheter leurs parts du fonds. Les structures les plus utilisées sont les fonds mutuels ouverts, les fonds mutuel fermés, les fonds d'arbitrage et les fonds négociés en bourse. On utilise plutôt le nom anglais "Exchange Traded Fund", ETF, pour faire référer le dernier type de fonds mentionné. Chaque structure correspond à des classes d'actifs particulières et les différentes conditions de liquidité du fonds correspondent aux différents niveaux de risques de liquidité.

Fonds mutuels ouverts (open-end mutual funds)

La structure la plus utilisée est celle du fonds mutuel ouvert. Son nom "mutuel" désigne le fait qu'il reçoit des capitaux de ses investisseurs de manière collective. Sa propriété "ouverte" permet aux clients de souscrire et/ou de racheter librement sans aucune contrainte autre que celle de la date de valorisation. Cette ouverture exige donc un système particulier d'évaluation de portefeuille qui doit servir de base aux transactions des clients. Chaque jour boursier⁵, le gérant de fonds demande à un auditeur, souvent extérieur, d'estimer le prix de chaque titre détenu et de sommer tous les titres pour obtenir la valeur du portefeuille global. Ensuite, cette valeur sous gestion, "Asset Under Management" (AUM) en anglais, est divisée par le nombre de parts dans le fonds afin d'obtenir le prix unitaire de chaque part.

⁵On donne un exemple d'un fonds journalier ici. Mais la fréquence d'ouverture peut être également hebdomadaire, mensuel, etc.

On utilise plutôt le terme anglais de la valeur d'actif net, "Net Asset Value" (NAV) , pour indiquer ce prix. Tous les ordres de souscription et de rachat reçus pendant cette journée vont être exécutés au prix NAV⁶. L'auditeur fait généralement ce calcul à la clôture des marchés. Par contre, les ordres sont reçus pendant la journée et donc avant ce calcul. Cela crée donc un décalage temporel car les investisseurs ne sont informés du prix de transaction qu'à la fin de journée, après le passage de leur ordre.

Les gestionnaires d'actifs utilisent cette structure pour leurs investissements dans de nombreuses classes d'actifs diversifiés. Selon Morningstar, un fournisseur de données connu dans l'industrie de la gestion d'actifs, les fonds mutuels ouverts investissent principalement dans les actifs suivants: actifs diversifiés, actifs alternatifs, matières premières, actions, dettes/obligations, actifs divers et actifs monétaires⁷. Nous présentons les poids de chaque catégorie de fonds en France dans l'annexe C.

Bien évidemment, le risque de liquidité dépend fortement du côté actif du fonds. Un fonds monétaire détient des titres beaucoup plus liquides que ceux détenus par un fonds "fixed income". Par ailleurs, les fonds de la même catégorie peuvent aussi s'exposer à différents niveaux de risque de liquidité. Par exemple, les fonds actions qui investissent dans les sociétés de la grande capitalisation, "large-caps" en anglais, sont bien plus liquides que les fonds actions spécialisés en petite capitalisation, les "small-caps". Surtout, les fonds qui méritent notre attention sont ceux qui gèrent des actifs très illiquides et de très long-terme, comme les fonds immobiliers et les fonds obligataires. Ces fonds ont un risque de "run" plus élevé puisque les gérants auraient plus de difficultés pour vendre les actifs illiquides afin de répondre aux rachats de leurs investisseurs. En effet, les fonds investissant dans les actifs illiquides sont ceux qui ont souffert de plus de "runs" (voir, [Goldstein et al. 2017](#) et [Fecht and Wedow 2014](#)).

Cependant, en dehors de la liquidité des actifs détenus, le passif d'un fonds contient aussi des facteurs déterminants du risque de liquidité. Un scénario de "fund run" est directement lié au comportement des clients. Un fonds illiquide, comme un fonds immobilier, peut ne rencontrer aucun problème même en période de crise si ses clients ne réagissent pas aux chocs externes et ainsi, si des sorties massives ne se produisent pas. En revanche, si les investisseurs sortent de manière trop brutale, même un fond monétaire, qui est en général liquide, peut rencontrer des difficultés pour revendre certains de ces actifs. Afin de gérer ce

⁶Pour certains fonds qui investissent en dehors de leur marché domestique, nous ajustons le timing d'évaluation pour éviter un arbitrage (e.g. [Chalmers et al. 2001](#)).

⁷Voir, "The Morningstar Category Classifications" dans http://im.mstar.com/im/newhomepage/Morningstar_Category_Definitions_US_June_2016.pdf. La classification de fonds n'est pas unique. Dans d'autres chapitres de cette thèse, nous utilisons également les méthodes alternatives.

risque, l'approche traditionnelle des gérants de fonds consiste à estimer une sortie moyenne des clients dans le futur proche et construire une allocation de disponibilités correspondant (voir, [Chernenko and Sunderam 2016](#); [Darolles and Roussellet 2018](#)). Cependant, cette estimation est imprécise et souvent erronée car les données manquent, et les méthodes de prévision des rachats des investisseurs sont peu efficaces.

Fonds fermés (closed-end funds)

Au lieu d'ouvrir le passif, la deuxième structure correspond aux fonds qui le ferment. Cette structure est ainsi fermée et appelée "closed-end". Ce type de fonds ne permet pas des souscriptions ou/et des rachats libres. Afin d'accueillir les investisseurs, le gérant définit une période initiale appelée la période de capitalisation. Les clients d'un fonds fermé peuvent y souscrire uniquement durant cette période⁸. Après avoir collecté les capitaux, le fonds ferme et investi dans les actifs cibles. Puis, il distribue les profits selon la décision du gérant pendant la vie du fonds et les capitaux seront remboursés à la fin lorsque le gérant liquide le fonds. Aucun rachat ne peut être exécuté avant la date de liquidation.

Le principal usage d'un fonds fermé est d'investir dans les sociétés non-côtées. Si le fonds intervient directement dans la gestion de ces sociétés, on parle de fonds de capital investissement, ou "private equity" en anglais. On parle de fonds "dette privée" si le fonds investi dans les dettes des sociétés non-côtées. En outre, on utilise aussi cette structure afin de gérer des investissements immobiliers ou des projets d'infrastructure. Comparé aux fonds ouverts, les fonds fermés s'engagent dans moins de classes d'actifs. Les capitaux ainsi que les dettes privées sont des actifs illiquides et ils demandent une détention de long-terme. La structure fermée peut être considérée comme une protection de liquidité pour ce type de fonds.

Les fonds fermés ont une distribution de risque de liquidité différente des fonds ouverts. Puisque les fonds ouverts laissent aux clients la liberté de sortir, le gérant subit le risque de liquidité. Cependant, dans le cas d'un fonds fermé, le gérant détient des actifs de long-terme sans que les investisseurs puissent sortir. Cela implique que le risque de liquidité est supporté à 100% par les investisseurs et c'est pourquoi le risque de passif n'est pas une préoccupation des fonds fermés.

Fonds d'arbitrage (hedge funds)

Les fonds d'arbitrage sont des véhicules d'investissement très utilisés. Ils ont une structure ouverte qui permet aux clients d'entrer et de sortir pendant toute la vie du fonds. Par cet

⁸Il est possible qu'un fonds fermé ouvre de nouveau s'il est trop demandé par les investisseurs.

aspect, les fonds d'arbitrage ressemblent aux fonds mutuels ouverts. Pourtant, des modifications sur la structure ouverte permettent de tenir compte de stratégies uniques des gérants. En effet, un fonds d'arbitrage offre la liberté de souscription/rachat sous certaines contraintes: (1) la période d'incessibilité, "lock-up period" en anglais: un fonds d'arbitrage peut définir une mauvaise "coupure" pendant laquelle les investisseurs ne peuvent pas souscrire ou racheter. Pendant cette période, la liquidité du fonds est ainsi protégée et le fonds devient temporairement fermé. (2) le seuil de rachat, "redemption gate" en anglais: les investisseurs peuvent sortir du fonds librement mais le montant du rachat ne peut pas dépasser à un certain seuil, disant "gate". Ce mécanisme permet aux fonds d'éviter de perdre leurs clients de manière trop brutale. (3) la réserve de disponibilités, "cash reserve": le gérant peut garder une proportion de "cash" dans son portefeuille. Si un investisseur veut racheter sa part, une proportion de cash doit être laissée dans le fonds. Cette provision permet au gérant d'avoir suffisamment de disponibilités afin de rembourser les sorties à venir, de répondre aux appels de marge ou de saisir des opportunités d'investissement/arbitrage. (4) le sous-portefeuille isolé, dit "side pocket": cela est plus une disposition qu'un mécanisme souvent constitué d'actifs très illiquides. Les titres dans ce "side pocket" sont exonérés de l'évaluation du fonds. Les études précédentes montrent que tous ces mécanismes renforcent la liquidité du fonds et améliorent sa performance (voir, [Teo 2011](#); [Aiken et al. 2015](#)).

Pour mieux comprendre l'intérêt de ces disponibilités, nous comparons les fonds d'arbitrage avec les fonds mutuels. Les fonds d'arbitrage ont deux sources additionnelles de risque de liquidité. D'un côté, les fonds d'arbitrage ont un effet de levier important; ils s'endettent beaucoup afin d'atteindre une meilleure performance. D'un autre côté, les fonds d'arbitrage utilisent beaucoup les produits dérivés. Les caractéristiques de ces actifs impactent la performance du fonds. Par exemple, les produits dérivés ont un profit de valeur asymétrique et la chambre de compensation demande un dépôt de garanti et des appels de marge sur ces produits. En conséquence, les fonds d'arbitrage ont besoin d'une protection de liquidité accrue, en comparaison aux fonds mutuels.

Côté passif, fonds d'arbitrage et fonds mutuel se ressemblent. Cependant, les clients qui investissent dans ces deux fonds sont différents. Dans beaucoup de réglementations financières, comme le code AMF en France, seuls les investisseurs "éligibles" peuvent souscrire aux fonds d'arbitrage. Les régulateurs demandent aux investisseurs d'avoir un bon niveau de connaissance du marché et des produits financiers. En plus, les investisseurs doivent être conscient des risques auxquels un fonds d'arbitrage s'expose. En outre, certaines grandes institutions financières, comme certains fonds de pension ou certaines compagnies d'assurance n'ont pas le droit d'investir dans les fonds d'arbitrage du fait de règles internes. Au con-

traire, presque tous les investisseurs peuvent souscrire à des fonds mutuels sans aucune contrainte réglementaire. En analysant les différences entre ces deux structures, on pourrait poser naturellement deux questions: (a) est-il utile pour un fonds mutuel d'introduire des mécanismes de liquidité comme pour les fonds d'arbitrage? (b) Est-ce que la composition de base d'investisseurs impacte le risque de liquidité, et surtout, le risque de "fund runs"?

ETF

Un ETF est un fonds d'investissement dont les parts sont cotées en bourse. Le principal objectif d'un ETF est de suivre un indice boursier. Normalement, un ETF cherche à garder la même composition de l'indice qu'il suit et son objectif est de délivrer la même performance que ce dernier. Il permet aux investisseurs d'avoir une position diversifiée avec des titres très liquides. Par ailleurs, un ETF est souvent divisé en plusieurs parts de petits montants. Cela permet d'attirer plus de clients, surtout les particuliers ("retails") qui n'ont pas de grandes capacités d'investissement contrairement aux clients institutionnels. Les ETFs sont considérés comme les outils de l'investissement passif.

Les ETFs ont un mécanisme de souscription/rachat particulier. Les teneurs de marché, aussi appelés participants autorisés ("Authorized Participants", APs), sont responsables de l'acquisition des titres pour l'ETF. Ils achètent tous les composants de l'indice suivi par l'ETF et ils les livrent à cet ETF. En échange, ils reçoivent des parts de l'ETF de valeur équivalente, qui s'appellent les unités créations. Par cette procédure, un ETF est créé. Désormais, les APs représentent les intermédiaires entre l'ETF et ses investisseurs. Ces derniers achètent ou vendent les parts de l'ETF aux APs. A chaque fin de journée, s'il y a plus d'achats que de ventes, les APs créent de nouvelles parts de l'ETF. Cela est équivalent à des nouvelles souscriptions dans les fonds ouverts. En revanche, si les investisseurs vendent plus que ce qu'ils achètent, les APs suppriment les parts. Cela pourrait être considéré comme les rachats dans les fonds ouverts. Les activités des APs garantissent une convergence de prix dans la mesure où la NAV de l'ETF s'aligne sur le prix des titres sous-jacents.

Il est intéressant de comparer les ETFs aux fonds mutuels. Une différence remarquable est le timing de l'évaluation. Un ETF suit un indice et son prix est publié dès que la bourse est ouverte. Dans une même journée, il y a une continuité des prix boursiers pour un ETF. En revanche, un fonds mutuel a un seul prix par jour qui est publié en fin de journée, comme expliqué au début de cette section. Par conséquent, il est possible d'observer les prix et les volumes d'un ETF avec une fréquence intra-journalière. Au contraire, cela n'est pas possible pour un fonds mutuels qu'à une fréquence journalière.

Une deuxième comparaison est d'opposer un ETF à un fonds mutuel indiciel. Ces deux

fonds suivent tous les deux une stratégie passive et sont considérés comme des produits concurrents. Généralement, les investisseurs préfèrent un ETF à un fonds mutuel indiciel pour les trois raisons principales suivantes: premièrement, un fonds mutuel a des frais de gestion plus élevés qu'un ETF. Ces "fees" couteux s'expliquent par le fait qu'un fonds mutuels a besoin d'une infrastructure lourde pour mener ses activités. La société de gestion doit sauvegarder tous les profils des clients, garder du personnel au service clientèle et construire une équipe "conformité" afin de répondre aux demandes réglementaires. Un ETF ne supporte aucun de ces coûts, ainsi ses frais de gestion sont bien moins élevés. Deuxièmement, un ETF offre des prix inter-journaliers et est ainsi plus liquide qu'un fonds mutuel. Troisièmement, puisque l'ETF investit dans un indice et ses clients échangent leurs parts sur un marché secondaire, les ventes des investisseurs impactent moins la valeur du fonds. Ainsi, il y a moins de risque de "runs" comme dans un fonds mutuel ou dans une banque. au regard de l'ensemble des désavantages de fonds mutuel par rapport à un fonds ETF, on peut se demander pourquoi les investisseurs ont-ils besoin de fonds mutuel indiciel/passif et quels types de service attendent-ils d'une telle société de gestion?

Parmi les quatre types de fonds que l'on vient de présenter, le fonds mutuel est l'objet d'étude idéal pour mieux comprendre le risque de liquidité au passif. Tout d'abord, nous avons vu que les fonds mutuels sont très sujets aux "fund runs". Ensuite, les fonds mutuels couvrent une plus grande diversité d'actifs et d'investisseurs. Enfin, les fonds mutuels n'utilisent aucun mécanisme de protection de liquidité comme les fonds d'arbitrage. Comprendre les souscriptions/rachats des clients et renforcer la liquidité du fonds est devenu un besoin essentiel dans l'industrie de la gestion d'actifs.

Risque de passif des fonds mutuels

Avant d'analyser le risque de "runs" en détail, il est nécessaire de présenter dans un premier temps les différentes notions opérationnelles d'un fonds mutuel. Un fonds est créé par une société de gestion. On l'appelle aussi gestionnaire d'actifs ou asset manager. Cette société peut être indépendante, comme Blackrock, Statestreet ou Vanguard. Mais, il est également possible pour un gestionnaire d'actifs de s'affilier à un conglomérat financier. Les exemples les plus connus des sociétés affiliées en France sont AXA Investor Managers, BNP Investment Partners ou encore Alliance Global Investors. Une particularité de cette affiliation est qu'une partie de capitaux souscrits provient de la maison mère. Par exemple, la moitié des souscriptions d'AXA IM sont faites par le groupe AXA lui-même et l'autre moitié vient des investisseurs externes. Grâce à la continuité de ces ressources internes, les

gestionnaires affiliés sont considérés comme moins vulnérables au risque de liquidité que les gestionnaires indépendants. Pourtant, malgré cette contribution importante en terme de capitaux, la maison mère n'intervient pas dans la gestion des fonds. En général, les gérants ont la liberté dans les décisions de leurs propres investissements.

Un fonds collecte les capitaux de nombreux investisseurs et ces capitaux sont gérés par le gérant du fonds. Dans certains cas, il y a d'autres personnes qui assistent le gérant principal. On appelle ces personnes les co-gérants. Les investisseurs du fonds sont souvent appelés les porteurs. Un gérant peut gérer plusieurs fonds et un porteur peut également investir dans plusieurs fonds. Dans de nombreuses situations, la société de gestion utilise un service intermédiaire afin de vendre ses parts de capital. Ce service intermédiaire s'appelle un distributeur. Il correspond au service de courtage dans la littérature (voir, [Bailey et al. 2011](#)). Une grande société de gestion a souvent des équipes séparées pour la gestion et la vente. Le gérant ne connaît donc pas la plupart de ses porteurs. Les clients sont issues de différents secteurs économiques, et souvent, la littérature les classe en deux grandes catégories, les particuliers et les institutions (voir, [Ben-Rephael et al. 2017](#); [Cao and Petrasek 2014](#); [Edelen et al. 2016](#); [Foster and Warren 2016](#)).

L'AUM global d'un fonds est divisé en plusieurs petites unités et le prix "NAV" correspond à la valeur d'une unité. Les clients paient trois types de frais à la société de gestion. Les premiers sont les frais de gestion, dits "AUM fees". Ses frais égalent à la valeur totale des parts détenues par un client multiplié par un taux de frais. La société va encore définir un seuil de performance. Si le rendement dépasse ce seuil, les investisseurs vont payer une partie de cette sur-performance. Par exemple, en définissant un seuil à 20%, si le gérant réalise un rendement de 25%, les investisseurs paient une partie de ce 5% ($25\% - 20\%$) de sur-performance au gérant, disons un pourcentage égal à 30%. Ces frais supplémentaires incitent les gérants à sur-performer, c'est pourquoi nous les appelons bonus d'encouragement, ou "incentive fees". Le seuil de 20% est le taux de "hurdle" et le taux de 30% de proportion correspond aux intérêts de "carry". Le dernier frais de gestion est le droit de souscription/rachat. Les investisseurs paient un montant à chaque transaction. Aux États-Unis, nous appelons cela le "load" (voir, [Johnson 2007](#)). En France, ces droits sont souvent exonérés, afin d'attirer plus de clients. Ces frais ne sont pas les mêmes pour tous les investisseurs. La société de gestion affecte les clients dans plusieurs catégories d'un fonds, disons les "parts" du fonds. Chaque part correspond à des frais de gestion spécifiques.

Les investisseurs ont la liberté d'entrer et de sortir s'ils le souhaitent. Ainsi, le gérant observe des flux entrants/sortants de son fonds. Les études précédentes appellent les flux entrants les "inflows" et les flux sortants les "outflows" (e.g., [Akbas et al. 2015](#); [Bergstresser](#)

and Poterba 2002; James and Karceski 2006; Kumar et al. 2015; Spiegel and Zhang 2013). Souvent, le gérant n'observe qu'un montant net, qui est l'écart entre les "inflows" et les "outflows". On les appelle flux nets. Chaque souscription/rachat peut être considéré(e) comme un flux individuel. La littérature de "household finance" le nomme "micro-transaction".

Les fonds mutuels ont un côté passif de court-terme puisqu'un client peut sortir au lendemain de sa souscription. Toutefois, ils ont un actif de plus long-terme. Dans une perspective macroéconomique, ces fonds accomplissent une fonction de transfert de liquidité, i.e. ils collectent les ressources de court-terme et financent les actifs de long-terme. Alternativement, nous disons aussi qu'il y a un écart de liquidité dans les fonds mutuels. De ce point de vue, un fonds se rassemble beaucoup à une banque. La liquidité de ses porteurs peut être garantie puisque les investissements de tous les clients sont mélangés et ils atteignent un équilibre dans le contexte du modèle de Diamond-Dybvig: la consommation marginale de liquidité égale l'offre de liquidité (voir, [Diamond and Dybvig 1983](#)). Cet équilibre peut être atteint car un fonds mélange les investisseurs qui ont différents besoins de liquidité. Ainsi, la littérature indique que cette mixité crée une assurance de liquidité (voir, [Johnson 2004](#), Section V, "insurance-by-pooling").

Grâce aux notions vues précédemment, nous définissons le risque de passif d'un fonds mutuels de la manière suivante:

Le risque de passif d'un fonds ouvert correspond au risque de perte de performance due aux comportements de ses clients.

Il faut souligner que les "inflows" ont des impacts aussi néfastes sur la performance que les "outflows". Si le gérant n'a pas suffisamment de "cash" pour répondre aux rachats, il doit vendre des actifs moins liquides à un prix désavantageux. De manière opposée, trop d' "inflows" peut aussi créer des problèmes. La recherche précédente indique que chaque gérant possède une taille de fonds par laquelle il est expert et a une parfaite maîtrise. Si l'AUM devient trop grand, une contrainte s'impose au gérant. Il ne sait plus quoi investir et place une part significative en "cash". Cela a pour effet de diminuer le rendement du fonds⁹.

La clé pour évaluer ce risque de passif est de comprendre le comportement des clients. Le gérant a besoin de réponses aux questions suivantes: pourquoi les investisseurs souscrivent/rachètent aux parts de fonds mutuels? Quels types de fonds préfèrent-ils? Quels sont les facteurs qui

⁹Voir l'article de **CBS news**, du juillet 11 2011: "*The Bigger Active Funds Get, the Worse Their Alpha*". <https://www.cbsnews.com/news/the-bigger-active-funds-get-the-worse-their-alpha/>

impactent le volume de souscription/rachat? En période économique défavorable, quels types de clients sont les plus à même de racheter? Est-ce que la classe d'actifs du fonds et/ou les caractéristiques du client joue un rôle déterminant dans la dynamique des flux? Dans un premier temps, nous nous basons sur des modèles théoriques pour trouver des réponses à ces questions.

Modèles théoriques sur le comportement des clients

La première étape dans la compréhension du comportement des investisseurs revient à répondre à deux questions essentielles: (1) comment les investisseurs réagissent au rendement du fonds et (2) quelle est leur attitude face au risque? Deux études théoriques précédentes concernent respectivement ces deux aspects. Le premier modèle, celui de [Berk and Green \(2004\)](#), décrit dans un contexte rationnel, comment les investisseurs choisissent les fonds à souscrire/racheter. Ce modèle explique certaines anomalies observées dans les comportements des clients. Par exemple, les investisseurs choisissent souvent les fonds qui ont le meilleur rendement passé malgré que cette sur-performance ne persiste pas dans le futur. Ils montrent que cette manque de persistance n'implique pas forcément la mauvaise capacité du gérant. Le modèle prévoit une relation entre la performance et le rendement qui est cohérent avec des évidences empiriques (e.g., [Chevalier and Ellison 1999](#); [Sirri and Tufano 1998](#)). Le deuxième modèle, celui de [Chen et al. \(2010\)](#) indique que la "run incentive" des investisseurs rend les fonds plus fragiles. Leur modèle prédit que les fonds qui détiennent plus de titres illiquides ont une plus forte relation entre les sorties et la performance. De plus, cette propriété est plus prononcée pour les fonds dont la composition des clients est plus dispersée. Dans cette section, nous présentons ces deux modèles de manière concise et discutons leurs implications pour la gestion de risque de passif. Les démonstrations des modèles sont présentées dans les annexes D et E.

Modèle Berk-Green

Le modèle Berk-Green s'intéresse à la façon par lesquelles les investisseurs évaluent la qualité du gérant de fonds. Le comportement d'un investisseur, i.e. ses souscriptions et ses rachats, est une réaction à cette évaluation. Si un investisseur pense qu'un gérant maîtrise son sujet, il va souscrire davantage à son fonds, sinon il en sortira. Un fonds mutuel est ainsi un actif financier dont l'attractivité dépend de qualités humaines. Cela le différencie de la majeure partie des actifs financiers pour lesquels les investisseurs n'évaluent que la qualité des actifs

en eux-mêmes. Dans cet esprit, le modèle Berk-Green fournit un cadre théorique adapté pour analyser ces produits particuliers.

Dans ce modèle, un investisseur ne connaît pas la capacité du gérant à offrir une sur-performance par rapport au rendement de référence. Il doit ainsi estimer cette capacité à partir des rendements historiques observés. Dans leur modèle, le rendement brut observé se décompose deux parties: (1) la capacité du gérant et (2) un terme d'erreur. Plus la deuxième partie est grande, plus il est difficile d'estimer la qualité du gérant. Ainsi, ce terme d'erreur représente la précision des rendements. L'investisseur ne reçoit que la partie nette du rendement moins le coût de transaction et les frais de gestion. Le coût de transaction augmente avec l'AUM du fonds. Cela correspond à une contrainte de capacité du gérant. Ainsi, on observe que plus l'AUM d'un fonds est élevé, plus il est difficile d'offrir une sur-performance.

Les investisseurs sont rationnels dans ce modèle: ils utilisent l'approche bayésienne pour mettre à jour leurs estimations sur la capacité du gérant. Les fonds recevant une bonne estimation vont continuer à accueillir des flux entrants jusqu'aux limites des capacités de maîtrise du gérant du fonds. A ce moment là, la sur-performance disparaît et les investisseurs ne souscrivent plus. Cela permet le marché des fonds d'atteindre l'équilibre pour lequel il existe une relation positive entre le rendement du fonds et ses flux, même si la sur-performance ne persistera pas dans le futur.

Ce modèle utilise les rendements passés du fonds comme la seule information que les investisseurs observent. Dans le même cadre d'analyse, [Berk and Binsbergen \(2016\)](#) inclue traditionnellement les différents facteurs de risque dans le modèle. [Brown and Wu \(2016\)](#) considère qu'un gérant peut gérer plusieurs fonds et que les investisseurs souscrivent ainsi à tous les fonds appartient à ce gérant s'ils perçoivent le bon signal. Pourtant, nous pouvons nous poser plusieurs questions quant au modèle Berk-Green. Premièrement, si les investisseurs ne sont pas rationnels, comment la relation entre les flux et les rendements observés est-elle modifiée? Deuxièmement, si un investisseur détient plusieurs fonds, le signal provenant d'un fonds va-t-il impacter d'autres fonds détenus par ce même investisseur? Enfin, est-il possible que les investisseurs n'aient pas les mêmes motivations d'investissement? Répondre à ces questions peut aider le gérant à anticiper les comportements des investisseurs et ainsi renforcer la liquidité de son fonds.

Modèle Chen-Goldstein-Jiang

Le modèle Chen-Goldstein-Jiang explique comment le "fund run" peut se produire. Puisque les flux sortants auront un impact négatif sur la performance future (voir, [Edelen 1999](#)), chaque investisseur surveille les comportement des autres investisseurs présents sont dans le même fonds. Ainsi, le comportement d'un client dépend en partie des transactions des autres clients. Si les investisseurs estiment que la liquidité d'un fonds est faible, ils peuvent décider de sortir même si les rendements sont bons. Ce comportement est différent de celui prédit par le modèle Berk-Green selon lequel un bon rendement attire les souscriptions et diminue les sorties. Dans ce modèle, la décision de chaque investissement dépend du comportement des autres investisseurs. Si de nombreux clients sortent du fonds, cela va inciter les autres investisseurs à sortir également.

Le modèle a deux périodes, appelés 1 et 2. Les investisseurs choisissent entre (1) rester dans le fonds jusqu'à la fin de la période 2 et (2) sortir de manière pré-maturée à la période 1. Le rendement d'un investisseur restant jusqu'en période 2 dépend du nombre de sorties pré-maturées car le gérant doit payer une prime d'illiquidité quand il vend ses actifs afin de répondre à ces sorties. Ainsi, le rendement des clients restants suit une fonction dépendent de trois paramètres: la prime d'illiquidité, le nombre de sorties pré-maturées et la qualité du fonds. La prime d'illiquidité correspond au fait que le gérant paie un coût pour vendre les actifs à la période 1. La probabilité de rester dans le fonds est positivement corrélée avec la qualité du fonds et négativement corrélée avec le nombre de clients sortant à la période 1. Ces relations déterminent l'équilibre de ce modèle: lorsque la performance est mauvaise, les fonds illiquides, i.e. pour lesquels les gérants paient une plus grande prime d'illiquidité, souffrent d'un plus grand nombre de sorties.

Cette théorie souligne une endogénéité de sortie: la sortie d'un investisseur a un impact négatif sur la performance du fonds et cet impact incite d'autres clients à sortir. Ainsi, le risque de "run" est plus important si le fonds est illiquide. Cependant, cette étude ne porte son analyse qu'au niveau du fonds. Nous voudrions poursuivre l'analyse de "fund run" à un niveau moins agrégé, i.e. étudier la décision de sortie des clients individuels. Dans cette perspective, nous cherchons à comprendre: (1) si certains investisseurs ont des comportements moutonniers et sortent parce que d'autres sont sortis. (2) Si les investisseurs répondent aux différents stimuli ou différemment à un même stimulus. (3) Comment ces comportements individuels vont déboucher sur un "fund run" et quels sont les facteurs déclenchants? C'est à l'ensemble de ces questions que nous cherchons à répondre via les travaux de cette thèse.

Plan de la thèse

Ces deux modèles théoriques nous font découvrir les comportements des investisseurs dans un cadre rationnel. Dans la continuité de ce sujet, nos travaux de recherche couvrent le côté empirique. Afin de répondre aux besoins industriels et réglementaires de la gestion du passif, nous examinons l'historique des micro-transactions des investisseurs et développons des modèles statistiques permettant de prévoir les flux entrants/sortants. En comparaison aux travaux précédents portant sur le côté actif du fonds, la difficulté du côté passif est le manque des données. Ainsi, la première étape de nos travaux est de construire une base de données sur le passif du fonds en collectant, en temps réel, les transactions effectuées par des investisseurs dans trois sociétés de gestion françaises.

Le premier chapitre de cette thèse présente le contexte académique propre à notre sujet. Nous effectuons d'abord, une revue de la littérature portant sur les comportements des clients des fonds mutuels. Ensuite, nous comparons les différentes sources de données utilisées par les études antérieures. Puis, nous présentons notre nouvelle base de données, qui procure des informations plus riches que celles trouvées dans la littérature. Cette base offre des opportunités uniques pour analyser les comportements des investisseurs. Elle nous permet de développer des originalités dans notre recherche à trois niveaux. D'abord, nous étudions les propriétés des séries temporelles des flux d'un fonds, là où la littérature antérieure s'intéresse surtout à l'étude de la coupe transversale de fonds différents. Deuxièmement, nos travaux suivent les transactions individuelles des investisseurs. Ils complètent les anciennes études qui n'observent que les flux agrégés au niveau du fonds. Enfin, notre base de données rend possible l'étude des liens et de l'hétérogénéité entre les investisseurs. Cela apporte une contribution à la littérature existant qui considère que les investisseurs sont homogènes et indépendants. Cette base de données nous permet de réaliser trois articles de recherche qui s'intéressent au lien entre les comportements des investisseurs et le risque de liquidité des fonds ouverts. Nous les présentons dans les chapitres 2, 3 et 4 de cette thèse. Chaque article examine cette base de données d'un point de vue unique, bien que les trois études ont toutes le même objectif d'aider le gérant du fonds à surveiller ses clients et renforcer la liquidité du fonds.

Le chapitre 2 vise à améliorer la prévision des flux futurs du fonds. Les propriétés statistiques des données conduisent à l'utilisation d'un modèle de comptage¹⁰. Inspirés par les études sur la modélisation de la volatilité, nous commençons par un modèle basique,

¹⁰Un modèle de comptage décrit les arrivées d'un événement, dans notre cas, l'évènement est une/un souscription/rachat. Une particularité de ce modèle est qu'il couvre des séries "integer" qui ont des valeurs discrètes et positives

i.e. un modèle Poisson indépendant et homogène. Nous ajoutons ensuite graduellement des différentes composantes à ce modèle. De cette manière, notre modèle capture tous les faits stylisés des flux observés. Nous trouvons que les investisseurs sont impactés par les flux précédents des autres clients. Ainsi, les arrivées de flux présentent une auto-corrélation. La littérature nomme cette propriété "auto-existant". Puis, nous trouvons que les flux ont une variance beaucoup plus grande que celle dans un modèle basique. Cela est expliqué par le fait que les investisseurs dans un même fonds sont hétérogènes. Ainsi, les séries des flux présentent une sur-dispersion. Ces deux faits stylisés sont générés par les caractéristiques et les comportements des clients, qui sont des éléments provenant du passif d'un fonds mutuel. Cela différencie notre travail de la littérature qui n'étudie que le côté actif du fonds. Nous introduisons alors la notion du **risque de liquidité au passif**.

Dans un deuxième temps, nous examinons les dynamiques des flux dans les différentes catégories d'investisseurs. Puis, nous analysons l'impact de certains facteurs explicatifs utilisés par les études antérieures. Nous observons que, même si les caractéristiques des actifs et les facteurs économiques ont des impacts non négligeables sur les dynamiques des flux, le comportement des clients reste un facteur significatif. Cette étude aide un gérant de fonds à obtenir des statistiques fiables sur les flux et réserver ainsi suffisamment de titres liquides pour satisfaire les rachats futurs des clients. Ensuite, la propriété "auto-existant" suggère que le gérant doit ajuster sa prévision des flux selon les derniers flux observés, de manière analogue à un modèle GARCH pour la prédiction de la volatilité. Cette étude répond directement aux besoins industriels et réglementaires de gérer la base d'investisseurs du fonds et de contrôler la liquidité au passif. Elle fait partie du projet "Modélisation et Gestion du Passif des Fonds", entreprise en collaboration avec plusieurs partenaires industriels. A l'issue de ce projet, un outil informatique de gestion de liquidité qui inclut notre modèle est développé et sera prêt à être commercialisé ¹¹.

Nous pouvons que les caractéristiques des clients ont des effets significatifs sur leurs flux. Cela nous incite à étudier les comportements des investisseurs au niveau individuel. Le chapitre 3 analyse les horizons d'investissement des clients individuels. Ces horizons sont modélisés par le modèle de survie qui explique la probabilité de rachat individuel à chaque période sur la base d'une grande variété de facteurs économiques. Nous observons que les horizons d'investissement sont fortement liés aux caractéristiques du fonds et de l'investisseur. De plus, les investisseurs sont moins rationnels en fin d'année car à ce moment ils sont plus susceptibles de commettre un biais comportemental, l'aversion à la perte.

¹¹Voir, <https://www.agefi.fr/asset-management/actualites/hebdo/20171012/gerants-en-passe-mieux-connaître-leur-passif-228922>

Enfin, nous trouvons que les investisseurs de long-terme sont plus sensibles aux conditions économiques que ceux de court-terme. Ces résultats introduisent la notion du **risque de sortie pré-maturée**, i.e. un investisseur sort d'un fonds plus tôt que l'horizon prévu au regard de ses caractéristiques. Notre recherche montre encore que ce risque est plus grand pour des investisseurs de long-terme.

Cette analyse de survie a plusieurs implications pour les praticiens. D'abord, à partir des horizons d'investissement prévus, nous pouvons calculer la durée moyenne du passif du fonds, pour la quelle la littérature a eut des difficultés à mesurer. Cela nous permet de mieux comprendre l'écart de liquidité entre l'actif et le passif. Afin de gérer ce problème, la société de gestion peut s'inspirer des techniques de l'ALM ("Asset Liability Management"), qui est beaucoup utilisée par d'autres institutions financières, comme les banques ou les assurances par exemple. Cela va également aider le gérant à choisir une meilleure allocation des actifs qui s'adapte aux comportements des clients. Par exemple, si la plupart des investisseurs restent dans le fonds pour des horizons courts, le gérant doit investir de façon plus marquée dans les actifs liquides. Deuxièmement, selon les résultats de cette étude, nous fournissons certains conseils commerciaux à la société de gestion. La prédiction des horizons d'investissement des investisseurs va aider le gérant à mieux cibler les clients "profitables". Par exemple, il faut présenter un fonds liquide à un client de court-terme et un fonds illiquide à un client de long-terme. En outre, la société de gestion pourra introduire des mécanismes de protection de liquidité comme la période d'inaccessibilité, si elle sait dans combien de temps chaque investisseur va sortir du fonds. Enfin, notre étude incite le gérant à surveiller le risque de "run" de manière individuelle. En période de crise, le gérant doit surveiller plus particulièrement les investisseurs de long-terme.

Notre base de données couvre une grande gamme d'investisseurs comme les banques, les compagnies d'assurance, les FoFs (fonds de fonds) et aussi les fonds de pension. Cela diffère de la littérature qui ne contient que la catégorie des investisseurs particuliers ("retails"). Nous posons ainsi une question associée: les investisseurs institutionnels sont-ils différents des investisseurs particuliers? Nous supposons que ces investisseurs adoptent une approche d'investissement plus sophistiquée: ils ciblent une allocation d'actifs et rééquilibrent leurs portefeuilles ("**rebalance**") vers l'allocation ciblée. Le dernier chapitre de cette thèse a contribué à ce sujet. Nous débutons cette étude par l'analyse des transactions d'investisseurs individuels. Nous trouvons que plus de 65% des investisseurs de notre échantillon détient plusieurs fonds qui opèrent dans des classes d'actifs différentes. De plus, les investisseurs exécutent beaucoup de transactions: 20 souscriptions/rachats par an. Ces deux observations sont des évidences de "rebalance". Puis, nous continuons en modélisant les transactions

des clients par un modèle probit à trois dimensions: souscrire, racheter et tenir. Nous identifions une relation négative entre l'intensité de l'achat et la performance du fonds. Cela suggère que les investisseurs adoptent l'approche de "constant-mix rebalance": ils vendent les titres appréciés et achètent les titres dépréciés, afin de maintenir l'allocation initiale des actifs. Enfin, nous étudions comment les investisseurs rééquilibrent leurs placements entre des fonds qui ont différents niveaux de liquidité et de risque. Cette étude fournit aux gérants de fonds un outil pour gérer les sorties individuellement. De plus, elle nous laisse réfléchir sur une question: quels types de services une société de gestion doit proposer aux investisseurs? D'après nos résultats, certains investisseurs savent déjà diversifier et rééquilibrer leurs portefeuilles. De ce fait, ils n'ont pas besoin d'un service de gestion passive et diversifiée. Une société de gestion doit donc leur proposer des styles de gestion plus sophistiqués.

Chapter 1

Background on Mutual Fund Liability Risk

1.0.1 Introduction

Our study focuses on fund investor's behaviour and its implication on the fund liability risk. We aim at providing new statistic tools to enhance the mutual fund liquidity and prevent the "fund run" scenario. This objective answers the regulatory requirements and industrial needs. This study relates to several strands of literature. In the first part of this chapter, we briefly present the main results of the previous studies. Then we make suggestions about how to extend these investigations. We find that, the data used by the literature contain only the inadequate information on investor's behaviour. It motivates us to collect a novel data-set which records more information on mutual fund investors and their behaviour. We present this data set in the second part of this chapter.

1.1 Literature review

Mutual funds play an important role in the modern financial system, which partially explains why it is receiving more and more attention in the academic literature. Among a diversity of previous studies, our work is most tightly linked to four strands of literature: (1) mutual funds performance and their risk exposure; (2) mutual funds' liquidity risk, (3) the fund run scenario and (4) fund flows. These studies highlight the essential institutional background on open-end mutual funds and the preliminary evidence on investor behaviours. Hence it is necessary to present several representative studies in these topics. We start by explaining the link between mutual funds performance and their risk exposures. Next, we describe mutual funds' liquidity risk and how fund managers mitigate that risk. Then we characterizing the famous fund runs that occurred in the financial history. An important step to avoid these crisis scenario in the future is to understand the properties of fund flows. Hence, finally we present the different characteristics of fund flows found in the literature.

Performance and risk exposure

Investors are sensitive to fund performance, by which they learn the quality of the fund and its manager. Therefore, a large strand of literature investigates the fund performance and its risk exposures. According to the common belief, the fund return persistence is a signal of fund managers' ability. However, [Carhart \(1997\)](#); [Bollen and Busse \(2004\)](#) among others, showed that the persistence in fund return is due to the fact that the fund holds momentum stocks and not necessary a proof of this ability. As a consequence, the momentum risk factor should be considered when assessing funds' performance. The Carhart's 4-factor

model, which adds the momentum factor to the traditional Fama-French 3-factor model, has become a standard tool to evaluate mutual fund return. Besides, the recent literature has also found other factors which affect funds' performance. [Banegas et al. \(2013\)](#) identify several macroeconomic variables enhancing the predictive power of fund return. [Jordan and Riley \(2015\)](#) develop a volatility-related factor which explains a large part of abnormal fund returns. [Christoffersen and Simutin \(2017\)](#) show that holdings of high beta stocks affect fund returns. [Cremers and Pareek \(2016\)](#) indicate that "patient" funds outperform others. Therefore they suggest that mutual fund should hold securities in a longer horizon than what in the usual case. Interestingly, [Chaudhuri et al. \(2016\)](#) document that the fund managers who hold a PhD degree could achieve a 20% out-performance.

In short, previous studies seek to explain the cross-sectional differences in mutual fund returns. This traditional approach could be extended in several directions. Firstly, fund returns might have time series properties which are not covered by cross-sectional factors. In this spirit, [Pastor et al. \(2017\)](#) focus on the time series proprieties of fund return and develop a turnover factor to account manager's capacity to seize the time-varying profit opportunities. They show that their method outperforms the traditional cross-sectional model when: (1) the fund holds many illiquid securities (2) the manager processes great investing skills. Hence their study provides a practical tool to assess the fund manager. Second, the fund manager might have the "timing" capacities, which highlight the fact that the he modifies his portfolio in response of different economic conditions. For example, when a fund manager expects the market to go up, he increases fund's market risk exposure to earn a high risk premium. On the contrary, when the manager expects the market to go down, he decreases fund's market risk exposure. The literature refers to this behaviour as the market timing capacity (e.g., [Liao et al. 2017](#); [Tchamyou and Asongu 2017](#)). Previous research also identifies the timing capacities of other risk factors, such as the liquidity (see, [Cao et al. 2013](#)) and the volatility (see, [Busse 1999](#)). These studies suggest that funds' risk exposures evolve over time and we need to take this evolution into account when we assess a fund's performance. Third, whilst a large amount of research has been carried out on the fund's asset side, i.e. the performance of securities in a fund's portfolio, fewer studies have attempted to investigate the link between the liability side, i.e. the characteristics and behaviours of investors, and the fund's performance. [Berk and Tonks \(2007\)](#) demonstrate that investors' redemptions make bad return persist. [Edelen \(1999\)](#); [Fang et al. \(2015\)](#) show that investors' purchases and redemptions disturb the fund manager and hence they decrease the fund return. As a consequence, understanding and predicting the behaviour of investors become a must for fund managers. [Cuthbertson et al. \(2016\)](#) provide a comprehensive

review of all relevant studies on the impact of behavioural issues on fund's performance. Rohleder et al. (2017) document that funds using different instruments, such as derivatives products, to manage this redemption risk outperform funds without these measures. A further understanding of clients' behaviour would help fund managers to better predict the future flows and allows us to develop reliable tools to manage the redemption risk.

Liquidity risk and its management

Traditionally, mutual funds are believed to invest in relatively more liquid securities than hedge funds, with the latter taking the large liquidity risk in chase of superior return (See: Sadka 2010). As a result, a liquidity gap might occur between asset and liability in hedge funds. Aragon et al. (2017) find evidence that hedge fund managers adjust their portfolio to control for this gap and thus mitigate the funding risk. Investors and regulators pay less attentions to the liquidity problem in mutual funds since they consider their portfolios to be relatively liquid. However, previous studies highlight that mutual funds also take a significant liquidity risk. Dong et al. (2017) reveal that the liquidity risk exposure explains cross-sectional differences in mutual fund returns. This study implies that if mutual funds hold less liquid securities than what investors imagine, and they do not adopt liquidity protection provisions such as "gate" and "lock-up period" as hedge funds do, a more serious liquidity gap might exist in some mutual funds. Yet, there is hardly an empirical investigation for this problem, since there is no public data of mutual funds' liability. Therefore, an important part of our work concerns to the collection of fund clients' transaction records in order to prepare a unique liability dataset.

Previous studies find evidence that some fund managers make efforts to manage the liquidity risk. Nanda et al. (2000)'s theory predicts that the randomness of a fund's flow has an impact on its fee structure. They demonstrate that, in equilibrium, funds should charge investors who redeem less frequently a lower management fee. Chernenko and Sunderam (2016) find empirical evidence that mutual funds use cash account to meet future redemptions. First the fund manager estimates the future outflows, then she/he sets her/his cash reserve accordingly. However, the estimation in practice is often inaccurate, leaving the manager with an inadequate cash level. Hill (2010) document that fund flows exhibit often the heavy tails in their distribution and this property generates the large estimation errors. In fact, Desmettre and Deege (2016) show that the flow prediction becomes far more reliable when these heavy tails are considered. In line with their research: we seek to explain the flow's extreme risk analysing investors' behaviours and we intend to develop an reliable

model to give a more accurate prediction of future flows.

Fund run crisis

The scenario a fund manager wishes to avoid the most is the "fund run". In such a case, clients redeem massively and rapidly, making the fund lose a significant part of its capital in a short time. For example, on 16th September 2008, American institutional money market funds have lost 20% of their AUM on average in a single day (see, Figure 1, Panel A in [Schmidt et al. \(2016\)](#))¹. The funds that hold many illiquid assets are the most exposed to the run risk. For example, the famous LTCM is a fund which has invested mainly in illiquid fixed income securities (see, [Edwards 1999](#)). Another example of the run on illiquid funds are the massive redemptions in Germany estate funds in 2006. During this year, the whole German real estate fund market lost 40% capital (see, [Fecht and Wedow 2014](#)). [Chen et al. \(2010\)](#)'s model explains that unexpected redemptions from an illiquid fund generate negative externalities to the investors who stay. Therefore, its investors are more likely to exhibit the run-like behaviour. However, the run risk is not only present in illiquid funds, it can concern to liquid funds too, such as in the aforementioned money market funds. While the bulk of this research focuses on detecting the funds most vulnerable to the risk of run, very few of them are trying to spot the clients most likely to exit. In practice, however, fund managers are more concerned by the latter as they must monitor the run-like behaviours of each individual client.

The acute consequences of a fund run is not limited to the fund company and its clients as in the LTCM's problem, it has an insidious influence on the well-functioning of the whole financial system. On the one hand, a fund run could transfer the liquidity problem from risky securities to other "safe" securities. [Kacperczyk and Schnabl \(2013\)](#) show that during the 2008 fund run, the money market securities which are risk-less in the normal circumstances became riskier. As a consequence, even the purportedly safe portfolios are actually risky. [Hau and Lai \(2016\)](#) describe a crisis propagation between "good" and "bad" equity assets held in the same funds during a fund run. Initially, the worse-performing equities trigger the massive redemptions. In turn, however, the fund managers are forced to liquidate the well-performing equities to satisfy all these withdrawals, introducing a pressure on the price of these assets and the risk is transferred from the "bad" securities to the "good" ones. On

¹The money market fund run of 2008 is a market-wide event, however, the fund run could also happen to an individual fund. For example, the **Janus Global Unconstrained Bond Fund (JUCAX)** has suffered 18.5 million dollars withdrawals on February 2015. See: <https://www.barrons.com/articles/bill-gross-janus-mutual-fund-saw-first-monthly-outflow-in-february-1425921142>.

the other hand, the fund runs have also macro-level impacts across markets. [Chernenko and Sunderam \(2014\)](#) document that during the 2011 fund run on the European money market, funds were forced to invest much less in the short-term lending market than before. Consequently, European corporate firms found a shortage in the short-term financing. [Shek et al. \(2017\)](#) display how the brutal outflows in the developed countries' fund market has affected the emerging countries' bond market. After experiencing the fund run in 2008, the fund managers have modified their target portfolio and decreased the allocation of illiquid assets. Many of them have even stopped investing in emerging markets, which has led to a significant decline in the overall supply of capital in the emerging debt market.

An efficient approach to study fund runs is to examine the micro-level data, i.e. the data containing the information about managers' and clients' behaviour. [Aragon and Strahan \(2012\)](#), as well as [Khandania and Lo \(2011\)](#), investigate fund managers' transactions during 2008 crisis. They display the evolution of fund risk exposure during a run and their work helps people to understand how liquidity problems could happen to funds. An investigation of investors' transaction would provide a more direct insight into the run scenario. For example, ? show how individual bank clients behave during two Indian banking crisis. They highlight the mechanism of the bank run. They present strong evidence on the lead-lag relationship: some clients follow other to withdraw. Moreover, they show clients in the same geographical zone are often affected by the same liquidity shock. Their findings are precious for fund managers to prevent a run crisis. However, there are still very few papers analysing individual fund clients behaviour because this type of data is not easily accessible.

Fund flow

We see from the previous two subsections that there are difficulties to mitigate the liquidity risk and prevent the fund run scenario. The primary reason lies in the lack of data on the mutual fund's liability hence there is no reliable measurement of investor's behaviour. Without individual information on fund clients and their behaviour, the only possibility is to analyse the aggregated fund flows at the fund level (See, e.g., [Agarwal et al. 2015](#); [Akbas et al. 2015](#); [Bollen and Busse 2004](#)). The fund level flows approximate the net purchase/redemption of each fund, usually at a monthly frequency. The primary objective of these studies is to identify the determinants of fund flows. Several explanatory factors are found. Firstly, fund flows are affected by the economic conditions. For instance, [Cao et al. \(2008\)](#) find that the volatility at the market level has significant predictive power of fund flows. Besides, fund flows are strongly related to the characteristics of fund company and

fund managers. [Aydogdu and Wellman \(2011\)](#); [Lou \(2014\)](#) document that spending more marketing expenses can increase the fund flows. [Kumar et al. \(2015\)](#) find that manager's reputation is an important determinant of fund flows. [Copper et al. \(2006\)](#) show that the manager's name is another important factor to attract flows. Previous research also finds that the Morningstar rating ([Guercio and Tkac 2008](#)) and fund's ranking ([Kempf and Ruenzi 2008](#)) have both non-negligible influences on fund flows.

Along various explanatory factors, the most significant one is obviously the fund return. [Berk and Binsbergen \(2016\)](#) test the predictive power of different return models, such as the CAPM, the Fama-French 3-factor model, the Carhatt 4-factor model, etc. They find that the CAPM has the largest predictive power on funds' netflows. It suggests that most investors are considering the market risk as the only important factors when they purchase the fund. Instead of focusing on the netflows, some literature studies the inflows and outflows separately, since they might react differently to performances. They find that inflows reacts quickly to good performances while outflows are far less sensitive to bad performances, hence there is a convex return-flow relationship (see, [Agarwal et al. 2004](#), [Chevalier and Ellison 1999](#)). However, sometimes this relationship turns to concave, i.e. outflows are more reactive to performances than inflows (see, [Goetzmann et al. 2003](#), [Sherman 2012](#), [Ding et al. 2007](#)), or even becomes linear ([Baquero and Verbeek 2005](#)). [Getmansky et al. \(2015\)](#)'s theory documents that, for hedge funds, the form of the return-flow relationship has a link to the share restrictions and asset liquidity.

In the mutual fund market, most of funds exhibit a convex relationship between the flow and the performance². However, in two cases this relationship is highly likely to changes. First, previous research only tests the aggregated flow's sensitivity to returns and it is not clear whether this relationship holds at the individual level. Second, many of them investigate only equity funds (See, e.g., [Ivkovich and Weisbenner 2008](#)) which are generally liquid. In contrast, in less liquid bond funds, the convex relationship becomes concave (see, e.g., [Goldstein et al. 2017](#) and [Chen and Nan 2017](#)). Moreover, [Leung and Kwong \(2018\)](#) demonstrate that the convex relationship does not hold for funds investing in the emerging market. It motivates us to examine and compare the flow-return relationship in different asset categories.

The previous literature on mutual fund flows and liquidity risk can be extended in four dimensions. Firstly, the literature pays inadequate attentions to fund flows' statistic properties. However, the time series stylized facts of fund flows have strong implications on fund's liquidity risk. The manager should set the cash level according to flows' time series dynam-

²See the survey of [Ma \(2013\)](#) and references therein.

ics. Secondly, the literature analyses the fund flows at a low frequency, such as monthly. However, the analysis at daily frequency is recommended since the fund manager makes his investing decisions on daily basis. Thirdly, most of the previous literature examines the flows at the aggregated level for each fund while a study at the disaggregated level might present different results. On the one hand, inflows and outflows might exhibit different properties, thus we should not pool them together. On the other hand, investors in the same fund might behave differently justifying the need to investigate fund clients individually. Finally, the literature uses a unique horizon to calculate fund returns. Many authors only test how fund flows react to the return in the previous period. However, investors' flows might be affected by returns at different horizons. For instance, long-term investors are likely to be more sensitive to long-term returns than short-term investors. To address these extended research questions, we collect a new fund client's micro-transaction data set containing richer information than ones in the literature.

1.2 Transaction-based fund flow database

1.2.1 Data-base collection

In this thesis, we work with three French asset managers to built a new "Mutual Fund Liability" data set by hand-collecting individual fund client's transaction record. We refer to these three companies as A, B and C. Since the regulator does not require fund companies to disclose the information on the liability side of mutual funds, there is no industry standard to record these transactions. Each fund family records the investor data in its own way. Consequently, the quality of liability information is heterogeneous across fund companies. For instance, the fund family might "omit" to record the investor's identity, either the name or the number, for several transactions. To built a comprehensive raw data file, we first define a standard format containing all essential information which we need for the further analysis. Then we require three fund companies to collect the data in this way. We present this format in the table below ³:

"Date" is the day where the investor gives his trade order. The transaction is executed at the price of that day but is cash-settled in 2 days later. "Investor" represents the investor's identity, generally a code number. "Category" indicates the investor's type, such as the bank or the insurance company. ISIN (International Security Identification Number) refers to the security identification number of the fund share class. "Class" is the asset class that the fund

³The data in this table are fictitious and are given as an illustrative example

Table 1.1: Raw Data Example

Date	Investor	Category	ISIN	Class	Amount	Direction	Inventory
2005-11-02	0091	Retail	FR5600290728	Equity	12.00	Purchase	92 011.00
2005-11-02	0093	Bank	FR5400330783	Fixed Income	3213.56	Purchase	1 118.32
2005-11-03	0091	Retail	FR0000330700	Money Market	2017.11	Purchase	2 456.13
2005-11-05	0085	Insurance	FR5600290728	Equity	200.50	Redemption	0.00
2005-11-07	0093	Bank	FR0000330700	Money Market	3000.00	Redemption	85 001.15

invests in. "Amount" corresponds to the amount of the transaction. "Direction" indicates the direction of a transaction. It can be either a purchase or a redemption. "Inventory" gives the investor's position in the fund.

Table 1.2 presents the elementary information for our three data providers in Panel A, the quality of the data in Panel B, as well as the summary statistics for our fund sample in Panel C. We see first from Panel A that all three asset managers are affiliated to one or multiple parent company(s). This affiliation is popular among French fund companies. The parent company is an important source of fund inflows but it does not get involved in the investment decision. However, the weight of the internal client, i.e. the parent company, differs from one company to another. Company A has approximately 80% fund purchase from its parent company and 20% from external investors. This proportion is inverted in Company B: 20% internal and 80% external. Company C has a balanced investor source: half internal and half external. All three companies provide a full and general fund service, i.e. their funds cover a diversified range of asset classes: equity funds, fixed income funds, real estate funds, money market funds, etc. This allows us to have a more general insight than the previous literature which covers only one or two fund class(es). For instance, [Chevalier and Ellison \(1999\)](#) study only equity funds, [Keswani and Stolin \(2008\)](#) focus only on money market funds. All these funds are no-load thus investors are freely to purchase and redeem. The three companies all charge a standard management fee: a basic fee proportional to the fund share position (AUM fee) plus an incentive bonus for superior performance.

Concerning the quality of our data, we see in Panel B that all three companies successfully identify the investor category. However, non of them record the clients' personal profile. The data quality of company B seems superior as it is the only one to provide information on investor's individual trade identity, history and category, for each trade.

Panel C presents the summary statistics on data's size. Company B records its client data since its creation in 1988. Company A's history starts in 2004. Company C only provides two years of data. The three companies in a whole provide a sample which covers

Table 1.2: Full fund sample description

Fund Companies	A	B	C
Panel A: company			
Affiliated	Yes	Yes	Yes
Fund type	Diversified	Diversified	Diversified
Investor source	20% + 80%	80% + 20%	50% + 50%
Load	No	No	No
Fee	Standard	Standard	Standard
Panel B: data quality			
Investor Identity	No	Yes	No
Investor Profile	No	No	No
Investor History	Yes	Yes	No
Investor Category	Yes	Yes	Yes
Panel C: summary statistics			
Period	2004-2014	1988-2016	2013-2014
Fund (ShareClass) Number	26	128	16
Investor (Account) Number	626	601	Unknow
Transaction Number	183 408	368 218	740 307

This table presents the three fund companies in our panel and the quality of their data. The general information of fund companies is given in Panel A. "Affiliated" indicates whether the company is affiliated to a parent company. "Fund Type" explains which classes of funds the company manages. "Investor source" shows the weight of the internal and external investors. "Load" indicates whether investors should pay a load to purchase or redeem. "Fee" explains how the company charges the management fee. The data quality is presented in Panel B. "Investor identity" shows whether the company records the investor's name or code number for each transaction. "Investor profile" refers to the personal profile of each investor, such as age, salary, marriage situation etc. "Investor history" indicates whether the company records the history of transactions of each client. "Investor category" reveals whether the company records the sector of the investor for each transaction. The summary statistics are shown in Panel C. We give the number of fund share classes, investors, and transaction observed over the sample period.

more than 150 fund share-classes⁴ and more than 1200 client accounts. In sum, our sample represents around 1.2 million micro-transactions of fund clients.

1.2.2 Comparison to previous Fund flow datasets

Our dataset differs from the previously used ones in the literature. Since the fund flow is not publicly available, previous studies often extract this information from the size and return evolution using the equation below:

$$F_t = S_t - S_{t-1}R_t \quad (1.1)$$

Where F_t stands for the "fund flow" between $t - 1$ and t . S_t is the fund size at t and R_t is the fund return between $t - 1$ and t . Without any flow, the fund size at t should

⁴In average, each mutual fund has 2-4 share-classes which charge different fees to clients.

be the fund size at $t - 1$ multiplied by the return between $t - 1$ and t . Therefore, the subtraction in Equation (1.1) approximates the flow between $t - 1$ and t . A large amount of studies adopt this approximation, such as [Ippolito \(1992\)](#) and [Chevalier and Ellison \(1999\)](#). However, it suffers from several shortcomings. First, the flow in Equation (1.1) is a net value and inflows and outflows are not separated, even through they might exhibit different characteristics (see, e.g., [Jank and Wedow 2010](#)). Second, fund size are often reported at a monthly frequency. Hence this method does not provide the short-term dynamic of fund flows. However, the daily flow information is crucial for a fund manager to set his liquidity reserve. Thirdly, this method assumes fund clients are homogeneous and does not provide the individual flow pattern.

In the last decade, few new dataset providing more disaggregated information on fund flows have been created. In the US, some fund companies are required to disclose their monthly "sell" and "redemption" in the N-SAR file. This database supports the study on the aforementioned convex return-flow relationship (see, e.g., [Ma \(2013\)](#)). However, there are still several limitations: (1) the flows are still reported in a monthly manner; (2) there is no information on clients' heterogeneity; (3) This dataset covers only a small number of funds (less than 1 000) belonging all to the equity class.

Some private data providers, such as ICI ("Investment Company Institute") and iMoneyNet, provide money market fund flows on a weekly basis for the US. This database allows people to study the short-term dynamic of funding risks. [Schmidt et al. \(2016\)](#) describe the mechanism by which 2008 money market fund run has occurred. Using the weekly data, they show the dynamic channel of a fund run: institutional investors react strategically to retail investors' to exit the fund. [Kacperczyk and Schnabl \(2013\)](#) find an increase of risk taking incentives for fund managers, immediately after the beginning of the financial crisis, i.e. bankruptcy of the Lehmann Brothers. Nevertheless, these studies cover only money market funds and we still do not have access to the individual trade information of fund investors.

Our dataset offers new opportunities to understand mutual fund's liability risks. Comparing to the datasets used by the literature, the ours has the following characteristics and advantages:

- We have daily observations enabling us to examine the short-term flow dynamic and its time series properties.
- Inflows and outflows are recorded separately, allowing to distinct investor's purchase and redemption behaviour.

- Our sample covers a large range of assets categories: equity, fixed income, money market, diversified, etc.
- Each fund flow is recorded individually. It allows us to study the heterogeneity of investors' behaviours.
- We know investor's sector/category, such as bank, insurance companies, retail investors, pension funds, etc. As reported by [Grinbatt and Keloharju \(2001\)](#), investor's category might have great predictable power on his behaviour.
- We have the possibilities to examine, for each investor, his investment horizon and his holding period return.

1.2.3 Non-normality of fund flow arrivals

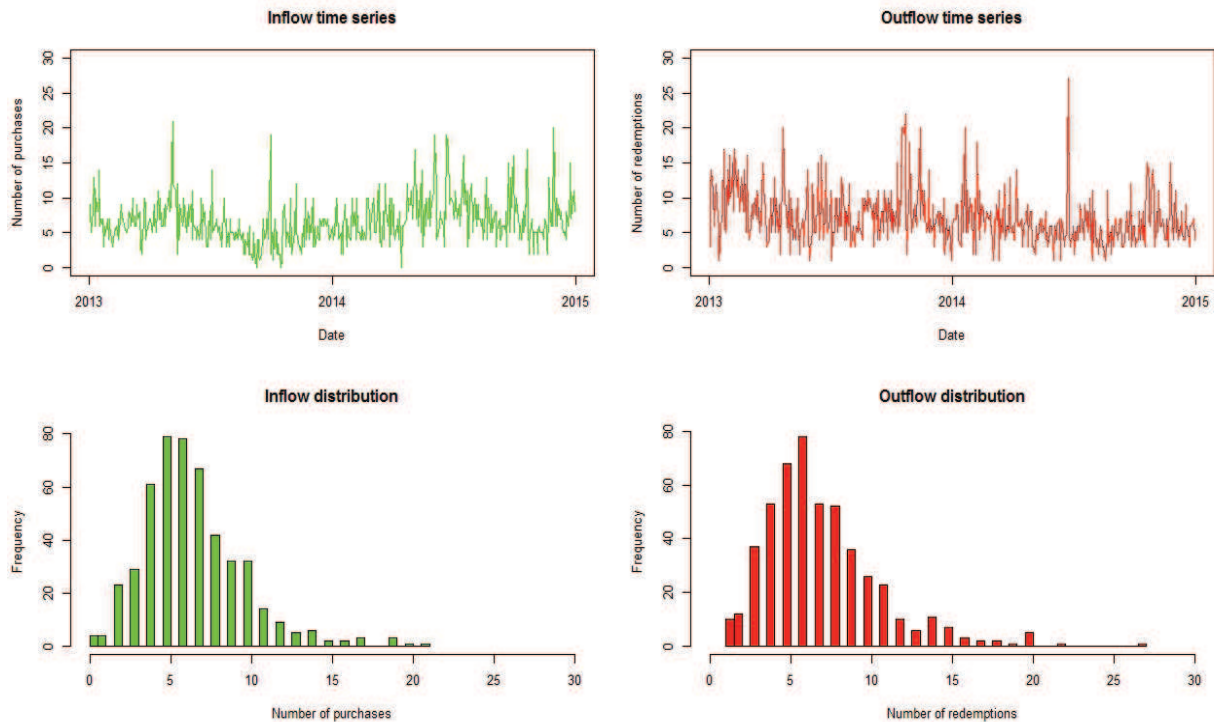
Our first concern is to improve the prediction of future fund flows. [Edelen \(1999\)](#) document that unexpected flows continuously disturb the fund management and decrease its performance. It implies that most of the fund managers under-estimate future outflows, and consequently, their cash reserves are inadequate to satisfy clients' redemption. Therefore, they sell relative illiquid assets at a disadvantageous price, aiming at recover this "cash gap". To set an accurate cash reserve, the fund manager should examine different statistic properties of fund flows, such as the tails of the flow distribution and the auto-correlation of the flow series. The former indicates the extreme outflows a fund manager might suffer and the latter would help the fund manager to predict the consequences of a liquidity shock.

For ease of presentation, we choose one fund to show flow's statistical properties in this chapter⁵. We present the times series patterns and the distribution of transaction numbers in Figure 1.1. Three main stylized facts are observed. Firstly, fund flows are discrete and always positive. When the transaction number is large (> 20 per day), a continuous distribution would give an acceptable approximation for this integer-value distribution. However, when the transaction number is low (< 20), the performance of a continuous model would worsen (see, [Liesenfeld et al. 2006](#)). This is exactly the case in our example where less than 20 daily redemptions arrive on average. This property motivates the use of the Poisson-type model. Besides, the two distributions in Figure 1.1 are both asymmetric. It implies that fund flows do not follow a Gaussian distribution.

⁵In the next chapter, "A Self-Exciting Model for Mutual Fund Flows", we choose also one fund for the illustration. However, two funds chosen by two chapters are not the same one.

Secondly, we observe from Figure 1.1 that both the inflow and the outflow series exhibit clustering effects. They can be confirmed by the significant auto-correlations of fund flows in Figure 1.2. This property should be integrated into the Poisson model.

Figure 1.1: Times series statistics of fund flows

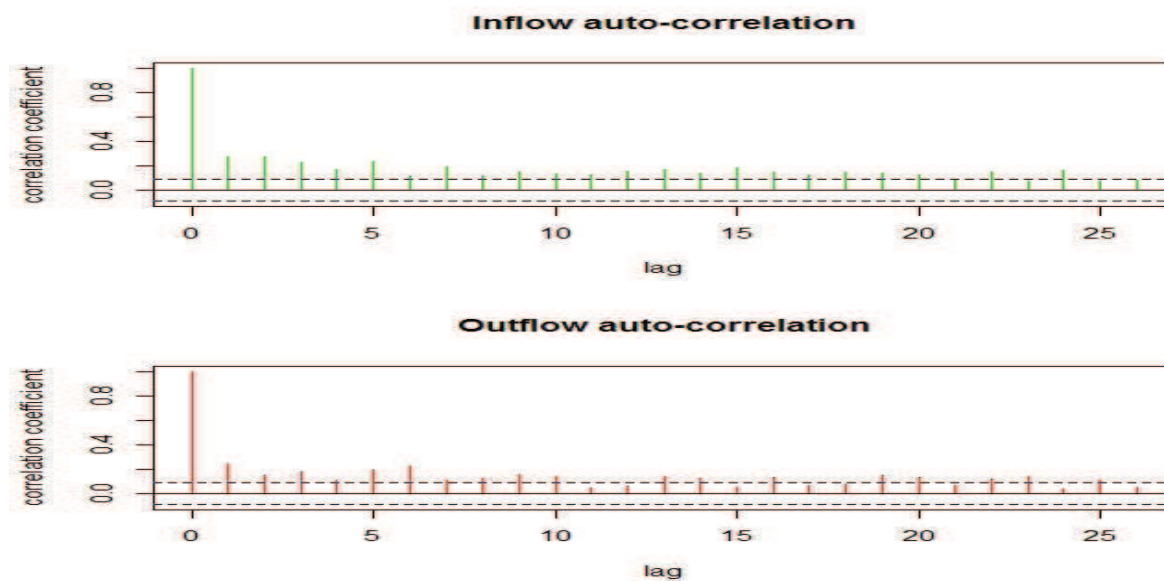


In this figure, the two top graphs give the daily evolution of flows over a 2-year period, from the 2nd February 2013 to the 30th December 2014. The two bottom ones give the distribution of the flow series during the same period.

Lastly, we can observe numerous extremely large "peaks" in fund flows (Figure 1.1). The frequency of the extreme flows seems to be higher than what a homogeneous Poisson distribution generates. The comparison in Figure 1.3 highlights this divergence in variance. We use the empirical intensity, set to the daily average of flows, to simulate two flow series according to the homogeneous Poisson distribution. We then compare the simulated series to our data. There is a clear evidence that the simulated series' distribution is much narrower than that of the historical data. Therefore, the simulated series do not reproduce the extreme scenarios in the data and the traditional Poisson model need to be adapted to capture this characteristic.

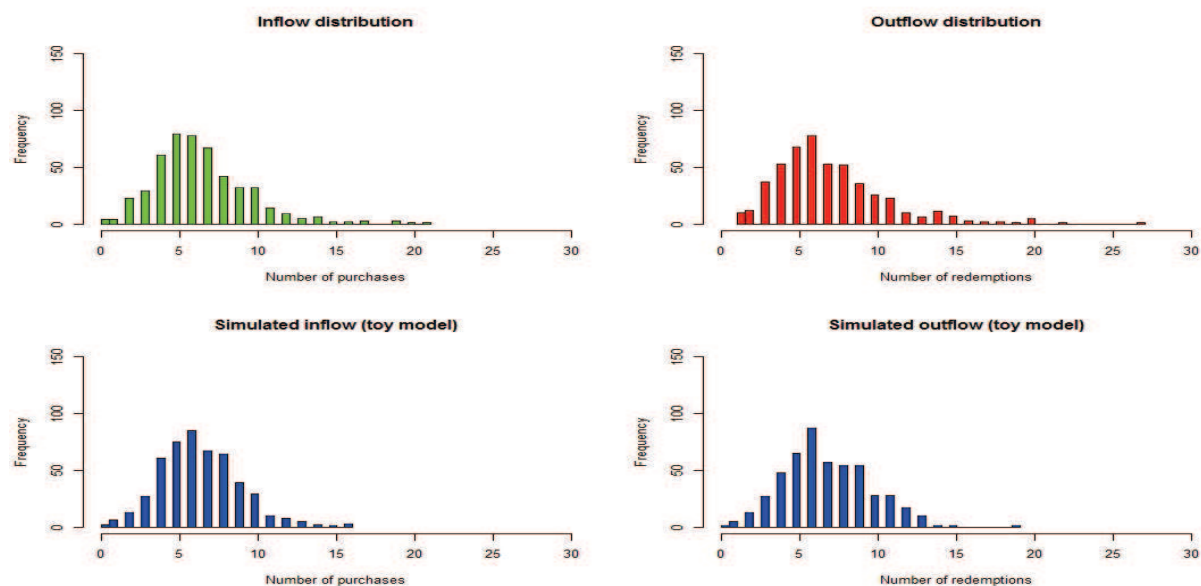
Overall, fund flows exhibit three key stylized facts: the discreteness, the clustering effect and the large dispersion. These properties make the flows' distribution deviating from a traditional continuous Gaussian model and motivate us to understand the source of these statistical patterns.

Figure 1.2: Auto-correlation of fund flows



This figure shows the auto-correlations of inflows and outflows of the fund example. We show inflows' auto-correlation in the left and outflows' at the right. The calculation of the auto-correlations follows the traditional approach which we use for continuous variables. However, coefficients' values are only approximative since the flow series are discrete. "ACF" stands for auto-correlation function.

Figure 1.3: Simulation of the "Toy" Poisson model



This figure shows the comparison between the flows data and simulation series of a homogeneous Poisson model. In this toy model, we simulate flow series with an intensity equalling to the sample average. We show the historical data on the top and the simulation on the bottom.

1.2.4 Heterogeneity among Fund Investors

Our second interest is to study the heterogeneity among fund investors and to explain the differences in their behaviours. Although a considerable amount of literature studies how investors trade, only few research investigates the differences among them. Previous studies such as Iyer and Puri (2012), Johnson (2004), and Ivkovich and Weisbenner (2008) cover only one category of investors: the "retail" client. In contrast, our dataset contains a large range of investor categories, retail clients as well as different institutional investors such as banks, insurance companies, FoFs and pension funds.

This diversity of the investor base has several implications to the fund liquidity management. Firstly, investors have different motivations in fund investing. For instance, those who invest in funds to earn superior returns, the "value" investors, might behave differently than those who use funds as cash management tools, the "liquidity" investors. The evolution of fund performance should affect more "value" investors and the liquidity condition should affect more "liquidity" investors. Secondly, Investors do not have the same investing horizon. Theses horizons are crucial for the liquidity management since they determine the average duration of fund's liability. The liability duration is a constraint for the fund manager and he should make his asset allocation accordingly, otherwise a liquidity mismatch between fund's asset and liability would appear and it is detrimental to fund's performance. Lastly, investors' purchase and redemption decisions might be time-varying. For example, they may redeem more at beginning or at the end of the year. It implies that the intensity of fund flows evolve over time and the liquidity risk changes accordingly.

We present the information of different investor categories In Table 1.3. There are 14 different sectors in total and almost each fund contains these 14 sectors. Firstly we need to explain several abbreviations in the table: "CGP" and "CI BTP" are French pension funds, "IRP" is a type of French insurance companies and "IFA" sands for independent financial advisor. Investor's transaction frequency seems to be linked to his sector. For instance, "Bank (private client)" and "Mutual insurance" are examples of active clients. They make more than 1 transaction per week. In contrast, family offices (IFA) have lower trade intensity: they execute only 1 trade over several months. Some investors are trading large amounts. For example, a single transaction of pension funds (CI BTP) might represent 5% to 10% to the fund AUM. On the contrary, insurance companies' trade sizes are often smaller 0.01% to the fund AUM.

Furthermore, we calculate two types of correlations for these sectors. We first compute the auto-correlation of each sector's flows, the "Inter-correlation", then we estimate the cor-

Table 1.3: Presentation of activity sectors

Sector	Type	Frequency	Trade size	Inter-correlation	Intra-correlation
Association & Foundation	Institution	Low	Large	Yes	No
Bank (own)	Institution	Low	Large	Yes	Yes
Bank (private client)	Institution	High	Small	Yes	Yes
CGP	Institution	Medium	Large	No	No
CI BTP	Institution	Medium	Small	Yes	No
Insurance	Institution	Medium	Large	No	No
Consultant	Institution	Medium	Medium	No	No
Corporate Firms	Institution	Medium	Large	No	No
IFA	Institution	Low	Large	No	Yes
IRP	Institution	High	Small	Yes	Yes
Mutual Insurance	Institution	High	Small	Yes	Yes
Group	Institution	Low	Large	No	Yes
Retail	Retail	High	Small	Yes	Yes
Asset management company	Retail	Medium	Large	No	Yes

The column "Type" indicates whether the investor is institutional investor or retail investor. "Frequency" indicates the trade frequency: "High" (more than 1 trade per week), "Medium" (1 trade per week to 1 trade per month) or "Low" (less than 1 trade per month). "Trade size" indicates average trade size: "large" (more than 1% of fund AUM per trade) or "small" (less than 1% of fund AUM per trade). We further calculate the auto-correlation of flows in each sector ("Inter-correlation") and the correlation between between each sector and the rest of the fund ("Intra-correlation"). "Yes" reveals there is a significant correlation coefficient and "No" indicates that the correlation is not significant.

relation between each sector and the rest of the fund, the "Intra-correlation". Insurance companies have the high inter-correlation, which might be explained by the insurance subscription cycle. In some periods, insurance companies receive continuously high subscription volumes, therefore, they continue to purchase the mutual funds. On the contrary, insurance companies often face sequent policy claims, hence they redeem their fund shares accordingly. Family offices (IFA) and pension funds exhibits some intra-correlations. A plausible interpretation is that these investors are highly skilled, hence they often react to other investors' fund flows (see, [Schmidt et al. 2016](#)).

In Table 1.4, we present the weight of each sector with respect to their number of trades. Over the sampling period, investors have been executing 1 291 935 purchases and redemptions pooled together.⁶ The investor base is diversified since no sector represents more than 15% of the overall transactions. Insurance companies are largest investors. Their trades represent 10.44% of the total number. It might be explained by the nature of their activity: they continuously receive the subscriptions and pay policy claims. "Group" represents parent companies of the fund family. Their trades represent 8.72% of total transactions. "Retail" contains all individual clients who invest in mutual funds through a brokerage service. They execute 6.49% of overall transactions observed. Other sectors are less important.

In summary, the diversity of the investor base implies a high heterogeneity among investors. Investors' category, motivation and liquidity needs all influence their trade be-

⁶The trade numbers in Table 1.4 are the "raw" numbers before any data cleaning process. In the rest of this thesis, the trade numbers are slightly modified for multiple reasons, such as deleting the outliers.

Table 1.4: Weight of each investor sector in the full sample

	Sector	Trade	Weight
1	Association / Fondation	1,923	0.45%
2	Bank (own)	3,155	0.73%
3	Bank (private client)	24,105	5.60%
4	CGP	105	0.02%
5	CI BTP	1,712	0.40%
6	Insurance	5,200	1.20%
7	Consultant	103	0.01%
8	Corporate Firms	10,985	2.55%
9	IFA	660	0.15%
10	IRP	10,134	2.35%
11	Mutual Insurance	134,277	31.18%
12	Group	111,951	25.99%
13	Retail	82,329	19.11%
14	Asset management company	44,108	10.24%
15	Total	430,645	100%

This table presents the weight of each sector in the sample, with respect to their trade number. The column "Trade" indicates the number of trade, either purchases or redemptions, executed by all clients in the sector. In the last row, we present total trades observed. The column "Weight" is the trade number divided by the total trade number.

haviours. Hence each investor has his own liquidity risk contribution to the fund. Therefore, a study at the individual level is preferred so that the fund manager might better forecast the timing and magnitude of future redemptions.

1.2.5 Conclusion

Although there is a growing attention of the financial literature on the mutual fund industry, the traditional fund flow data-set limits the ability of previous investigations. Most of the previous research focus only on the asset side of the fund liquidity risk, whereas the liability risk is often ignored. Then, the literature studies the fund flow and fund run risk at an aggregated level, however, the individual analysis at client level merits more attentions. Our dataset contain rich information about individual fund investors' profile and their behaviour. It enables us to understand more about the "liability risk" and therefore to build quantitative tools to manage it. In the each of three following chapters, we use our data-base in a particular way. Nevertheless, all three studies intend to understand the fund investors' behaviour and their implications to fund liquidity risk.

Chapter 2

A Self-Exciting Model for Mutual Fund Flows: Investor Behaviour and Liability Risk

This chapter is linked to the research paper "A Self-Exciting Model for Mutual Fund Flows: Investor Behaviour and Liability Risk", co-written with Serge DAROLLES, Gaëlle LE FOL and Yang LU.

This paper analyses the purchase and redemption behaviour of mutual fund investors and its implications on fund liquidity risk. We collect a novel set of proprietary data which contains a large number of French investors holding funds with various degrees of asset liquidity. We build a Self-Exciting Poisson model which captures fund flows' statistic properties such as clustering effects and over-dispersed variance. The model improves the forecast accuracy of future flows and provides reliable risk indicators like Flow Value at Risk. Accordingly, we introduce the notion of liability risk where investor's behaviour increases mutual fund liquidity risk. We further decompose fund flows into investor categories. We find that investors exhibit high heterogeneous behaviour, and a lead-lag relation exists between them. Finally, we control flow dynamics for various economic conditions. We show that although flows evolve under different economic conditions, investor's behaviour stays the main significant determinant of flows' randomness. Our findings encourages fund manager to adopt an ALM (Asset-Liability-Management) approach: considering this client risk while making the capital allocation decision.

Key words: open-end funds, redemptions, fund run, self-exciting proceress, Asset-Liability Management

Introduction

Daily open-end mutual funds collect investors' capital to invest in a diversified range of securities, while allowing investors to freely purchase/redeem whenever they wish. As a consequence, mutual funds are exposed to a liquidity transformation risk. They receive short-term liabilities (clients' capital) and invest in longer-term assets. The capital pooling process may ensure part of clients' liquidity needs but the management of this liquidity insurance requires an accurate estimation of the timing and the amount of redemptions. However, fund managers sometimes incorrectly estimate this risk, suffering a liquidity mismatch because they do not have enough cash to satisfy investor's redemption needs. This problem has triggered numerous failures for mutual funds¹ and deserves in-depth studies on fund liquidity risks. A large amount of research investigates the fund asset side, i.e. analysing the liquidity of financial instruments funds are holding. Yet few study focuses on the liability side due to the limitation of fund client data.

We propose in this paper a new approach to estimate the funding liquidity risk supported by the fund manager using individual data. We build a new fund liability database by collecting individual investor's purchases and redemptions from existing open-end mutual funds. Our sample covers diversified fund classes and investor categories. At first glance, this database allows us to monitor the daily fund size variation and allows us to examine the disaggregated components of fund liability. We compute investor purchase (inflow) and redemption (outflow) separately to compare their distinct dynamics, whereas previous studies often aggregate them. Furthermore, we observe the behaviour of investors from each category, and assess its individual liquidity risk contribution to a fund. This database motivates us to develop a new client-management tool for fund companies in order to enhance the liquidity level of mutual funds.

Our model captures various stylized facts of fund flows time series on the one hand and provides an improved forecast to future fund flows on the other hand. We give the economic interpretation for each stylized fact observed and introduce accordingly the notion of liability risk. Furthermore, we analyse the linkage and the heterogeneity among fund investors. We study fund flows at the fund level and develop a self-exciting Poisson model to count fund flow arrivals on a daily basis. This model includes two major statistical properties of fund flows: the self-exciting captures the clustering effect and the over-dispersion aims to adjust for the accurate level of variance. This study helps us identify the source of a liquidity crisis, and how liquidity shocks transfer through investors. Lastly, we study the liability risk under

¹C.f. "Biggest Mutual Fund Failure Since 2008" in **ETF.COM**

various economic conditions. We compare the trading activities of investors in funds with different returns and asset classes.

We obtain several important findings. First we find that these stylized facts are direct consequences of investor behaviour. Self excitement is coming from correlation in transactions between investors, and over-dispersion is generated by the heterogeneity of fund client-base. The model succeeds in decreasing back-testing errors by 65% compared to the inaccurate static model. It provides a reliable calculation on extreme risk ratio such as flow Value at Risk. Second, we disaggregate global flows across investor's categories. On the one hand, we identify the lead-lag effect among fund clients. During the period of massive redemptions, some investors are leaders - redeeming first -, while some others are followers - exiting the fund in a second wave. We also find that the lead-lag relationship is not symmetrical, meaning that leaders do not react to followers. On the other hand, investors of the same fund belonging to different categories cannot be considered as identical since their flows are highly heterogeneous. Finally, we control for the economic factors, such as fund return and asset classes. We find that even though various economic factors have non-negligible influences on fund flows, impacts of fund liability (and client's behaviour) is always strongly significant.

Our paper contributes to the literature in three ways. First, our result is related to the literature on mutual fund liquidity risk. [Dong et al. \(2017\)](#) find that the liquidity risk exposure can explain the cross-section of mutual funds returns. It advises people to monitor fund managers' behaviour since they have a strong incentive to take additional risk to earn higher management fees (see, [Chevalier and Ellison \(1999\)](#)). This problem is even more acute in illiquid fund classes. [Goldstein et al. \(2017\)](#) highlight the fragility of corporate bond funds with low liquidity reserve. [Fecht and Wedow \(2014\)](#) describe how real estate funds are vulnerable to the liquidity risk. Illiquid funds are not the only ones, however, to be dealing with liquidity risk; liquid funds are exposed too. [Kacperczyk and Schnabl \(2013\)](#) show that money market funds, which are considered as extremely liquid investments, suffered a high liquidity risk during the 2008 financial crisis. These studies evaluate the fund's liquidity by studying the securities they are invested in (asset-side liquidity). We show that there is another component to the fund liquidity risk, namely the liability risk. The new liability risk differs from asset-related risk exposures and is present in all fund categories.

Second, our results contribute to the growing literature on mutual fund flows. Previous research seeks to identify the economic factors which determine the volume of fund flows. Commonly used factors include fund return ([Berk and Green 2004](#); [Berk and Binsbergen 2016](#)), fund risk level ([Spiegel and Zhang 2013](#)), market volatility level ([Cao et al. 2008](#)),

fund companies' marketing expense (Lou 2014), etc. Furthermore, psychological factors might also affect fund flows. Kumar et al. (2015) document the significant influence of fund manager's name and reputation on fund flows. Kamstra et al. (2017) show that sentiment index has a predictive power on fund flows. Although previous authors pay a great attention to this topic, their studies remain in the static framework² and ignore the dynamic patterns of fund flows. Barber et al. (2016) document that fund investors exhibit a time varying behaviour, and fund flows evolve accordingly. Furthermore, Cashman et al. (2008) show that previous findings might be biased as they ignore flows' non-linearities. We naturally wonder whether there are other statistical issues. We aim at filling this gap by analysing the time series of fund flow data. We build a dynamic model which includes flows' stylized facts such as clustering effect and stochastic volatility. This model improves the prediction of future fund flows, and can be used as a reliable tool to manage fund liability risk.

Our findings also shed light on the fund investors' behaviour by showing that they adjust their holding for fund risk exposure. Our approach puts emphasis on fund clients' linkage which is frequently ignored in the literature. Keswani and Stolin (2008) find that institutional investors behave differently from retail investors and exhibit a smart money effect. Chalmers et al. (2001) show that some investors can detect the mis-pricing of mutual funds and profit from this arbitrage opportunity. Puy (2016) document that mutual funds investors from the same geographical zone might suffer from the correlated liquidity shocks, which generate correlated redemptions. Friesen and Sapp (2007) examine the timing ability of fund investors. Brown and Wu (2016) identify that fund investors exhibit cross-fund learning ability within the same fund family. Focusing on stressed scenarios and Fund Run crisis, we show clear evidence that liquidity shocks are contagious: during a run scenario, some investors would run out the fund, mimicking other investors. Fund managers can use our results to monitor their investors to prevent a flurry of redemptions.

The remainder of this paper is organized as follows: we present our database in Section 1. Section 2 describes the modelling process. For ease of presentation, we choose one fund as a demonstrative example to build the model but present the analysis of the full sample at the end of the section. Section 3 compares different investors belonging to the same fund and Section 4 investigates funds with various asset classes. Finally, we discuss our results, and conclude.

²These previous studies explain the cross-section of fund flows by fund characteristics. Therefore we consider that these studies are more static than ours

2.1 The Data

Most asset management companies collect and store only a fraction of the data that could on the behaviour of their clients. However, the digital transformation of the asset management industry and the evolution of the asset management regulation force them to process more data. While these data can be very useful for the commercial development of funds, it should also be considered for optimal liquidity risk management purposes. In this section, we describe the different stages that led to the development of the database on the historical clients' behaviour.

2.1.1 Analysing investors' behaviour

Accurately identifying who the final investor, i.e. who makes subscription or redemption decisions, is always useful to any management company. Indeed, a good knowledge of liabilities, including its composition and the investors' profiles, allows more targeted commercial pushes and a more effective communication during crisis. However, getting this detailed information is not straightforward as the classical fund distribution model involves many intermediaries between management companies and final investors.

Despite the difficulty of tracking orders within a complex distribution network, the clients' data issue is becoming very popular in the asset management industry mainly for two reasons. The first reason comes from the regulatory constraints imposed to asset managers. The Autorité des Marchés Financiers (AMF) recalled in February 2017 that "the knowledge and analysis of liabilities is an essential component of the identification of risks by management companies". Future European regulations (MIFID II, Priips) also affect the distribution network as well as the producer-distributor relationship, and thus push for a better knowledge of funds liabilities. The MIFID II directive in particular could be an opportunity to set up a reporting of distributors to producers broken down by investor characteristics. The second reason of this increasing interest is related to the evolution of the asset management industry business model. The emergence of FinTechs and their direct distribution model upset traditional distribution channels and strongly compete with established asset managers. The direct access to end-clients allows them to gather more detailed information on their clients and to adapt their investment offer to their characteristics. Therefore, the aggregation of individual investments gives them a better view of the fund liabilities structure and allow them to better anticipate the fund flows during stressed periods. Any fund manager whose distribution channel complexifies the link between asset managers and clients have or will have a clear disadvantage.

Several avenues for effective monitoring of liabilities are already mentioned in the AMF guide³ published in June 2017. They are essentially qualitative, through a better understanding of the relationship between managers and investors. Nevertheless, a statistical analysis of subscriptions/redemptions conditional on clients' characteristics can improve this analysis. Moreover a quantitative approach can serve as a prerequisite for a more qualitative, targeted and ad hoc approaches. Developing a quantitative approach, however, requires access to resources, data and know-how not available to every management company. A quantitative approach can also be very useful at the industry level, but the redemptions/subscriptions time series must be large enough - both in the cross-section and in the time dimension, to be statistically representative. This can only be assessed by pooling a large dataset, involving considerable collection, anonymisation and standardisation data work we propose to do in the section.

2.1.2 Merging fund managers' database

The creation of a historical database for subscriptions/redemptions from information directly provided by management companies is a preliminary step to our research. To the best of our knowledge, there exists no database of individual investment decisions available to academic researchers. Indeed, as this information is highly strategic for management companies, they can be reluctant to share it and public information only concerns the past evolution of total flows at the fund level. While such aggregated information can be very useful for studying the relation between performance and fund size for example, any liquidity analysis requires a more precise view on funds liabilities. The related structure can vary over time according to the clients' inflows/outflows and only disaggregated data can give useful information on these potential trends.

Any statistical modelling of subscriptions/redemptions requires to work on a large and heterogeneous dataset. Asset management companies entrusted their clients full trading activity to us to this end. More precisely, data are provided by three medium size French asset management firms, referred to as Companies A, B and C in the following, and all affiliated to large bank or insurance groups. These firms sell funds mainly in France and Luxembourg. Fund shares are all denominated in Euro and the client-base covers both retail and institutional investors. They have indeed diversified investors and invest in all asset classes. Among all the fund categories they manage, we choose a representative sample of funds following two requirements. First, the final dataset must include funds which

³See, "Journal officiel de l'Union européenne: Règlement (UE) 2017/1131 DU PARLEMENT EUROPEEN ET DU CONSEIL, du 14 juin 2017 sur les fonds monétaires"

invest on different asset classes with different market liquidity level. Second, the sample must cover the principal funds' styles considered in the existing literature, such as Equity and Fixed Income funds. According to these two requirements, we decide to target four different categories⁴ which are: money market, large-cap equity, small and mid-cap equity and fixed income. These categories represents the most important part of the total asset under management, both at the industry level for the three funds companies considered in this study.

The raw database of investors' flows over the period from January 2010 to December 2014 was transferred from the asset management firms to us. For each fund, we hand-collect information on all investors purchases/sales trades, the associated customer identifier, the number of shares involved, the corresponding price and the date of the transaction. Since disaggregated fund flows are not mandatorily disclosed to regulators, there is no incentive to disclose such information making our database quite unique. Indeed, previous studies of fund flows use less precise data such as aggregate monthly net flows (see, e.g., [Chevalier and Ellison \(1999\)](#), [Ippolito \(1992\)](#))⁵, disaggregated monthly inflows/outflows (e.g., [Jank and Wedow \(2010\)](#), [Keswani and Stolin \(2008\)](#)), and recently daily or even ultra-frequency flows⁶ (e.g., [Johnson \(2007\)](#), [Schmidt et al. \(2016\)](#)). All mutual funds in our sample are open-end funds. Although there could be small deviations between funds, the basic rule is the same. Investors have the possibility to purchase or redeem fund shares whenever they want and the fund provides periodic - in most of case, daily - price or Net Asset Value (NAV) per share. All investors' transactions - purchases or redemptions - would then be executed at the NAV price. This contrasts to other investment vehicles such as Exchange Traded Funds (ETFs) which provide to investors an intra-day bid/offer price, or closed-end funds which do not provide redemption possibility before the final liquidation date.

The next step in building this database is to collect information on clients' characteristics. It is common knowledge that retail investors do not behave as institutional investors. The differences dictate their trade size and frequency, but also the type of funds they invest in as well as the triggers for inflows and outflows. Besides the willingness to share clients' data, a significant effort was made by asset management firms to create homogeneous groups of investors and to define a common nomenclature. Indeed, none of the asset management

⁴We use the Morningstar's "Global Board Category Group" classification as fund category. This classification is based on asset types and geographical zones, as "Europe Large Cap Equity" or "US Money Market". However, it is not based on managers' style, as "Growth", "Value" or "Market".

⁵The fund flow was not available in the public database. Thus one can extract from the AUM and fund return.

⁶[Engle \(2000\)](#) refer the ultra-frequency the datas which record information of every transaction, in an order book or in other fields.

companies is using the same client typology so we had to create a new unique classification and apply it to all investors. We end up with 8 types of investors⁷, ranging from institutional investors through private banks or independent wealth management advisors to retail investors. The granularity level of our database has several advantages. First, this classification allowed us to draw conclusions about the average behaviour of each category of investors by observing the behaviour of a large number of investors in each category. For example, we are able to compare the investment horizon of institutional in small and large cap stocks to that of public clients by calculating average holding times for each type of investors. Second, the disaggregated approach allows us to calculate separately the amount of subscriptions and redemptions by type of clients, enabling management companies to identify trends in their liability structure. Let us consider the case in which a significant proportion of institutional clients are replaced by retail clients leaving the total asset under management identical. The modification in the fund's liability structure is silent on aggregated data and only detectable with individual flows (see, [Christoffersen and Xu 2017](#) for a study of the impacts of the liability structure on fund flows.). Third, the statistical treatment of disaggregated data also makes it possible to follow the history of a given investor within the same management company. We are thus able to detect the arbitrages between asset classes or types of funds made by a given investor or a group of investors. For example, we can analyse if some investors reduce their exposure to a given asset class to invest in money markets. This would not be possible with aggregated data. Finally, our disaggregated data can also help addressing issues related to contagion. Let us consider two clients - an institutional and a retail investors, within a given fund. The most sophisticated investor can quickly react to a deterioration of the financial environment and thus, reduces its risk exposure. We can measure the delay for each category of investors as well as the contagion to other investors, if any, and see if it can help predicting future outflows. These analyses can be performed for any fund or at the industry level.

In a nutshell, our dataset reaches the most disaggregated level of flow data recorded on a daily basis. All transactions are recorded with all necessary basic information: 1: Date; 2: Fund liquidation price (NAV); 3: Anonymous client number; 4: Client's sector; 5: Amount of trades. This dataset reaches our main objective, that is to work on a panel of funds with various exposure to liquidity risk. Money market funds, for example, are invested in liquid assets, with little exposure to liquidity risk. Moreover, if large-cap equity funds have also little exposure to market liquidity risk, it increases for small and mid-cap equity funds.

⁷The number of sectors in this chapter is different from the Chapter 1 (15 sectors). Because some sectors possess only a small number of flows, we merge them into 8 for the simplicity of the model estimation.

At the less liquid end of the spectrum, fixed income funds are much more vulnerable to deteriorating liquidity because they invest in illiquid assets. On the liability side however, all the selected funds offer daily liquidity to investors. It is the liquidity mismatch - the market liquidity of assets relative to the structure of the liabilities - that matters. As a consequence, it is important to better understand and measure the real consumption of liquidity on the liability side. It is then possible to assess the risks related to the open nature of investment funds, and the risks related to the presence of the same investors in different funds. This can create contagion channels on the fund liability side, and only the disaggregated approach addresses these issues.

2.1.3 Final database characteristics

We choose 12 funds within four management styles from the very large universe of funds provided by the three management companies A, B and C. Our sample contains almost one million trades - 930 000 transactions (578 000 purchases and 357 000 sales). An extension to all funds in the full sample would increase the size of the sample to several billion transactions. We present basic information in Table 2.1. Each fund is assigned to a Morningstar global category. Selected funds have several fund share classes to ensure a minimum level of investors' heterogeneity. Moreover, all these funds have been launched more than 10 years ago implying that the structure of the funds is already quite stable. The total asset under management on all 12 selected funds exceed 10 billion euros, with the presence of some large - money market - funds.

The largest fund in this category manages more than 3.4 billion USD, while the smallest fund is in the mid and small cap Equity category with 34 million USD under management.

We present the summary of these flow data in Table 2.2. The trading activity is heterogeneous across funds, with a number of transactions ranging from 522 (Funds 9) to 197 037 (Funds 1). There are also large discrepancies across funds' types as well as across management companies. Most of the funds have more than one trade per day on average, which differs from previous studies using similar type of data (e.g., [Ivkovich and Weisbenner 2008](#); [Johnson 2004](#)). Indeed, they have in general less than 1 trade per day and in and outflows happen rarely together. The difference of investors' behavior between the United States and Europe can explain this empirical fact. In the United States, most individual investors tend to make their own investment decision while preparing their retirement portfolio (401k plan) (see, e.g., [Paula and Croce \(2016\)](#)). On the contrary, the vast majority of European individual investors delegate the decision making process to some financial intermediaries

such as insurance company, fund of funds managers or pension funds. Consequently, funds' investors are financial intermediaries investing for their clients and fund flows are more aggregated in Europe.

Table 2.1: The fund sample (12 selected funds)

Fund	Company	Category	N shares	Inception Date	Fund Size
Fund 1	A	Equity Large Cap	3	1998/10/02	329 723 439
Fund 2	A	Equity Mid/Small Cap	2	1991/9/6	376 326 122
Fund 3	A	Money Market	2	1985/12/31	450 074 000
Fund 4	A	Fixed Income	3	1990/02/05	935 044 376
Fund 5	B	Equity Large Cap	5	2006/4/25	280 424 002
Fund 6	B	Equity Mid/Small Cap	4	1994/5/11	333 368 999
Fund 7	B	Money Market	4	2006/4/25	3 415 839 000
Fund 8	B	Fixed Income	5	2006/3/8	354 900 000
Fund 9	C	Equity Large Cap	3	2001/01/09	295 271 161
Fund 10	C	Equity Mid/Small Cap	2	1997/2/14	34 287 000
Fund 11	C	Money Market	2	2013/1/4	1 319 876 994
Fund 12	C	Fixed Income	1	2001/11/30	388 074 000
Total	3	4	43		10 714 336 092

This table shows the elementary information about our fund sample. It contains 12 open-end mutual funds from three fund families. We classify funds according to the Morningstar's "Global-Board Category". "N Shares" is the number of the share-classes that each fund have. Clients in the different share-classes (within the same fund) pay different management fees. "Inception Date" is the creation date of the oldest share class. "Fund Size" is dated at 16/12/2015.

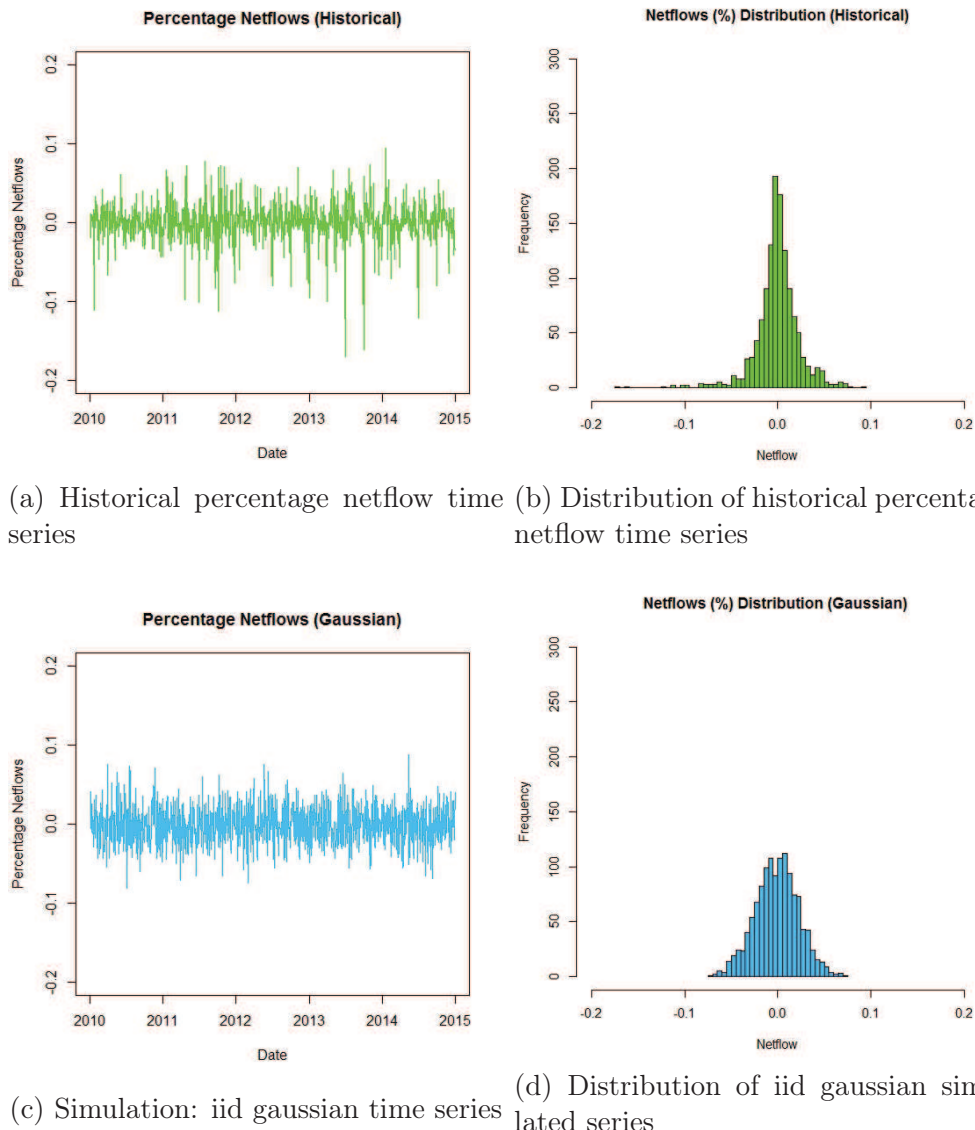
For the simplicity of presentation, we first take Fund 7 as an illustrative example before presenting the results for all other Funds later on. Fund 7 is a money market fund managed by Company B. As other funds in the same category, investors are very active and a significant number of flows occurs every day. First, as the fund manager only observes the net flow time series, we focus on the relative netflows taking the difference in value between in and outflows and dividing by the fund' size (see Figure 2.1). The next step is to see if we can gain from separating in and outflows as in Figure 2.2.

Table 2.2: Summary statistics for fund flows

Fund	Category	Period	N Days	Number			Freq		
				<i>Inflows</i>	<i>Outflows</i>	<i>All</i>	<i>Inflows</i>	<i>Outflows</i>	<i>All</i>
Company A				41 358	59 536	100 894			
Funds 1	EquityLargeCap	2013-14	497	521	1 097	1 618	1.04	2.21	3.25
Funds 2	EquityMidSmallCap	2013-14	497	761	863	1624	1.53	1.74	3.27
Funds 3	MoneyMarket	2013-14	497	36 779	54 044	90 823	74.00	108.74	182.74
Funds 4	FixedIncome	2013-14	497	3 297	3 532	6 811	6.60	7.11	13.70
Company B				37 186	41 872	79 058			
Funds 5	EquityLargeCap	2010-14	1252	2 581	2 192	4 773	2.06	1.75	3.81
Funds 6	EquityMidSmallCap	2010-14	1252	2 481	1 610	4 091	4.94	3.20	8.15
Funds 7	MoneyMarket	2010-14	1252	31 465	36 720	68 185	25.13	29.33	54.46
Funds 8	FixedIncome	2010-14	1252	659	691	1 350	0.53	0.55	1.07
Company C				3 392	4 732	8 124			
Funds 9	EquityLargeCap	2010-14	1 249	1 468	1 400	2 868	1.18	1.12	2.3
Funds 10	EquityMidSmallCap	2010-14	1 233	1 151	2 459	3 970	1.35	2.20	3.56
Funds 11	MoneyMarket	2010-14	1 115	210	312	522	0.43	0.63	1.06
Funds 12	FixedIncome	2013-14	493	563	561	1 124	0.46	0.45	0.91
Total				81 936	106 140	188 076			

Table 2.2 gives the summary statistics on fund flow data. We follow Morningstar's "Global-Board Category" to classify funds into four groups. The column "Days" is number of days when the fund is open to trade. "Number" counts the number of flows. We present inflows and outflows separately and then we give their sum ("All"). "Freq" is the flow frequency which equals to flow count divided by the number of days.

Figure 2.1: Inaccurate fits of traditional iid model

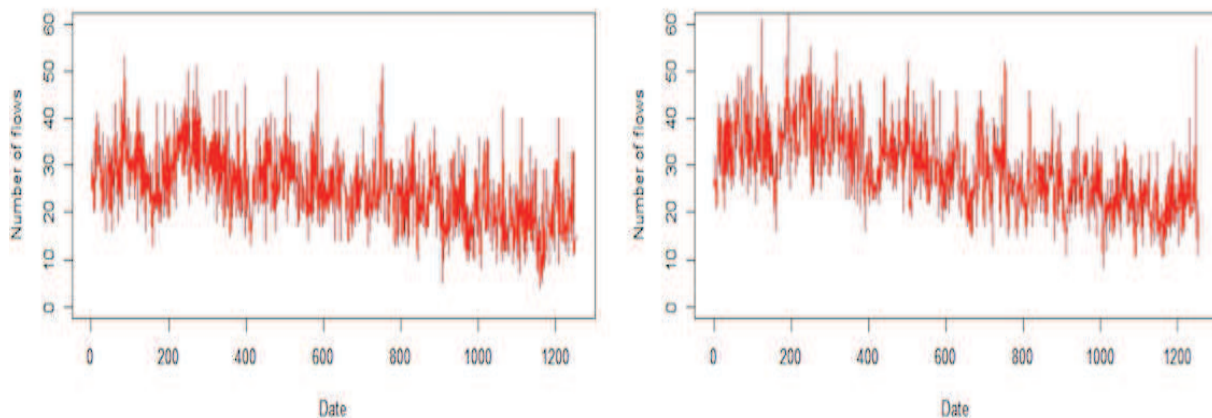


(a) Historical percentage netflow time series (b) Distribution of historical percentage netflow time series

(c) Simulation: iid gaussian time series (d) Distribution of iid gaussian simulated series

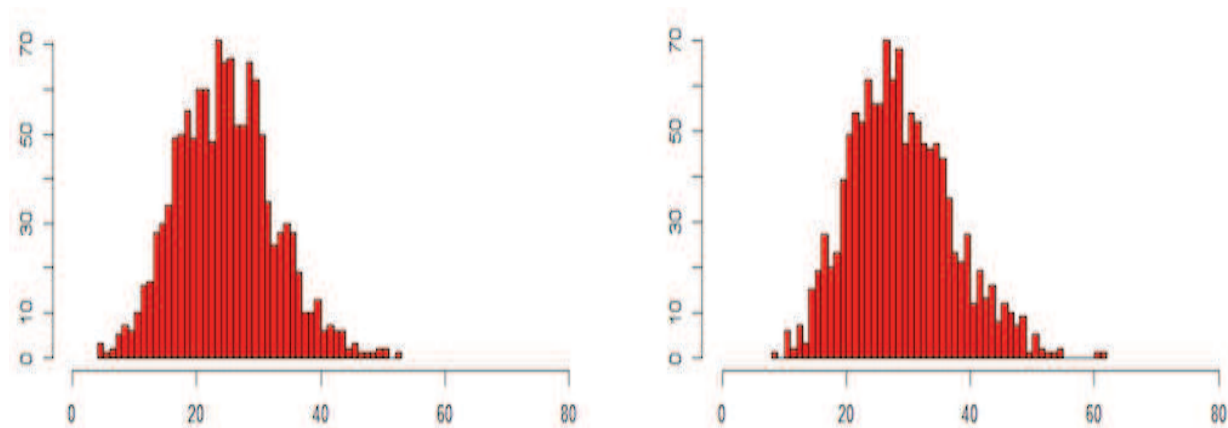
Figure 2.1 shows the comparison between the historical data series and a Toy model. We first convert the net-flow series (inflows minus outflows) into percentage value by dividing them by fund's size. We present net-flows' time series for a 5 year-period, from 2010 to 2014, in Sub-Figure (a). Then we give their distribution in Sub-Figure (b). We use their average and variance to generate a Gaussian distribution. We use this distribution to simulate time series with a same length (in Sub-Figure (c)). Lastly, we give the distribution of simulated series in Sub-Figure (d).

Figure 2.2: Fund flow times series



This figure presents flow time series of a money market mutual fund. We present inflow (left) and outflow (right) separately. The data covers a 5 year-period, from 2010 to 2014.

Figure 2.3: Distribution of fund flow numbers



In this figure we present the distribution of fund flow number counts. We present inflow (left) and outflow (right) separately. The data covers a 5 year-period, from 2010 to 2014. The X-axis shows the flow number and the Y-axis shows the frequency (number) of observation.

There are several evidences that the series in Figure 2.1 do not follow a gaussian distribution: the distribution is not symmetric and has fat tails. Moreover, the percentage of netflows appears to be clustered. These features are even more pronounced on the disaggregated series. Besides, the evolutions as well as the distributions are clearly not the same for the two series.

Table 2.3: Summary statistics on a demonstrative example:

Fund	Category	Type	N	Mean	Variance	Q1	Q2	Q3	Q4	Q5
Fund 6	Money Market	Inflow	1252	25.13	59.80	4	19	25	30	53
		Outflow	1252	29.33	67.65	8	23	29	35	62

Table 2.3 presents the summary statistics on Fund 6's flows. We present inflows and outflows separately. The sample covers a 5 year-period (1 252 days), from 2010 to 2014. We calculate sample's average, variance and quantiles. "Q1" to "Q5" represent the first to the 5th quantile.

The distribution summary statistics given in Table 3 confirm that is the distributions of inflows and outflows are different from one another, they both are asymmetric with fat-tails. Moreover, with the average being more than twice as big as the variance, by no mean these two distributions can be considered as Poisson distributions.

Our objective is to propose an accurate model for in and outflows counts that reflects the empirical properties of the data. For that reason, we look for bi-variate discrete model of counts that can account for fat-tails and clustering.

2.2 Modelling investors' behaviour

We present the model specification in this section. Our goal is to propose a simple specification compatible with the different stylized facts observed in practice. The objective is to develop a standard method, and to keep the possibility to fund managers of using more sophisticated internal models. The analogy can be made with the calculation of the value at risk (VaR) of a portfolio. We can see the different stages of modelling such as the calculation of a Gaussian VaR, then a GARCH VaR, etc. This gradual approach allows to evaluate at each step the contribution of the new parameters to be included in the model. We apply this approach to the estimation of the number of subscriptions and redemptions observed for a given fund, regardless of the type of client.

2.2.1 Subscription/redemption counts estimation

This section applies an over-dispersed, autocorrelated, bi-variate Poisson model to describe the number of subscriptions and redemptions observed per day. The simplest statistical counting model is the Poisson model. It assumes that the number of subscriptions or redemptions observed each day, denoted by N_t , is the realization of an independent random variable identically distributed according to a Poisson law of parameter. The Poisson model captures the fact that the variance increases with the mean and in fact they are both equal. Moreover, the average number of event per period of time is constant. In our specific case of subscriptions and redemptions, these two conditions seem unreasonable. In fact, empirically we observe that the variance of the number of subscriptions and redemptions are much greater than their averages. We will stress this point in the first subsection analysing this problem of over-dispersion. Moreover, during turmoil periods, the longer the time without redemption the greater the chance that a fund will experience a redemption. Also, whether a fund is experiencing outflows on any given day is independent of what happens in other funds, contradicting a common belief that outflows tend to cluster. Finally, we can not only past redemptions (subscriptions) matter to explain future redemptions (subscriptions), but also past subscriptions (redemptions). We will tackle the problem of autocorrelation in subsection 2.2.3 and the cross effects in subsection 2.2.4.

2.2.2 Capturing over-dispersion

One of the stylized facts observed on the transaction data goes against the theoretical properties of the Poisson's law - in particular, the variance is equal to the mean. Indeed, it

is observed empirically that the variance of the number of subscriptions or redemptions are greater than its average. Using a simple Poisson law would therefore have the disadvantage of misaligning the variability of the series of interest and underestimating for example the probability of liquidity stress scenarios. It is therefore essential to take this into account by working with models to reproduce the empirical characteristics of the series. For this, we use a "generalized" Poisson model including the over-dispersion property: the sample variance is larger than its mean.

We adapt the over-dispersion by adding an additional parameter s into a traditional Poisson distribution to generate different levels of variance. The flow number N_t is then generated by a "Double Poisson" distribution with two parameters: the intensity coefficient λ represents the average number of flow arrivals; s is the parameter which generates different variance levels. Under this distribution, the flow variance is $V(N_t) = \frac{\lambda}{s}$. If s is smaller than 1, $V(N_t)$ will be larger than λ , hence the over-dispersion appears. On the contrary, if s is larger than 1, the flow variance is smaller than the flow average, thus an under-dispersion presents. When s equals to 1, the distribution reduces to a homogeneous Poisson distribution⁸.

The estimation of the two parameters λ and s for the different bottoms of the sample shows that the over-dispersion is effective, with parameters s different than 1. The estimated values of this parameter, and therefore the levels of over-dispersion, are higher for redemptions than for subscriptions.

This representation has the advantage of being very simple to estimate, but it also has the disadvantage of being purely static, i.e. the law of observation at a date t does not depend on the observations on the previous dates. This has a troublesome direct implication. The best estimate of the number of subscriptions or redemptions on the following date is equal to the average subscription or redemptions, which is constant over time. A study of the corresponding time series shows that there are phenomena of concentration of subscriptions and/or redemptions over given periods. Similarly, to volatility, a period of high redemptions appears to increase the likelihood of future large redemptions. The persistence phenomena we observe in subscriptions and redemptions cannot be captured using a static representation and we have to include a dynamic component.

2.2.3 Capturing autocorrelation

In line with the ARCH representation, where the volatility at a given date depends on the square of the returns observed on the previous dates, we propose to replace the constant

⁸See, [Efron 1986](#) for details of the Double Poisson distribution. There are other choices to generate the over-dispersion property. See, [Zhu 2012](#) for other alternative choices.

parameter of the Poisson law λ by a parameter λ_t which varies in time as a function of the previous observation of the time series. The specification chosen is the following:

$$\lambda_t = \lambda_0 + \rho N_{t-1}, \quad (2.1)$$

where λ_0 is a constant intensity and the additional ρ parameter captures the temporal persistence in subscriptions or redemptions. When $\rho > 0$, we observe that an increase in the number of transactions on the past date will have a positive impact on the intensity λ_t , and therefore increase the average number of transactions on the following date. This channel creates both persistence and clusters in the time series of transactions, just like volatility in the ARCH representation. The inclusion of additional delays in the specification above is straightforward. In this study, we limit ourselves to the inclusion of a single delay in order to keep the model as simple and as parsimonious as possible. In the case where $\rho = 0$, we are back to the simple Poisson case where the past has no influence on the number of transactions at the current date. Therefore, the test for persistence in the series observed corresponds to the statistical significance of ρ . Our estimation results show that for the vast majority of funds, the persistence parameters ρ are significantly different from zero, and that the most important persistence levels are observed for redemptions. The risk of observing order concentrations over short periods of time is therefore more important for redemptions than for subscriptions.

From a practical point of view, the interest in dynamic models is to provide non-constant subscription or redemption forecasts. As soon as the parameter ρ is significant, the forecast of future redemptions depends on current redemptions, and that of future subscriptions for current subscriptions. This forecast can of course be used by the manager of a fund which has the possibility to anticipate what will be the amount of the repayments on the following date. It is therefore possible for him to start adjusting the size of his portfolio so that he can easily cope with his clients' repayment orders.

2.2.4 General bivariate specification

As mentioned above, redemptions and subscriptions are not behaving exactly the same way and as a consequence, should be modelled separately. Moreover, here we want to see if there are any cross effects between the two characteristics. With this in mind, we propose a bivariate approach including additional parameters. We consider N_t^{in} the number of subscriptions in a fund at the date t , and N_t^{out} the number of redemptions for the same fund at the date t . We assume that the intensity of the Poisson laws describing the subscriptions

at the date t , λ_t^{in} and the one describing the redemptions λ_t^{out} satisfy the following two equations:

$$\lambda_t^{in} = \lambda_0^{in} + \rho^{in-in} N_{t-1}^{in} + \rho^{in-out} N_{t-1}^{out}, \quad (2.2)$$

$$\lambda_t^{out} = \lambda_0^{out} + \rho^{out-in} N_{t-1}^{in} + \rho^{out-out} N_{t-1}^{out}, \quad (2.3)$$

where the two parameters ρ^{in-out} and ρ^{out-in} capture the dependencies between subscriptions (or redemptions) on the date t and redemptions (or subscriptions) on the previous date. All other parameters interpretations remain the same as in the previous representation.

It is now possible to discuss the financial interpretation of the four parameters ρ included in the most general version of the model. ρ^{in-in} can be interpreted in terms of reputation. Past subscriptions increase the average number of subscriptions on average. $\rho^{out-out}$ captures the panic effects. Investors, seeing a number of significant outflows coming true interpret this as a negative signal and also tend to exit the fund. ρ^{in-out} measures the manager's ability to stabilize the size of his fund, for example by triggering commercial actions to offset past outflows by a larger number of subscriptions. Finally, ρ^{out-in} captures the behaviour of investors who leave the fund following the massive influx of other investors. This can be seen as a contrarian behaviour of some investors who leave the fund when "too good" performances attract too many new investors. They anticipate capacity issues and a deterioration of future performance as the size of the fund increases.

In terms of liquidity risk management, the presence of positive cross-effects is rather beneficial since it tends to stabilize the fund's assets under management. The most critical case is a value $\rho^{in-out} < 0$, which means fewer subscriptions on average when redemptions increase. Thanks to this representation, the negative effects of past exits can be split in two. It can indeed increase future outflows or decrease future inflows. The empirical analysis will show which of the two effects is the strongest.

Again, if all the ρ parameters are zero and $\lambda_0^{in} = \lambda_0^{out}$, we get back to the simple (univariate) Poisson distribution. When ρ^{in-out} and ρ^{out-in} are null, we have the over-dispersed persistent Poisson representation with no cross effect.

In a matrix format, we have :

$$\begin{pmatrix} \lambda_t^{in} \\ \lambda_t^{out} \end{pmatrix} = \begin{pmatrix} \lambda_0^{in} \\ \lambda_0^{out} \end{pmatrix} + \begin{pmatrix} \rho^{in-in} & \rho^{in-out} \\ \rho^{out-in} & \rho^{out-out} \end{pmatrix} \times \begin{pmatrix} N_{t-1}^{in} \\ N_{t-1}^{out} \end{pmatrix} \quad (2.4)$$

2.3 Empirical results

We first apply the bivariate model described in the previous section to aggregated inflows/outflows and four different fund categories. The objective of Subsection 2.3.1 is to check if the fund category or if the management company can have an impact on investors behaviour. Next we concentrate on money market funds in Subsection 2.3.2 and then on disaggregated flows in Subsection 2.3.3. The objective there is to exploit information from disaggregated data to check whether different investors types result into different investors behaviours.

2.3.1 Are all liquid funds equally risky?

There is a clear possibility that fund flows' characteristics are related to the asset classes invested by the fund manager. We intend to test for this effect by comparing the value taken by the parameters of the model described in Equations (2.2)-(2.3) when applied to funds belonging to different categories. We then estimate and compare estimation results for funds within and between categories. Table 2.4 presents the estimation results on aggregated inflows/outflows for 12 funds spread out among 4 categories and 3 asset management companies. The objective of the exercise is to compare the differences in the clients trading behaviour - inflows and outflows separately - depending on the associated management company, while all funds offer the same daily liquidity.

Table 2.4: Count model estimation for 12 selected funds

Fund & category	Company	λ_0^{in}	λ_0^{out}	ρ^{in-in}	ρ^{in-out}	ρ^{out-in}	$\rho^{out-out}$	S^{in}	S^{out}
n1 EquityLargeCap	A	0.52***	1.87***	0.47***	0.00	0.00	0.15*	0.76***	0.84***
n2 EquityMidSmallCap	A	0.99***	1.59***	0.34***	0.00	0.00	0.08*	0.87***	0.86***
n3 MoneyMarket	A	17.25***	23.06***	0.26***	0.34***	0.20***	0.65***	0.15***	0.14***
n4 FixedIncome	A	4.62***	5.40***	0.30*	0.00	0.00	0.25***	0.77***	0.65***
n5 EquityLargeCap	B	1.94***	1.55***	0.00	0.06*	0.01	0.10***	0.67***	0.73***
n6 EquityMidSmallCap	B	2.31***	1.59***	0.53***	0.00	0.03	0.45***	0.75***	0.75***
n7 MoneyMarket	B	9.20***	10.49***	0.30***	0.28***	0.20***	0.47***	0.56***	0.65***
n8 FixedIncome	B	0.20***	0.29***	0.18***	0.00	0.00	0.08*	0.42***	0.44***
n9 EquityLargeCap	C	1.04***	0.97***	0.01	0.00	0.00	0.15***	3.05***	2.82***
n10 EquityMidSmallCap	C	0.37***	1.56***	0.56***	0.05**	0.17***	0.18***	0.56***	0.70***
n11 MoneyMarket	C	0.29***	0.37***	0.15**	0.08*	0.09	0.35***	0.39***	0.31***
n12 FixedIncome	C	0.40***	0.43***	0.10*	0.00	0.00	0.02	0.12***	0.25***

This table presents the estimation the Self-Exciting Fund Flow model for a diversified fund sample, which contains 12 open-end, non-load mutual funds from three different fund families. These funds cover 4 categories: Equity Large Cap, Equity Mid/Small Cap, Fixed Income and Money Market. This classification is created by Morningstar Data Service. It indicates the asset that the fund manages. The model is a time series Poisson count model, which contains 3 elements: the baseline intensity, the clustering effect and the dispersion. We present each estimated coefficient with its statistical significance.

We present in Table 2.4 the estimation results. We display the fund's category and the company name in the first two columns followed by the eight parameters estimators in the following columns. We observe a huge variety on flow characteristics. Indeed, the average level of flows are very different from one category to another, but also within each category. The less actively traded funds (Funds 11) has a baseline intensity of inflows of only 0.29 (resp. 0.37 for outflows). On the contrary, the most actively traded fund (Funds 3) has a baseline intensity of inflows of 17.25 (resp. 23.06 for outflows). Moreover, we observe that the fund company has a substantial impact on the average flow level. Company A has the largest flows and Company C the smallest ones. This strong "company effect" is linked to the fact that the company scale determines the fund size and its client-base. A large and famous fund company receives more flows than any unknown small one. Furthermore, the fund category also influences the flow level. Liquid categories such as Money Market and Equity Large Cap funds receive on average more flows than illiquid funds belonging to Equity Small Cap and Fixed Income categories, for example. The result tends to prove that clients use liquid funds to manage their liquidity needs. As a consequence, they are more active on "liquid" funds when cash is needed.

We now focus on the dispersion component captured via parameter S , estimated on both inflow and outflow series. The estimation of our model confirms that, as we observe in Section 2.1.3, the majority of funds suffers from over-dispersion with a parameter S being lower than 1. Indeed, S is larger than 1 for Funds 9 only meaning that this funds is the only one to exhibit under-dispersed flows. This supports our choice of a model which can cover both over- and under- dispersion of count series. Moreover, there is no clear monotone relation between the management company and the dispersion level, or between the asset class invested by the funds and the dispersion level. We can conclude at this stage that over-dispersion is not company nor style-dependant. Each funds seems to exhibit an idiosyncratic dispersion behaviour that depends on the clientele only. Indeed, funds in the same class might have diverse dispersion levels. For instance, Funds 9 and 5 are both large-cap equity funds with similar baseline intensities but while Funds 9 exhibits under-dispersion, Funds 5 flows are over-dispersed.

Turning now to dynamic effects, we can see that the autocorrelation of flows vary also quite a lot. On the one hand, the highest ρ^{in-in} reaches 0.56 (Funds 10) and the highest $\rho^{out-out}$ coefficient is 0.65 (Funds 3). These results are directly related to the clustering behaviour in funds trading observed in Section 2.1.3. On the other hand, some funds exhibit no significant autocorrelation at level 5% in their flows, such as Funds 12 for example. Let us now focus on the $\rho^{out-out}$ estimators for the 12 funds. We observe that this parameter

is significant for the three money market funds in our sub-sample, while it becomes not significant at 5% for two of the three Fixed Income funds. Therefore, it appears that clients take into account the difficulty that the manager may encounter in managing the offered liquidity and adapt their behaviour accordingly. Fixed Income fund clients are usually less sensitive to outflows of other investors than money market fund clients. The large value obtained from money market funds can also be interpreted as seasonality and show that clients are releasing funds at some regular time periods. The first conclusions of this empirical study are that not only the clients of the funds integrate well the liquidity dimension in their investment policy but also that the daily liquidity offered by the fund is not used in the same way according to the type of the considered funds. Although fund company and fund category influence strongly the fund flows, we find that this influence is not so clear concerning the autocorrelation of the flow level and dispersion. We find clear evidence that funds in the same category have different correlation parameters. Funds 2 and 6 are both mid/small-cap equity funds with similar baseline intensities. However, their ρ^{in-in} parameters are very different and if Funds 6 $\rho^{out-out}$ is 0.45, this parameter is not significant for Funds 2. We find a similar situation for Funds 8 and 12. They are both fixed income funds with the similar λ_0 . However, Funds 12's $\rho^{out-out}$ is zero while Funds 8 has a significant clustering effect in outflows. If we now look at the cross effects between subscriptions and redemptions, we observe that these effects are significantly different from zero and positive for two of the three money market funds and not significant for all the other funds except for Funds 10. They therefore play a stabilizing role and in particular offset past outflows by a larger number of new entries. This stabilizing mechanism does not hold for funds highly exposed to liquidity risk.

To further confirm these results on "liability risk", we estimate the parameters in a larger sample. We consider all large-cap equity and fixed incomes funds data of Company B in 2014, ending up with 51 equity funds and 23 fixed income funds in total. We estimate Equations (2.2)-(2.3) and present the distribution of $\rho^{out-out}$ and S^{out} in Figure 2.4. We clearly observe that funds in the same asset class, and so with the same style, present a wide variety of liability risks. In particular, if the correlation risk in outflows is on average quite low for Equity Funds, it still can be large for some particular funds. The same apply for S^{out} and we see Sub-Figure (a) that S^{out} is mostly low but can even reach 0.9. On the contrary, Fixed income funds present low to high clustering effects as we see in Sub-Figure (c), where the outflow autocorrelation is zero for some funds.

In summary, we have analysed several economic factors which might impact the flow risk. Especially, the fund company and fund category have strong determinant power on

fund flows on average. However, we find that this influence is limited on flows dispersion and autocorrelation where the parameters differ for funds within the same category or managed by the same company. Accordingly, we argue that these two risks are more generated by the liability-side, i.e. by investors' behaviour so that the fund managers should make efforts to monitor this liquidity risk component.

Figure 2.4: Distribution of Model's Parameters: Equity and Fixed Income Funds (an extended sample)

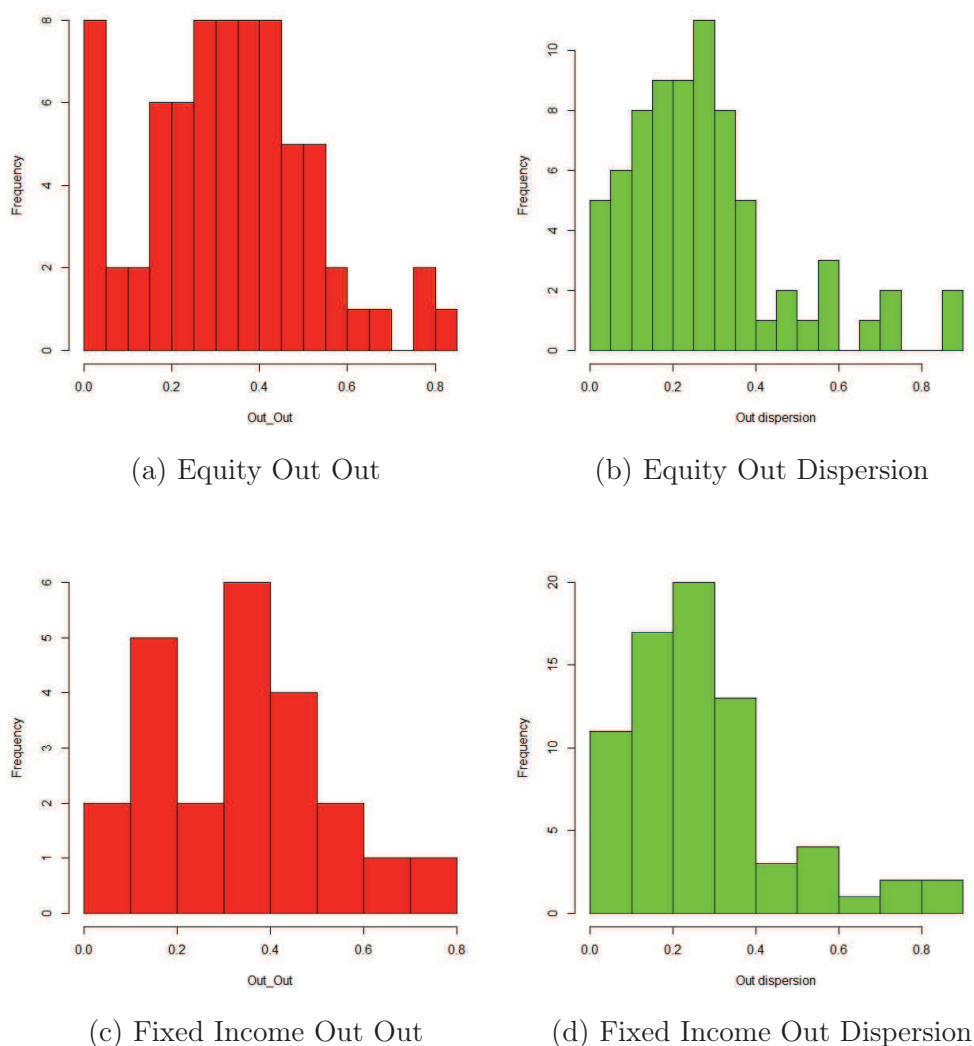


Figure 2.4 presents the distribution of model's parameters in a multiple fund's sample. It contains 51 equity funds and 23 fixed income funds. All of them are of the fund company B. We present the distribution of two risk parameters: *Out - Out* (red at left) and *Out - Dispersion* (green at right). We show the equity funds on the top and the fixed income funds on the bottom.

2.3.2 Are money market fund predictable?

We present in this Subsection some results relative to a given money market fund considering different characteristics reflected in different model specifications. Our benchmark model is the full model of the previous section and the competing models are some constraint specifications of that model. The idea is to measure the importance of each individual feature encompassed in the full model. This model corresponds to Eq. (2.4) page 57 that we rename (2D3) for the rest of the paper.

The first competing specification considers over-dispersion and autocorrelation (dynamic direct effect) but no cross effects (dynamic cross effects) and we have ρ^{in-out} and ρ^{out-in} that are both zero in our general bivariate model Eq. (2.2)-(2.3). In that case which corresponds to Eq. (2D2) below, we can estimate each equation separately, i.e. Eq. (2.2) on inflows and Eq. (3) on outflows. The second competitor assumes over-dispersion but no dynamic effects, direct or indirect. As a consequence, in this model all the ρ -parameters are zero. Here again, we do have two equations: one for the inflows and one for the outflows. The last competitor assumes that in- and outflows should not be modelled separately and we merge the two series into a single netflows series. This final specification resumes to a unique Poisson model with overdispersion as in Eq. (1D). Finally, we get the following equations for the λ function.

$$\begin{pmatrix} \lambda_t^{in} \\ \lambda_t^{out} \end{pmatrix} = \begin{pmatrix} \lambda_0^{in} \\ \lambda_0^{out} \end{pmatrix} + \begin{pmatrix} \rho^{in-in} & \rho^{in-out} \\ \rho^{out-in} & \rho^{out-out} \end{pmatrix} \times \begin{pmatrix} N_{t-1}^{in} \\ N_{t-1}^{out} \end{pmatrix} \quad (2D3)$$

$$\begin{pmatrix} \lambda_t^{in} \\ \lambda_t^{out} \end{pmatrix} = \begin{pmatrix} \lambda_0^{in} \\ \lambda_0^{out} \end{pmatrix} + \begin{pmatrix} \rho^{in-in} & 0 \\ 0 & \rho^{out-out} \end{pmatrix} \times \begin{pmatrix} N_{t-1}^{in} \\ N_{t-1}^{out} \end{pmatrix} \quad (2D2)$$

$$\begin{pmatrix} \lambda_t^{in} \\ \lambda_t^{out} \end{pmatrix} = \begin{pmatrix} \lambda_0^{in} \\ \lambda_0^{out} \end{pmatrix} \quad (2D1)$$

$$\lambda_t = \lambda_0 \quad (1D)$$

In order to measure the quality of the model, we compute the mean square error (MSE) between the predicted flows and the observed ones using the three above specifications for the intensity function. The higher this criteria, the poorer the quality of the prediction.

Table 2.5 [Resp. Table 2.6] reports the parameters' estimators corresponding to the specifications described above for the Money Market Fund 7. We choose this particular

fund because it shows a complex structure with both a high level of autocorrelation and a large dispersion of flows. In this case, we can expect to measure more easily the impact of misspecification.

Table 2.5: Models' estimation (for the selected money market fund)

Specification	λ_0^{in}	λ_0^{out}	ρ^{in-in}	ρ^{in-out}	ρ^{out-in}	$\rho^{out-out}$	S^{in}	S^{out}
Step 1: (1D)	54.46***	same					0.44***	same
Step 2: (2D.1)	25.13***	29.32***					0.46***	0.51***
Step 3: (2D.2)	13.42***	12.75***	0.46***			0.56***	0.55***	0.65***
Step 4: (2D.3)	9.20***	10.49***	0.30***	0.28***	0.20***	0.47***	0.56***	0.65***

Table 2.5 presents the estimated coefficients of all models in Section 2. Each step represents an adaptation that we add to the aggregated homogeneous Poisson model. We note the significance level at the right-top of each coefficient (*p<0.1; **p<0.05; ***p<0.01).

Table 2.6: Improvements in model's quality (for the selected money market fund)

Specification	N	Adaptation	MSE-In	MSE-Out	MSE-All	%	AIC
Step 1: (1D)	2	"Over-Dispersion"			239903.1		
Step 2: (2D.1)	4	Separate Dynamic	74821.25	84638.42	159459.7	34%	-315870.4
Step 3: (2D.2)	6	"Self-Exciting"	60588.69	59129.57	119718.3	25%	-316328.2
Step 4: (2D.3)	8	"Crossed-Effects"	55940.79	57183.28	113124.1	6%	-316478.2

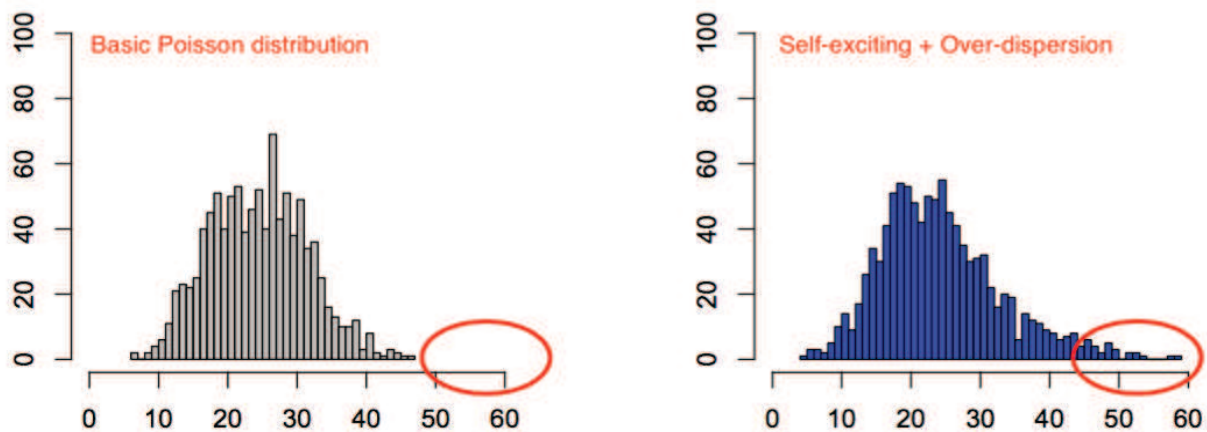
This table shows the summary of models' quality. We count the number of parameters (N) in each model (Specification) and calculate the Mean Square Errors for inflows, outflows. Then we sum two MSE up to obtain a quality ratio for the full model (MSE-All = MSE-in + MSE-out). Lastly we compute the AIC ratio for the model. We do not present AIC ratio for the uni-variant model since it is not comparable to other bi-variant models. "N" is the number of parameters in each step model. "Adaptation" is the stylized fact that we include in each step. "%" is the errors that each step decreases in percentage of the previous step's MSE-ALL.

When we first constrain the cross-effect parameters to be null, the MSE increases by 6%, moving from 113 124 to 119 718. Even if this increase is rather small, this misspecification is not neutral on the other dynamic parameters which increase too. This loss of quality in the prediction is even more important if we constraint the two remaining dynamic parameters to be zero. Indeed, the MSE increases too, but this time it goes up to 159 489, meaning that the increase is 33% compared to the model with no cross-effects and goes up to 41% if we compare to the general model. These results show the importance of including all these effects in the specification and particularly, the autocorrelation effects. Finally, we compare these results to that of the Poisson model with overdispersion applied to the aggregated flows. The MSE is moving up once again but to reach 239 903 the increase is of 34% compare to the model with autocorrelation effects and of even 112% compare with the full model. These results clearly show that not only in and outflows should be modelled separately, but

also that some dynamics features should be incorporated into the specification, and more particularly the direct ones.

To continue the comparison of the statistical properties of the constrained and unconstrained models, we use the specifications given in Eq. ((1D)) - the Poisson process on netflows - and in Eq. (2.4) - the full model, to generate two simulated count data. We use the two smoothed empirical distributions that we get out of the two simulated samples to observe and compare the marginal distribution generated by the two dynamic specifications. The results are displayed in Figure 2.5. We observe on the right of the plot that the implied marginal distribution has a fatter right hand tail for the dynamic model, meaning that the probability to observe extreme flows is higher. In this example, the worst situation with the static Poisson model model is 47 outflows while it jumps to 60 when we take into account the autocorrelation of outflows. We conclude from that observation that the static model is underestimating by 6.5% the potential outflows.

Figure 2.5: The Liability Risk



This figure compares the distribution of two simulated outflow series: one is generated by a homogeneous Poisson model, the other is generated by Model (2D.3) which captures all stylized fact of fund flows. Two model's parameters are estimated from the same fund.

2.3.3 Consequences of clients heterogeneity?

Our second investigation focus on the consequences of the "heterogeneity" in the behaviour of fund investors. In this section, we model fund flows for each of the 8 investor categories described in section 2.3. The objective is to examine the finer components of client's risk. Clients' heterogeneity may imply that the composition of fund client-base can have a great impact on its liquidity risk. If this is true, then two funds - a fund composed by one type

only of investors and another fund with heterogeneous investors - should exhibit different dynamics in their flows. Therefore, in this section, we test whether investors are identical and independent. We first identify the lead-lag relation between two investors in the same fund. Then, we disaggregate one fund's flows into sectors to study whether clients have similar or different behaviour.

Measuring lead-lag effects between investors

We want to prove that fund flows' property are related to client linkage: one investor might react to other investors' massive redemptions to exit the fund. However, there are at least two alternative explanations for the Self-Exciting property observed on funds flows. The first possibility is spurious correlation coming from too many heterogeneous investors aggregated in a single fund. This aggregation may generate the auto-correlation in the time series whereas there is no causality among investors' activities (see, [White 1980](#)). The second possible explanation comes from the existence of common factors which make fund flows correlated (e.g., [Duffie et al. 2009](#)). In this situation, the Self-Exciting property is not necessarily generated by client's behaviour. To confirm that our behavioural explanation is the most credible one, we study fund flows at individual level. We choose two investors (sectors) in Fund 6 and examine the lead-lag effect between them. Any significant lead-lag effect will support our argument in favour of the existence of a "correlation risk". Fund flows' Self-Exciting property is more likely to be due to the investors' linkage than the two alternative explanations.

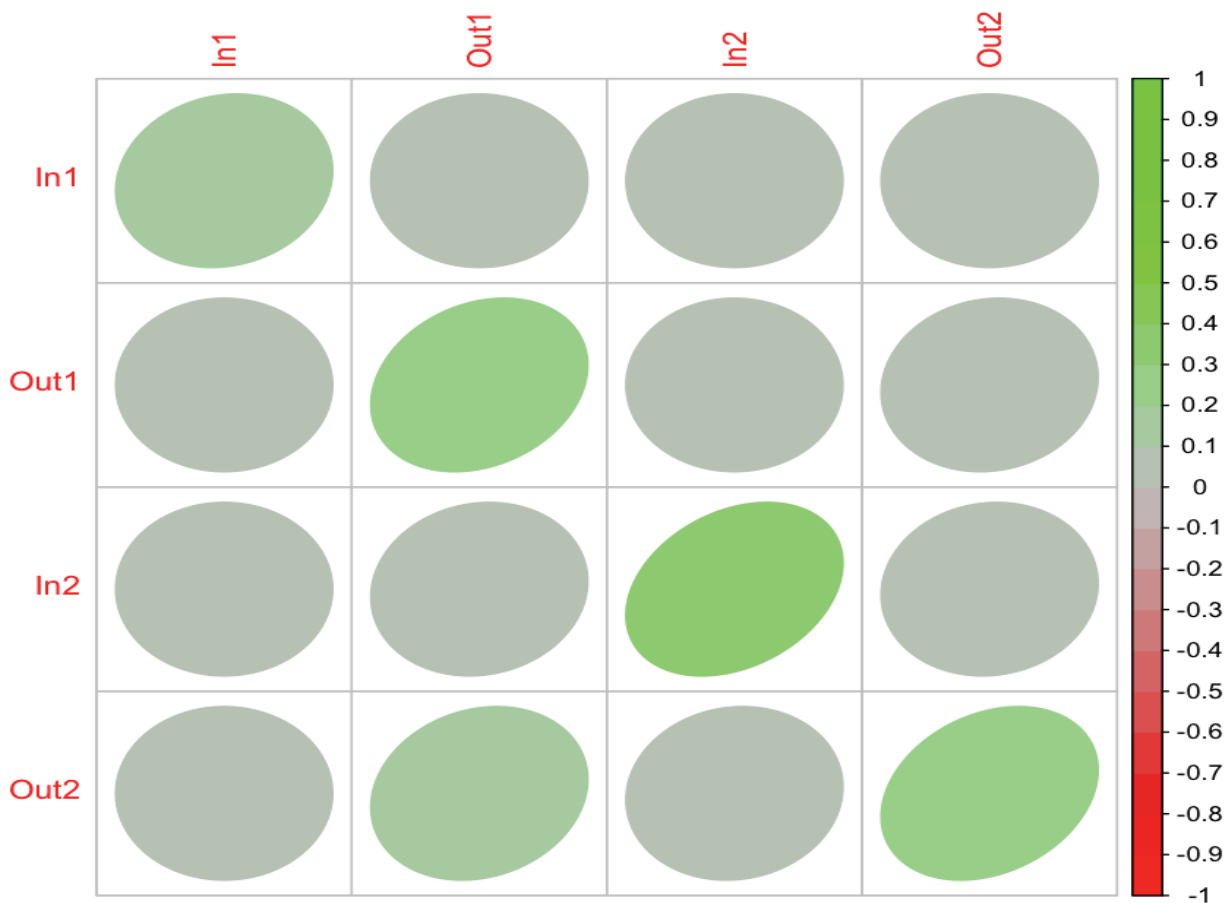
We stick to the analysis of Fund 7 as in the previous Subsection. Two sectors are then chosen since they exhibit largest flow numbers. Sector 1 gathers insurance companies and Sector 2 collects pension funds. Due to their activities, these two sectors frequently receive/pay cash from/to their clients. They both have a short-term and volatile liquidity needs. Consequently, they invest in the money market fund to manage their cash account. They are suitable target to study since they are active investors who generate high flow numbers. We include the liquidity contagion component to the full model to highlight the lead-lag effect. The modified model becomes:

$$\begin{pmatrix} \lambda_t^{in1} \\ \lambda_t^{out1} \\ \lambda_t^{in2} \\ \lambda_t^{out2} \end{pmatrix} = \begin{pmatrix} \lambda_0^{in1} \\ \lambda_0^{out1} \\ \lambda_0^{in2} \\ \lambda_0^{out2} \end{pmatrix} + \begin{pmatrix} \rho^{in1-in1} & \rho^{in1-out1} & \rho^{in1-in2} & \rho^{in1-out2} \\ \rho^{out1-in1} & \rho^{out1-out1} & \rho^{out1-in2} & \rho^{out1-out2} \\ \rho^{in2-in1} & \rho^{in2-out1} & \rho^{in2-in2} & \rho^{in2-out2} \\ \rho^{out2-in1} & \rho^{out2-out1} & \rho^{out2-in2} & \rho^{out2-out2} \end{pmatrix} \times \begin{pmatrix} N_{t-1}^{in1} \\ N_{t-1}^{out1} \\ N_{t-1}^{in2} \\ N_{t-1}^{out2} \end{pmatrix}, \quad (4D1)$$

where *in* and *out* indicate inflows and outflows, 1 and 2 indicate Sector 1 and Sector 2. In this formula, each sector flow series not only react to investors in the same sector (Self-Exciting and Crossed-effect components), but also to the other sectors (Liquidity Contagion). For example, Sector 1's inflow λ_t^{in1} is composed by the baseline intensity λ_0^{in1} , the Self-Exciting $\rho^{in1-in1}$, the Crossed-Effect $\rho^{in1-out1}$ and two contagion components driven by $\rho^{in1-in2}$ and $\rho^{in1-out2}$. Each flow series react to itself and other three flow series. Overall, there are $4 \times 4 = 16$ correlation parameters. In the following, we use a covariance matrix representation to plot model's estimation in Figure 2.6. We find that all Self-Exciting coefficients are significant as the diagonal coefficients are different from 0. However, all Crossed-Effects are insignificant in these two sectors. We identify a contagion effect from Sector 1 to Sector 2. Since the coefficient $\rho^{out2-out1}$ has a non-zero value. It highlights the scenario that the massive redemptions from Sector 1 lead the following redemptions of Sector 2. However, the reciprocal effect does not hold. The coefficient $\rho^{out1-out2}$ is zero. Sector 1 does not react to Sector 2.

This lead-lag effect between the insurance sector and pension sector confirms the correlation risk in mutual fund. Under a stressed scenario, some investors follow other to redeem. This correlated exits make the fund more exposed to the liquidity risk. Furthermore, we find that the liquidity contagion between investors is asymmetric meaning that some investors are leaders while some others are followers and leaders do not react to followers. Therefore, it is important for the fund manager to know the role of each client. If a leader makes a big redemption, the manager should prepare for sequent redemptions of "followers". In contrast, a redemption of followers has less impacts on other clients.

Figure 2.6: Flow Contagion between Investors



In this figure, we present the correlation coefficients of Model (4D.1), estimated in two representative investor categories in one money market fund, over a 5-year period. Each line of the matrix shows the coefficients of one flow series: *In1* is the inflows of the first investor, *Out1* is the outflows of the first investor, *In2* is the inflows of the second investor and *Out1* is the outflows of the second investor. Each column presents flows' reaction to one (other) flow series (ρ). The color indicates the value of the parameter: darker the green is, larger the coefficient is.

Heterogeneity among fund investors

We continue our investigation by examining whether all individual investors (sectors) in the same fund are identical. This study of investors' heterogeneity helps to manage the liquidity risk in two aspects. On the one hand, it helps the fund manager to identify each component of liability and to obtain each investor's risk contribution. On the other hand, since the heterogeneity is a large source of modelling errors, studying fund flows in disaggregated approach might improve the model's performance.

We first break down Fund 7 by sectors. There are eight investor sectors in this fund and each sector gathers investors who exercise the same activity, such as banking, insurance, brokerage, etc. We apply the full model to each sector in the following manner. We consider each sector taken individually as Investor 1, while the others form all together Investor 2. The model then highlights four risk components: (1) the flow average (λ_0) indicates the activism level of the sector; (2) the Self-Exciting (ρ^{In-In} and $\rho^{Out-Out}$) components present investors' reaction to other clients within the same sector; (3) the liquidity contagion shows whether investors follow other sector's flows, we re-note the parameters as ρ^{In-All} and $\rho^{Out-All}$; (4) the Dispersion (Ss) components highlight the heterogeneity level within the sector. Since Crossed Effects are often negligible, we do not present them in this empirical exercise.

Table 2.7 shows the estimated coefficients of the disaggregated test. We find that investors in this same fund are not identical. Fund flows' dynamics in different sectors are highly heterogeneous. Some sectors possess the high flow volume in average, such as Sector 1, 3 and 5. We observe high baseline intensities in these investors. In contrast, Sector 7 and 8's flows are negligible. λ_0 s in these two sectors are almost 0. Majority of sectors exhibits the self-exciting property. Yet, the correlation risk is higher in some sectors. ρ^{In-In} and $\rho^{Out-Out}$ in Sector 4 and 8 exceed 0.5. In other sectors, investors are less correlated. We find the "liquidity contagion" in Sector 4, 5, 6 and 8. These investors are likely to be the followers when other redeem massively. Lastly, we find that all sectors in this fund present over-dispersion property, however, the dispersion level decreases strongly comparing to the fund level dispersion. In Table 2.5 we observe that dispersion coefficients are 0.56 (in) and 0.65 (out) at fund level. However, most of dispersion parameters increase to around 0.8 (except for Sector 2)⁹ at sector level. It suggests that classifying investors in categories helps us to better monitor the liability risk since flows' variance decreases.

⁹Remind: a higher value of coefficient parameter S indicates a lower level of dispersion.

Table 2.7: Disaggregated sector model

Sector	λ_0^{in}	λ_0^{out}	ρ^{in-in}	ρ^{in-all}	$\rho^{out-out}$	$\rho^{out-all}$	S-In	S-Out
1	4.13***	4.59***	0.12***	0	0.25***	0	0.75***	0.89***
2	0.89***	1.56***	0.41***	0	0.29***	0	0.55***	0.67***
3	3.11***	2.78***	0.33***	0	0.33***	0	0.77***	0.78***
4	0.55***	0.40***	0.55***	0.01*	0.54***	0.02**	0.89***	0.91***
5	3.04***	1.76***	0.24***	0.03*	0.36***	0.06***	0.81***	0.91***
6	0.18***	0.01***	0.01	0.08***	0.01***	0.01***	0.98***	0.98***
7	0.03	0.08*	0.03	0.03*	0.02	0.01	0.91***	0.85***
8	0.42***	0.04	0.66***	0	0.56***	0.03***	0.71***	0.71***

This table gives the estimation results of the disaggregated study of all sectors in one money market fund. The study covers a period of 5 years. We number these sectors from 1 to 8 and apply the same model (4.1) to them. We note the significance level at the right-top of each coefficient (*p<0.1; **p<0.05; ***p<0.01). The parameter $\rho^{out-out}$ indicates the reaction of each sector to other's flows. During the "Fund Run", it parameters indicates the risk of contagion between investors (sectors). The dispersion parameter S has a threshold 1: when S is larger than 1, the flow series exhibit the "under-dispersion"; when S is smaller than 1, the "over-dispersion" presents. When correlation coefficients (ρ) equal to 0 and dispersion parameters (S) equal to 1, the model reduces to the traditional iid Poisson distribution.

In the previous section, we show that the heterogeneity among investors is a large source of modelling errors. Therefore, we expect that a disaggregated study would mitigate the errors and improve model's performance. We do a comparison study for model's quality and present it in Table 2.8. We first calculate MSE for both inflows and outflows for each sector and then we sum them up to obtain the total MSE. We compare three models' performance. In the first column, we give the MSE for the aggregated model (2D.3) in the previous section. This model does not provide individual sector's information. In the second column, we apply the same model to each sector. We label it as "Separated Model". The last column ("Contagion Model") gives the MSE of the disaggregated model in this section. We find that, from the aggregated model to separated model, the total MSE in inflows decreases by $60588 - 42209 = 18379$. The reduction is $59129 - 41780 = 17349$ for outflows. This decrease of MSE represents an improvement around 30%. It implies that the fund manager has a great interest to monitor investors at individual level. Finally, we find that the model achieves another slight improvement from separated model to contagion model.

Table 2.8: Model's improvements in disaggregation

Sector/ Model	Aggregated Model	Separated Model	Contagion Model
Sector 1 In	-	9 316.84	9 204.86
Sector 1 Out	-	7291.20	7028.33
Sector 2 In	-	2 006.78	1 909.60
Sector 2 Out	-	2 060.34	1 960.51
Sector 3 In	-	11 077.48	10 922.76
Sector 3 Out	-	11 722.92	11 520.27
Sector 4 In	-	4 488.72	3 792.61
Sector 4 Out	-	4 491.56	3 816.45
Sector 5 In	-	10 970.30	9 822.00
Sector 5 Out	-	11 364.64	10 194.96
Sector 6 In	-	505.07	496.70
Sector 6 Out	-	588.38	557.73
Sector 7 In	-	121.98	122.58
Sector 7 Out	-	140.01	140.92
Sector 8 In	-	3 722.24	3 229.24
Sector 8 Out	-	4 121.50	3 512.10
Total MSE In	60 588.69	42 209.45	39 500.38
Total MSE Out	59 129.57	41780.6	38 731.31
Total MSE	119 718.3	83 990.04	78 231.69

This table presents MSE ratio of three different models estimated for the same fund. We compare the aggregated model (2D.3), separate model and the contagion model (4.1). The aggregated model and separate model use the same specification. Their difference is that we estimate the aggregated model in the full fund sample but we apply the separate model for each sector in the fund. Therefore, the aggregated model give a result at the global level and the separate model presents coefficients for every individual sector. The contagion model includes the "lead-lag" effect in addition. We calculate the MSE for each sector and sum them up for the whole fund, except for the aggregated model does not provide information for individual sectors. We list errors of inflow and outflow separately, "In" indicates inflow's error and "Out" indicates outflow's error.

In summary, investors in the same fund are not identical, nor independent. During the disaggregated study, we identify the "liquidity contagion" in some investors. These investors are led by others to redeem their fund share. When massive redemptions happen, the fund manager should make efforts to persuade these followers to stay. Hence the liquidity risk would not worsen. Furthermore, we find that individual investor sectors present heterogeneous flow dynamics. Each of them bring the different liability risk to the fund. It suggests an individual investigation of fund client behaviour for the future research.

2.4 Controlling for other risk factors

Fund flows have strong links to the fund's asset side. Factors such as manager's performance, fund's risk level and fund's asset class affect investor's trade decision. These links challenge our notion of "liability risk" since the randomness of fund flows can be solely generated by the asset-related factors and there is no influence from the liability. In this section, we implement two robustness checks to confirm the flow risk has liability-side components. We first control for several economic factors by adding them to the Poisson model. Second, we estimate the model in a larger sample to show that funds in the same asset class can have strongly distinct flow risks.

We begin the robustness check by controlling for several "asset-related" risk factors. Our test contains four elementary variables which might affect investors' decision. "Dec" is a dummy variable which equals to 1 if the observation date is in December. The literature documents that investors have tax concerns in this month and they change possibly their behaviour (see, e.g., [Ivkovich and Weisbenner 2008](#)). "Rate" is the short-term interest rate. It is a proxy for the funding cost of institutional investors. We use the french 3-month director rate downloaded from Datastream. "MKT" is the performance of the MCSI European market index' performance. "R" is the fund return at the previous day. We calculate the return by taking the log-difference of two day's NAV value. We add factors' impacts to obtain a modified flow intensity " λ_t^{Factor} " according to the following formula:

$$\begin{aligned}\lambda_t^{InFactor} &= \lambda_t^{In} \times e^{\beta'_{In}X} \\ \lambda_t^{OutFactor} &= \lambda_t^{Out} \times e^{\beta'_{Out}X}\end{aligned}\tag{5.1}$$

Where $\lambda_t^{In/Out}$ is the previous reduced-form intensity in Equation (2D.3). $\lambda_t^{In/OutFactor}$ is the new intensity with factors. β is the vector of factor sensitivities and X is the vector of explanatory factors. In this specification, we include factors by a multiplicative effect:

the new intensity equals to the old intensity multiplies by the $e^{\beta_{In/Out}^t X}$. It differs to the "Self-Exciting" component of the model, where the previous flow numbers ($N_{t-1}^{In/Out}$) have an additional effect: we add the $\rho \times N_{t-1}^{In/Out}$ to the baseline intensity λ_0 . We choose this "multiple function" to avoid the intensity becomes to negative when the factor impact is strongly negative. The estimation follows the same procedure as Equation (2D.3) where we replace the old intensity by the $\lambda_t^{In/OutFactor}$.

We estimate the factor model (5.1) and we present the result in Table 2.9. We list at the first column (**Non-Factor**) coefficients of previous reduced-form model for a comparison purpose and we give the new model's estimation at the second column (**Factor**). We present the liability-related coefficients in Panel A and explanatory factors in Panel B. We observe clearly that previous parameters keep almost the same values (there are only acceptable slightly deviations). After controlling these asset-related factors, the liability-related elements have still significant impacts. The correlation and dispersion risk remain in the fund. We need to emphasis that we can not directly compare the economic significances of reduced parameters and explanatory factors since we include them into the model in different manners (additional effect and multiplied effect).

The interpretation of factor sensitivities is out of the scope of this paper. We would not give a full discussion about these parameters. We merely wish to highlight one surprising coefficient, " $R - Out$ ". Being contrary to the common knowledge, an increase of fund return creates more redemptions! We speculate that this relation is due to the fact that we calculate flow and return in short-horizon. It differs to the literature which tests the flow-return relation in longer horizon, like the monthly frequency. We leave this question to the future research.

Table 2.9: Factor Flow Model

	$\lambda_t^{Factor} = \lambda_t \times e^{\beta'X}$	
	Estimated coefficients	
	(Non-Factor)	(Factor)
Panel A: reduced parameters		
$Base_{in}$	8.15***	9.75***
$Base_{out}$	10.84***	10.49***
$In - In$	0.32***	0.28***
$In - Out$	0.30***	0.25***
$Out - In$	0.19***	0.17***
$Out - Out$	0.46***	0.40***
$S - In$	0.56***	0.57***
$S - Out$	0.65***	0.66***
Panel B: explanatory factors		
$Dec - In$		0.07
$Dec - Out$		0.06
$Rate - In$		0.02***
$Rate - Out$		0.01***
$MKT - In$		0.00
$MKT - Out$		0.00
$R - In$		0.18***
$R - Out$		0.20***
Observations	1252	1252
-2LogL	1654042	16493.71
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

Table 2.9 presents our robustness test which controls for several economic factors. We multiply the previous reduced form intensity by the impact of economic factors ($e^{\beta'X}$). If fund flows are not sensible to these factors, the factor part of the model equals to 1 (e^0) and the model reduces to previous model (2D.3). We list this previous model in the first column for comparison propose and we give the coefficients of non-factor model in the second column. We present the time series coefficients in Panel A and the economic factors in panel B.

2.5 Conclusion and Discussion

Focusing on the liability of mutual funds and its components allows us to answer several new research questions. Firstly we learn that fund investor's behaviour affects strongly fund liquidity risk exposure. A highly correlated and heterogeneous client-base generates a large extreme redemption risk. Second, mutual fund investors are neither identical nor independent. Therefore the fund manager should better monitor his client risk at more deep level. Lastly, the liability risk presents in all types of mutual funds, from the most illiquid fixed income funds to the high liquid money market funds. Hence the asset side should no longer be the only concern of portfolio managers, their clients bring also an important component of the liquidity risk. An asset-liability management approach is preferred than an asset-only approach.

We propose several suggestions based on our study. The first one is related to the fund selling. The fund manager should monitor two components of liability risk, clients' linkage and heterogeneity, when he builds his client-base. A fund with homogeneous and independent investors would receive more regular flows. A possible solution is to offer the fund to a "target" client group. For example, we can sell funds to only one investor category but in a diversified geographical zone. Therefore both two risk components are minimized. Then, our analyse proves that fund's liquidity risk exposure is time-varying. Based on investors' activity level, the fund manager should adjust his cash reserve to prevent the change on future redemptions. Finally, since both asset and liability of a fund contains the liquidity risk, we advise the fund manager to invest in assets whose risk is less correlated to liabilities'. We should avoid the scenario where both asset and liability sides of a fund suffer the liquidity shocks simultaneously.

For the statistical analysis of fund investor behaviour, it is clearly that we are still in the early stage. There are two potential directions to continue this research. One way is to "go up": collecting a larger data-base which contains a larger number of funds. It would allow us to study more about the cross-sectional dispersion on the liability risk. Moreover, numbers of hi-dimensional statistic tools might be useful to analyse the more enormous flow data. The other way is to "go down": examining the individual investor account. An investigation on investors' personal account might bring clearer evidence on their preference and habit on fund investing.

Chapter 3

Survival Analysis Of Mutual Fund

Investors:

Evidence From Micro-Transaction Data

This paper studies clients' investing horizons in mutual funds. A unique data set enables us to monitor all individual redemptions in one non-load French fund company and we adopt the Cox proportional hazard rate model to explain the redemption behaviour. We first find evidence that investors behave rationally under normal circumstances but exhibit the loss aversion bias under extreme returns or at the end of the year. Secondly, we identify that investors' categories have strong impacts on their investing horizons. We explain this finding by the fact that investors in different categories have different levels of liquidity needs. Finally, we make a comparison between long-term and short-term investors and find that long-term investors are more value-oriented since they are more sensitive to economic conditions. Overall, we highlight that investors bring a "pre-mature" redemption risk: exiting the fund earlier than expected. Furthermore, this risk is higher in long-term investor sub-sample. Combining all results, we suggest fund managers to monitor fund investors individually. We also advise mutual funds to introduce some liquidity provisions, such as the "withdrawal gate" or the "lock-up period" in hedge funds, to mitigate the redemption risk and enhance mutual fund's liquidity.

Key words: survival analysis, fund run, redemption risk, investing horizons

Introduction

Open-end mutual funds allow investors to freely redeem their fund shares so as to attract clients who have liquidity constraints to invest in long-term assets. However, this facility introduces a new liquidity risk that the fund manager has to deal with. Not only are frequent and active redemptions harmful for the fund performance, but also it can become critical when many investors are redeeming at the same time. In these situations, analysing investors' behaviour becomes key. Why are investors redeeming? Are these redemptions rational? What are their trigger factors? Are some investors redeem earlier than others and why? Addressing these questions would improve mutual fund's liquidity management.

We investigate fund investors' behaviour and its impacts on the fund liquidity risk in this paper. A particular focus is given to the investing horizon, i.e., the time between the entry and exit of one investor. The lengths of these horizons influence the fund liquidity risk exposure. If most clients exit the fund early, the average duration of fund's liability would be short. Consequently, the fund manager is forced to invest less in long-term assets, otherwise he would face a liquidity mismatch problem. All fund managers, even some world-wide famous ones, are worried about this early redemption problem. For instance, one of world's best fixed income fund managers, Bill Gross, experienced massive redemptions in August 2014. This problem forced him to close his fund and to leave the PIMCO company¹. Fund managers need to figure out which investors are most likely to redeem early and how many are they in a fund. This problem has become a central issue for the fund industry.

In this paper, we use a unique proprietary dataset to study the dispersion in individual fund investing horizons. We name these horizons as the "surviving times". We choose the Cox proportional hazard rate model to explain investor's redemption intensity and we adopt a semi-parametric approach to estimate it. This approach contains a parametric part which identifies explanatory factors and a non-parametric part which leaves baseline intensities free to capture the heterogeneity of the investor sample. Our empirical test proceeds in two stages. First, we estimate redemption rates in the full investor sample. We use both static variables, such as fund class and client category, and dynamic factors, such as fund return and economic conditions. We introduce the notion of the "pre-mature" closure risk, which describes the scenario where investors choose to shorten their surviving times. In the second stage, we classify fund clients into long-term and short-term sub-samples and estimate their hazard rates separately. We analyse their divergence by comparing their sensitivities to the

¹cf. "Pimco Total Return Fund suffers 15th month of outflows" in <https://www.reuters.com/article/us-investing-pimcototalreturn-flows/pimco-total-return-fund-suffers-15th-month-of-outflows-morningstar-idUSKBN0G41GN20140804>

aforementioned explanatory factors.

We establish three key empirical findings. Firstly, We find evidence that investors have a time-varying behaviour: redemptions are rational under normal circumstances but fund clients exhibit the loss aversion bias at the end of the year or under the extreme returns. We identify a negative relationship between the redemption rate and the fund performance during the first three quarters of the year. It supports the theories which predict that rational investors keep funds with good returns (see, [Berk and Green 2004](#); [Brown and Wu 2016](#)). However, in two situations, this negative relation is not held. We observe that, in the fourth quarter of the year, the return-redemption relationship switches to positive. Furthermore, extreme gains also make investors more possible to exit the fund. This evidence suggests that investors exhibit the "loss aversion": a behaviour bias where investors sell better-performed funds immediately but keep the "loser" funds for the longer time.

Secondly, we identify significant sector effects on investor decisions. The sector refers to the activity sector where each fund investor belongs to. We find that banks, private banking clients and insurance companies stay in funds for the longer periods than the average level. On the contrary, the FoFs (fund of funds), mutual insurances, and corporate firms have shorter surviving times. We explain this observation by the fact that investors from different sectors have different levels of liquidity needs. For instance, corporate firms need to pay salaries, dividends and short-term debts. It leads the short investing horizon of this sector. Another example is the FoFs, which need to answer their own redemptions. Therefore their surviving times are also shorten. On the other side, banks and private banking clients have less short-term liquidity needs. Accordingly, we find that they stay in funds for longer horizons. The sector effects are robust after we control for the effects of the asset side liquidity.

Finally, we find notable divergences between the long-term and short-term investors. We test the model separately in the long-term and short-term sub-samples and we identify three major distinctions between them. First, long-term investors' behaviour is more explainable and short-term investors' trades are more noisy. We obtain this conclusion by comparing the R-square of two sub-sample tests. Second, we find that long-term investors are more sensitive to economic variables such as fund performance and macroeconomic conditions. It confirms our hypothesis which predicts that long-term clients are value-oriented so they make more efforts to monitor their investments. Third, we observe that long-term investors exhibit the higher loss aversion. This behavioural bias does not present in the short-term sub-sample. Overall, these findings imply that the long-term investors bring the larger "pre-mature" closure risk to the fund. Hence in stressed scenarios, the fund manager would loss

more long-term clients.

Our paper contributes to several different strands of literature. First, there is a growing literature on the rationality of mutual fund investors. [Berk and Green \(2004\)](#) develop a fund investor theory where investors chase funds with good returns under the rational context. Furthermore, [Berk and Binsbergen \(2016\)](#) find empirical evidence that investors do not only follow good returns but also consider the market risk's exposure. [Franzoni and Schmalz \(2017\)](#) extend Berk-Green model by adding investors' learning on the market risk. In contrast, some other studies indicate that fund clients exhibit several behavioural bias such as the disposition effect, the over-confidence, the narrow framing and the local bias (See, e.g., [Bailey et al. 2011](#); [Chang et al. 2016](#); [Ivkovich and Weisbenner 2008](#) for the description of these behavioural bias.). However, Previous research rarely investigates the individual transactions and authors focus more on fund purchases rather than redemptions, whereas the redemption behaviour has more direct consequences on the liquidity risk. We examine the investor rationality in individual fund redemptions. We first identify during which periods and under which circumstances investors exhibit the loss aversion, then we compare which group of investors exhibit more behavioural bias.

Furthermore, we contribute to the literature which attempts to identify the determinants of fund flows. Previous studies find that the fund performance is the most significant factor for fund flows (see, [Berk and Binsbergen 2016](#)). Besides, the market volatility (e.g., [Cao et al. 2008](#)), the Morningstar rating ([Guercio and Tkac 2008](#)), fund's ranking ([Kempf and Ruenzi 2008](#)), the media coverage ([Solomon et al. 2014](#)), and even the geographical information ([Leung and Kwong 2018](#)) have all significant influences on fund flows. Whilst much interest is taken in fund flow determinants, few literature examines how investors' characteristics impact fund flows. We provide in-depth analysis showing that the investors' sector is one relevant factor to individual fund flows. Contrast to the literature which uses more variables related to fund assets, such as fund return performance and volatility, the investors' sector is a variable which relates to the fund liability, i.e. investor's characteristics and behaviour. Moreover, we explain how investors in the same fund generate different fund flows while previous research does not distinguish them.

Our last research question focuses on the heterogeneity among fund investors. On the one hand, people are interested in the comparison between investors in different investing vehicles. [Akbas et al. \(2015\)](#) contrast mutual fund investors to hedge fund investors. They indicate mutual fund investors are "dump" and hedge fund investors are "smart". [Guercio and Tkac \(2002\)](#) compare mutual fund investors to pension fund investors. [Levy and Liberman \(2015\)](#) compare return-flow relation of active funds to passive funds. On the other

hand, for mutual funds, a large number of studies investigate the differences between retail and institutional investors. [Schmidt et al. \(2016\)](#) document that institutional investors behave more strategically than retail investors during the 2008 financial crisis. [Edelen et al. \(2016\)](#) highlight that institutional ownership generates more market anomalies. [Foster and Warren \(2016\)](#) report that institutional investors use different fund selection criteria than retail investors. In the literature, although considerable attentions are paid to fund clients' heterogeneity, researchers compare rarely different motivations in fund investing. We contrast value investors to liquidity investors and link this difference to the investing horizons. In a recent investigation, [Christoffersen and Xu \(2017\)](#) argues that the loss of "sensitive" investors would modify one fund's return-flow relation. We show that the long-term investors are these "sensitive" ones.

Our paper is closely related to [Johnson \(2004\)](#). We continue his study on how fund clients' investing horizons influence the liquidity risk and we extend it in three major directions. Firstly, we collect a more diversified sample containing various investor types and fund classes, whereas [Johnson](#) investigates only equity funds and retail investors. However, previous studies highlight that investor's category and fund asset class have the significant predictive power on investor's behaviour. For instance, [Ben-Rephael et al. \(2017\)](#) show that institutional investors carry more attentions on performance monitoring than retail investors. [Chen et al. \(2010\)](#)'s theory predicts that investors behave differently in illiquid funds than in the liquid funds. Therefore, it is necessary to examine a more diversified sample than [Johnson](#)'s. Secondly, our tests examine the dynamic aspect of redemption behaviour, while [Johnson](#) focus more on the static prediction of investing horizons at client account's opening. Finally, [Johnson](#) analyses the difference in liquidity consumptions of long-term and short-term investors and we complete his study by comparing factors sensitivities of these two groups. Our study provides a more useful practical guidance for fund liquidity management. A fund manager with a diversified client-base should understand the heterogeneity among investors and monitor the liquidity risk continuously when economic conditions evolve.

We organize the remainder of this paper as follows. Section I describes the database in which we compute the surviving times and give their preliminary statistics. In Section II we introduce the survival model and its estimation approaches. Then we develop our hypothesis about how and why investors redeem in Section III. Section IV presents the economic determinants of the investing horizon. Section V compares the long-term investors to short-term investors. Then we check the robustness of our tests and we conclude at the last section.

3.1 Data

3.1.1 Summary statistics

Our investigation employs a comprehensive data source: the client record of a non-load French fund company. In 2017, the company manages six billion euros in around 50 funds with 128 share classes. It is affiliated to two insurance companies. However, the fund company has an independent management team and its funds are open freely to all investors, either the ones belong to the two parent families or ones from the outside. The fee structure follows industry standards. The company offers a large range of fund classes, such as equity, fixed-income, diversified or money market, and its investors come from different activity sectors, such as banks, insurance companies, retail investors, FoFs, family offices and pension funds. This diversity of the sample makes our results enough generalized, even our sample contains only one fund company.

The fund company transferred to us the electronic copy of its client transaction's record from January 1997 to July 2017. The micro-transaction data records each individual trade where a client purchases or redeems his fund share and contains elementary information of each trade: (1) the date; (2) the fund liquidation price NAV, representing Net Asset Value; (3) the client identity; (4) the client's activity sector; (5) the amount of trade; (6) the fund (share-class) ISIN code. For example, we can identify that: on December 15th 2011, client number 888 from the insurance sector, purchased 5 000 euros of a fund whose share class ISIN is 77896521 ².

Table 3.1 displays the summary statistics for our fund sample. Since the sample covers a long period of more than 20 years, we present statistics year by year to see data's evolution. Over the sample period, 6172 days are opened for transactions, out of 8776 calender days. Within this time, the company has been through three financial crisis respectively in 2001, 2008 and 2011. The company manages less funds in the early days and the fund number starts to increase strongly since 2010. The number of clients increased significantly from 121 in 2004 to 290 in 2006 and stays stable around 300 after. The number here is only the number of client accounts, however, the total number of clients is larger than it for two reasons: (1) some investors come back after the account closure; (2) the company uses sometimes an old account identity for new investors. Clients' transaction volume follows the same trend: after an increase started in 2004, it remains stable after 2006. We observe that the trade number also follows this dynamic. Concerning to clients' category, before

²See the institutional background of mutual fund investing in Chapter 1.

2001, the companies' client-base contains only five to six sectors. Thereafter, the client-base become more diversified and now it contains more than 10 sectors.

Table 3.1: Summary statistics for the fund sample (one company)

Fund	before	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	After	Total
Fund number	6	8	10	13	18	18	21	26	23	30	35	35	44	63	64	70	72	79	86	82	128
Client number	23	25	30	34	86	98	121	230	290	306	306	300	290	360	331	326	324	339	342	311	4491
Trade number	3032	4791	7726	9258	11561	16079	16407	21146	27024	25163	27768	21939	20753	25699	30626	30860	29924	30057	28195	23099	413422
Purchase	1548	2546	4171	4992	6221	8790	8550	10787	14457	13151	12285	9554	10122	13377	15738	14913	14706	14424	13463	10805	205900
Redemption	1484	2245	3555	4266	5340	7289	7857	10359	12567	12012	15483	12385	10631	12322	14888	15947	15218	15633	14732	12294	207522
Date	251	251	273	302	306	304	305	309	278	253	253	256	298	310	308	310	308	305	306	306	6043
Sector	5	5	5	5	10	10	11	12	12	12	13	13	12	13	13	13	13	12	14	13	220
Share	6	8	10	13	18	18	21	26	24	31	36	36	46	67	71	77	82	90	102	99	886

This table presents summary statistics for the fund sample. The data are provided by one French non-load fund company and the sample covers all its clients' micro-transactions over a 20-year period, from 1997 to 2017. We present statistics year by year. In this table, "Fund number" is the number of fund share classes appearing in the sample during each year. "Clients number" presents the number of clients who have executed transactions, "Sector" shows number of investor categories, "Share" indicates the number of share classes. "Trade number" refers to the number of transactions executed. We present thereafter "Purchase" and "Redemption" separately. "Sector" is the number of client sectors.

Being compared to the traditional mutual fund databases, such as Morningstar Direct and LIPPER, which record only fund level information, our data provide an additional possibility to examine transactions at the individual level. We present descriptive statistics for the micro-transactions in Table 3.2. Most investors invest in a small number of funds over the sample period, except for some clients invest in much more funds: the last quartile value of "Number of funds" is 56. Inflows and outflows are symmetric, according to the trade sizes. The average size of individual trade is larger than 2 million euros. However, this statistic is not representative since the calculation is strongly affected by some extreme observations. Then we calculate the inter-trade durations. The average duration between two trades is 100 days. The inflows' durations are shorter, 93 days on average, than outflows' average, 113 days. Both in- and out- durations are highly dispersed. The most frequent investor trades on a daily basis and the most patient one has stayed in the fund for years (1726 days). We calculate the individual holding period returns as the average NAV return of each holding period. On average, investors realize a 0.0049% daily return. This average measure is taken after we drop the extreme situations, i.e. the returns below 3th percentile and above 98th percentile. However, when the extreme returns are included, investors might earn 1.45% or lose 1.64% in a single day. The extreme situations are unreported. The large dispersion in statistics above suggests that investors are highly heterogeneous, and hence we need to consider their different characteristics in our study.

Table 3.2: Summary statistics for micro-transactions

Variables	Mean	Variance	Q1	Q2	Q3	Q4	Q5
Number of funds	5.20	35.13	1.00	1.00	3.00	7.00	56.00
Trade size	2.67 M	74.68 B	588.69	133 303.95	489 697.40	1.55 M	100.06 B
-In size	2.86 M	81.48 B	504.90	140 027.20	520 636	1.71 M	100.08 B
-Out size	2.93 M	1 102 012 B	464.00	0.11 M	0.49 M	1.50 M	149.97 B
Trade duration	100.74	63 396.18	0	2.60	47.90	130.26	1 726.06
-In duration	93.82	69 173.47	0	1.25	35.41	148.77	2 409.50
-Out duration	114. 90	9 802.28	0	3.05	51.25	157.92	2 147.00
Average holding period return	0.0049%	0.0012%	-0.0035%	0.00%	0.00015%	0.00073%	0.0062%

This table presents summary statistics for individual investors' trades. "Number of funds" indicates how many funds each investor holds during the sample period. "Trade size" is the average amount of transactions. We then present the size of purchases ("In") and redemptions ("Out") separately in the following lines. M stands for million and B stands for billion. "Trade duration" refers to the average duration between two trades for each investor. The purchase and redemption sub-samples are given after. Durations are measured in calendar days. "Holding period return" is the individual holding return of each fund investment average holding return. This average return measure is taken after we drop the extreme situations, i.e. the returns below 3th percentile and above 98th percentile.

3.1.2 Preliminary survival analysis

Among the various patterns of individual transactions, our main focus is the individual investing duration in mutual funds. To differ it from the aforementioned inter-trade duration, we label the investing duration as the "surviving time". The term "surviving" comes from the econometric method "survival model" which we use. We calculate the surviving time as the difference between the initial purchase and the final redemption of each investor. We could identify the initial purchase because before this trade the account inventory was 0. Similarly, we identify the final closure as the redemption which clears the account inventory back to 0.

Table 3.3 gives the statistics for our preliminary study on the surviving times. We first show the full sample and then we split our data into sub-samples for investor sectors in Panel A and asset classes in Panel B³. The average surviving time of the full sample is 330 days. However, the surviving time can be extremely long, such as 7 287 days, or extremely short, such 1 day, in some situations. The average fund share of an individual investor is only 0.02% of the fund's AUM. It implies that funds in our sample have the diversified investor bases: an individual investor does not represent the significant portion in the fund. During the whole investment horizon, investors make 30.32 intermediary trades on average. It indicates that investors adjust their fund shares over their surviving periods. However, we do not study these intermediary transactions in this chapter.

In Panel A we present 5 representative sectors: the private banking clients, the insurance company, the corporate firm, the mutual insurance and the fund of funds (FoFs). Private banking clients are "patient" investors: they stay in mutual funds 2 times longer than the average, however, they execute the same number of intermediary transactions. On the contrary, FoFs are less patient clients. They stay in mutual funds for only 97 days, which is lower than a third of the full sample average. Surviving times of the rest three sectors are more closed to the average level: insurance companies stay for slightly longer than one year, corporate firms and mutual insurances stay for slightly shorter than one year. The divergence exists not only in their surviving times but also their average holding returns. Among these five sectors, private banks and corporate firms make the largest profit in mutual funds. They achieve respectively 0.04% and 0.05% daily average returns, which are about as twice higher as other sectors' return (0.02%). Furthermore, corporate firm's investment represent the largest portion in a fund's client base. On average, their investments represent 1.06% of fund's AUM, whereas this number in other sectors is only negligible. All sectors make

³There are 16 sectors and 6 asset classes, as presented in Table 3.1. However, we choose only 6 representative sectors and 3 asset classes here. Other sub-samples are too small to provide reliable statistics.

around 30 intermediary transactions during their surviving time, except for the insurance sector which trades a half less.

In Panel B we show the sub-sample statistics for three fund class sub-samples. Investors in money market funds have the shortest investment horizons, 91 days on average. It might be explained by the fact that investors consider these funds as short-term liquidity management tools, hence they do not take the long-term perspective in these funds. The class "absolute performance" contains the funds managed 100% by the active strategies and thus they do not follow any benchmark. Most of investors stay in these funds longer than 1 year. The average surviving time in fixed income funds is only 393 days which is a potential risk for the fund management as many fixed income securities are relatively illiquid. Investors earn a 0.04% average daily return in fixed income funds. In other two fund classes, the average daily return is much lower. Respect to the investing size at account opening, fixed income funds receive the largest client. The average size reaches to 0.04%. Individual transaction sizes in other two classes are smaller. It implies that fixed income funds have the least diversified investor base. In average, investors make 34.44 intermediary trades during their survival times in money market funds. This number is much higher than ones in other two fund classes.

Through this preliminary analysis, we find that surviving time varies in different sub-samples. It seems to be linked to different characteristics such as the investor sector and the asset class. Moreover, fund return might also have impacts on investors' decisions. This gives us the motivation to identify the factors explaining the difference in surviving times.

Table 3.3: Preliminary analysis of surviving times

	surviving time		Quantiles					<i>Holding return</i>	<i>Opening size</i>	<i>Trade number</i>
	<i>Mean</i>	<i>Variance</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>			
Full sample	330	451 393	1	2	28	354	7 287	0.01%	0.02%	30.32
Panel A	Sectors:									
<i>Private bank</i>	707	1 027 713	1	24	254	1 024	5 578	0.04%	0.02%	30.42
<i>Insurance</i>	575	597 630	1	53	251	828	4 712	0.0014%	0.02%	16.50
<i>Corporate firm</i>	253	2 424 464	1	3	87	328	3 746	0.05%	1.06%	33.43
<i>Mutual insurance</i>	271	297 571	1	2	36	321	6 544	0.02%	2.04%	29.91
<i>FoF</i>	97	139 312	1	2	4	24	7 287	0.01%	0.01%	29.61
Panel B	Asset Classes:									
<i>Fixed income</i>	393	344 788	1	5	186	509	5 908	0.001%	0.04%	7.03
<i>Money market</i>	91	87 331	1	2	6	44	7 287	0.0005%	0.01%	34.44
<i>Absolute performance</i>	450	473 923	1	24	237	627	6 606	-0.005%	0.0001%	5.13

This table presents the descriptive statistics for the surviving times. The surviving time stands for individual investing horizon which counts the the number of calender days between the initial purchase and the account closure for each individual investor. We list their mean, variance and five quartile values. Then we calculate the holding return which is the log-difference between the NAV value at the purchase and at the closure. We convert all returns to the daily basis for comparison purpose. "Opening size" is the investor's initial inventory divided by fund's AUM. "Trade number" is the number of intermediary trades during each survival period. We use it as an indicator of trade frequency. We start by showing the full sample statistics on the top, then we give information for representative client sectors in Panel A and asset classes in Panel B.

3.2 Empirical method

3.2.1 Cox proportional hazard rate model

We adopt the "Survival Model" which is a standard choice to study the duration-type data in the literature. This approach is widely used in the financial modelling for various topics such as the default risk assessment (see, [Duffie et al. 2009](#)), hedged fund failures analysis (e.g., [Darolles et al. 2013](#), [Liang and Park 2010](#)), bank distress prediction ([Cox et al. 2017](#)), and also mutual fund redemption examination (see, [Johnson 2004](#) and [Ivovich and Weisbenner 2008](#)).

We first introduce several key notations. We use i to index the individual investor and hence the random variable T_i represents the surviving time of investor i . Accordingly, the survival function $S_i(t) = \mathbb{P}(T_i > t)$ gives the probability that this investor survives (stays in the fund) still after time t . It is more convenient to study the survival probability by identifying its complement, the hazard rate $h_i(t)$, which specifies the instantaneous rate of redemption of investor i at time t , conditional on the fact that investor survives till the time t . More formally, it is defined as:

$$h_i(t) = \lim_{\Delta t \rightarrow 0} \frac{[(t \leq T_i < t + \Delta t) | T_i \geq t]}{\Delta t} \quad (2-1)$$

This hazard rate is equivalent to the exit rate, the account closure probability or the redemption intensity. In the rest of paper, we use all these equivalent terms to indicate this conditional probability.

We follow [Cox \(1972\)](#) to introduce the vector of economic factors $\mathbf{X}_{i,t} = \{x_{i,t}^1, x_{i,t}^2, x_{i,t}^3 \dots x_{i,t}^j\}$ with length j to explain the redemption intensity as:

$$h_i(t) = h_0(t) e^{\mathbf{X}'_{i,t} \boldsymbol{\beta}} \quad (2-2)$$

Where the i represents the individual investor, t represents the observation date and j indicates the explanatory factor. Each factor takes value for i and t fixed. For instance, $x_{i=1,t=6}^2$ stands for the value of x^2 for investor 1 at the date 6. $h_0(t)$ is the baseline hazard rate. It is the redemption rate when all factors are 0. We give its formal definition in the end of this sub-section where we explain the estimation method. $\mathbf{X}'_{i,t}$ denotes the transpose of the variable vector and $\boldsymbol{\beta}$ is the vector of the factor loadings with the same length. We can write $\mathbf{X}'_{i,t} \boldsymbol{\beta}$ in a linear form:

$$\mathbf{X}'_{i,t}\boldsymbol{\beta} = \beta^1 x^1_{i,t} + \beta^2 x^2_{i,t} + \beta^3 x^3_{i,t} + \beta^4 x^4_{i,t} \dots + \beta^j x^j_{i,t} \quad (2-3)$$

We need to clarify that the model we choose explains only the cross-sectional dispersion among surviving times. Although there is a time index, the estimation approach makes restraint assumptions about the relation between $h_0(t)$ and t . The time series statistical are not studied in this paper ⁴.

Equation (2.2) shows that the redemption rate of investor i at time t as the product of: (a) a baseline intensity $h_0(t)$; (b) the factor impact $e^{\mathbf{X}'_{i,t}\boldsymbol{\beta}}$. We follow the literature to call these factors as covariates. They add multiplicative effects to the baseline hazard rate. Hence people call this model as "Cox Proportional Hazard Rate Model". Conventionally, people rescale all covariates to ensure they are centred at 0. By doing so, $h_0(t)$ becomes the average redemption rate at the time t (see, e.g., [Liang and Park 2010](#)).

Under this approach, we do not interpret factor sensitivity $\boldsymbol{\beta}$ directly. Instead, we study how explanatory factors drive the risk probability $h_i(t)$ away from the baseline rate $h_0(t)$. We define the hazard ratio for the factor x^j as:

$$HR = \frac{h_0(t)e^{\beta^j}}{h_0(t)e^0} = e^{\beta^j} \quad (2-4)$$

If a hazard ratio is greater than one, the factor increases the closure risk since it is associated to a higher redemption intensity than that of the baseline level. Similarly, if a hazard ratio is less than one, the fund account becomes less risky because it is associated with lower redemption intensity, relative to the average account. In fact, only the sign of β , rather than its value, is directly informational: the positive sign for risk increasing and the negative sign for risk decreasing. For example, if the redemption intensity has a sensitivity $\beta = 0.784$ to one factor, the closure risk becomes larger since investor's exit probability increases by $e^{0.784} - 1 = 119.02\%$. Or, on the contrary, if another factor has a β of -0.656 , it reduces the liquidity risk since the closure rate is decreased by $1 - e^{-0.656} = 48.107\%$.

We keep the semi-parametric approach used by [Johnson \(2004\)](#) to estimate the model. Under this approach, a mild assumption about the baseline hazard rate is made. We first

⁴There are several alternative methods to study the time series properties of survival rate. (1) when we suppose $\log\{h(t)\}$ is a constant c , the surviving time follows an exponential distribution with density $p(t) = ve^{-ct}$. In this case, the redemption rate is constant. Obviously, it is not consistent with our data. (2) We might let $\log\{h(t)\}$ increasing linearly with time, $\log\{h(t)\} = c + \rho t$. In this case, the hazard rate follows a Gompertz distribution. (3) When $\log\{h(t)\} = c + \rho \log(t)$, it leads a Weibull distribution of the hazard rate. These three approaches focus more on the time series properties of redemption intensities. They might serve as the reduced form forecasting tool. However, there is an inconvenience is that these methods do not identify the factors which affect the hazard rate.

subdivide time into reasonably small intervals and then we assume that $h_0(t)$ is constant in each interval but it increases over time. This assumption leads the hazard rate to follow a piece-wise constant exponential distribution and the survival rate decreases naturally with time. It is consistent with the fact that the longer an investor stays in the fund, the more possible she/he would exit.

The semi-parametric approach is chosen since it fits better our data than two alternatives, parametric approach and non-parametric approach. In parametric approach, people assume an explicit form for $h_0(t)$, such as a constant. This assumption ignores the high heterogeneity in the data. In non-parametric approach, people focus only on factor sensitivities ($\mathbf{X}'_{i,t}\boldsymbol{\beta}$) and leaves the $h_0(t)$ completely unspecified. It would be complicate for us to compare the behaviours of different investors ⁵.

In this paper, we keep this semi-parametric approach consistently for all regression tests and robustness checks. The model first assumes the benchmark redemption rate for each time interval, then it highlights how factors drive the hazard rate away. If a factor indicates a characteristic of the fund investing, such as the asset class, it shows how this type of investment differ from others. Or, if a factor is an economic variable, such as the market return, its coefficient shows how investors interpret this economic condition. We give the description of all variables in the next sub-section.

3.2.2 Presentation of covariates

In Table 3.4, we present all variables being used in our tests. We follow the econometric literature to call these variables as "covariates". Some of them are static factors which are collected when the investors open their accounts. They indicate characteristics of investors or funds and their values stay constant thereafter. On the contrary, other covariates are dynamic variables which might evolve over time and we could not forecast them at account opening. We use the term "**Pre-Mature Closure Risk**" to represent the situation where some dynamic variables lead investors to exit the fund early than predicted. We put similar covariates into a panel for a better presentation.

Panel A displays the return-related variables. We use various different return metrics, which provide richer information on the investor behaviour than the previous literature which often uses only one single return measure. We first contrast the individual holding return to the fund return. The individual holding return $R^{holding}$ is computed as the log-difference between the fund NAV at each account opening and the actual NAV, rescaled

⁵See the discussion in Section 3 of [Ivkovich and Weisbenner \(2008\)](#) for the comparison of three estimation methods.

to the annual basis . Investors in a same fund might achieve different holding returns. Then, we compute the fund return, which is identical for all investors in the same fund, in different horizons: R^{Daily} is the return of fund in the day before the observation date; R^{weekly} is the average return during the week before each account closure; $R^{Monthly}$ is the average return during the previous month; Lastly $R^{Quarterly}$ is the average return during the previous quarter. We have checked the correlation of these four returns, which is weak and would not generate the co-linearity problem. The previous mutual fund research does not investigate returns in different horizons. However, studies on the household finance indicate that returns in different horizons might have different impacts on investors' decisions (see, [Grinbatt and Keloharju \(2001\)](#) for the equity market). We further highlight whether a daily return is higher than 10% or lower than -10%. We label these returns as "Extreme-Gain" and "Extreme-Loss". The behavioural finance theory documents that investors may be irrational facing extreme returns.

Table 3.4: Covariates' presentation (for the survival analysis)

Variable	Type	Description/calculation
Panel A: return covariates:		
$R^{holding}$	Dynamic	The individual holding period (Nav-)return
R^{daily}	Dynamic	Daily fund (Nav-)return
R^{weekly}	Dynamic	Weekly average fund (Nav-)return
$R^{monthly}$	Dynamic	Monthly average fund (Nav-)return
$R^{quarterly}$	Dynamic	Quarterly average fund (Nav-)return
Extreme-Gain	Dynamic	Dummy: if the daily return is higher than 10%
Extreme-Loss	Dynamic	Dummy: if the daily return is worse than -10%
Panel B: time controls:		
Dec	Static	Dummy: the month December
Q4	Static	Dummy: the fourth quarter
Period 1	Static	Dummy: the period before the 2000.12.31
Period 2	Static	Dummy: the period between the 2002.01.01 and the 2008.06.30
Period 3	Static	Dummy: the period from the 2009.01.01 to the 2011.04.30
Period 4	Static	Dummy: the period from the 2011.12.31 to the 2014.12.31
Crisis 2001	Static	Dummy: the period from the 2001.01.01 to the 2001.12.31
Crisis 2008	Static	Dummy: the period from the 2008.06.30 to the 2008.12.31
Crisis 2011	Static	Dummy: the period from the 2011.05.03 to the 2011.12.30
Panel C: other controls:		
Size	Static	The initial purchase amount of the investor
Trade frequency	Dynamic	Number of trades divided by the surviving time
Variance	Dynamic	Monthly variance of fund returns
Net previous flow	Dynamic	The net fund flow in the previous day
Panel D: macro controls:		
Global activity	Dynamic	The total number of trades in the fund company
Short term rate	Dynamic	3-month french benchmark rate (Euribor)
Rate spread	Dynamic	The bid-ask spread of the short-term rate
Market return	Dynamic	The return of MSCI equity index

(continued)

The table above is the 1st part of Table 4. The second part is in the next page.

continued-Table 3.4

Variable	Type	Description/calculation
Panel E: fund asset class:		
Money Market	Static	Dummy: money market funds
Fixed Income	Static	Dummy: fixed income funds
Absolute Performance	Static	Dummy: funds adopted the pure alpha strategy
Panel F: investor category:		
$Bank^{own}$	Static	Dummy: bank investors
$Bank^{private}$	Static	Dummy: banks' "private wealth management" client
Insurance	Static	Dummy: insurance companies
Corporate	Static	Dummy: corporate firms
Mutual	Static	Dummy: mutual insurance
FoF	Static	Dummy: fund of funds

This table lists all covariates used in this paper. Most variables are provided by the fund company, except for the macroeconomic variables which we collect from the Datastream. We list the name of each covariate in the first column and we indicate either the variable is static or dynamic in the second column. "Static" shows that the variable is constant during the surviving time. The "Dynamic" type indicates that the variables evolve over time. We give a brief description in the last column.

We present the variables to control for the time effect in Panel B. We split the whole sample period into multiple parts since our data cover a long period over 20 years. "Period 1" represents periods before the 31st December 2000, "Period 2" is the period from the 1st January 2002 to the 30th June 2008, "Period 3" is from the 1st January 2009 to the 30th April 2011, and "Period 4" lays from the 31st December 2011 to the 31st December 2014. Under this specification, we consider the period after the 31st December 2014 as our benchmark. The coefficients of these 4 period dummies would show how investors behave before, being compared to the benchmark period. In this approach, four periods are approximately separated by three recent financial crisis. We consider after each crisis, the global activity level of investors would change.

Next, we indicate whether an account was closed during the crisis. Dummy variables $Crisis_{2001}$, $Crisis_{2008}$ and $Crisis_{2011}$ represent the periods during the three recent past financial crisis. The literature argues that the method to choose the crisis period has an impact on research results (e.g. [Dungey et al. 2015](#)) and it is difficult to define the days where the crisis were happening. In our paper, we use only a simple approximation. We consider: the full year of 2001 is the period of the first crisis; from the 1st July 2008 to the 31st December 2008, the second crisis was happening; the third crisis covers the period from the 1st June 2011 to the 31 December 2011. After, we also control the end-of-year effects. "Dec" shows if a trade takes place in December. This specific month's observations might be biased by tax issues. "Q4" is the fourth quarter of the year. Previous literature indicates that at end of year, investors' risk aversion strongly increase (see, [Ben-Rephael et al. 2012](#)) and hence they might change their behaviour.

Panel C gives other control variables. "Size" is the trade amount. We rescale the money value into the percentage of the fund AUM. The common knowledge suggests that investors with large amount have more "skin in the game". Therefore they should make more efforts to monitor the fund performance. "Trade frequency" is the number of trades each investor executes divided by his total surviving time. "Variance" is the monthly variance of the fund daily returns. "Net previous flow" is the net fund flow at the day before the account closure. We use this variable to control for the situation where investors follow other to exit the fund.

Panel D lists the variables which we use to control for the macroeconomic conditions. The variable "Global activity" is the total daily number of clients' transactions that the fund company receives. It controls for the overall activity level of fund investors. The covariate "Short term rate" is the the French three month reference interest rate, the Euribor. We use this rate as a proxy for the funding cost which could impact investors' redemptions. Then, "Rate spread" is the bid-ask spread for the previous interest rate. We use it as an approximation for the market liquidity level. "Market return" is the return of the MSCI equity index and it shows the market evolution. We download these three variables from Datastream.

In Panel E, we present dummy variables to identify whether a fund is a money market fund, fixed income fund or "absolute performance" funds. Investors in the rest of funds, such as equity funds or diversified funds, have similar surviving times, thus we consider that them have the "benchmark" surviving times. The model shows whether investors in other funds lengthen or shorten their surviving times, being compared to the benchmark level. Similarly, we show some investor sectors in Panel F: "*Bank^{own}*" indicates if the investor is a bank; "*Bank^{private}*" indicates if the investor is a private banking client; "Insurance" is the normal insurance company; "Mutual" represents the mutual insurance company. The normal insurance company issues more life insurance products and the mutual insurance company issues more short term products, as healthcare insurance. "Corporate" stands for corporate firms and "FoF" indicates investor is a fund of funds.

3.3 Hypotheses on investor's behaviour

We are interested in three research questions regarding the fund investor's behaviour. Our first concern is to investigate how a fund with diversified client-base differs from a fund with single client category. In the previous studies on mutual funds, [Johnson \(2004\)](#) and [Ivkovich and Weisbenner \(2008\)](#) study only retail investors, [Dahlquist et al. \(2017\)](#) examine

only pension participants. However, the modern delegated asset management requires the fund manager to receive different types of investors. We wonder whether investor's type has the impacts on their behaviours, hence our first hypothesis is:

Hypothesis 1: *The surviving time is linked to investor's activity sector. (H1).*

We include six aforementioned sector dummies in our test. These sectors are chosen since they represent an important source of fund companies' client base and the preliminary analysis in Section 1 shows these sectors deviate their surviving times from the average level (c.f. Table 3.3). The others sectors in our sample are considered to have the benchmark surviving time. If we obtain a significant positive sector coefficient, this sector has a shorter surviving time than average, whereas a positive coefficient indicates a longer surviving time. If the coefficient is not significant, there is no sector effect.

We are expected to observe significant coefficients at least in some sectors. It implies that a diversified client-base needs to be managed differently than a simple client-base and the fund manager is required to monitor investors at the individual level. Otherwise, investors behave homogeneously and there is no need to distinguish them.

The (H1) focus on the static side of investor behaviour, then in the second step, we turn to the dynamic side. An investor's surviving time should not only be determined by his type, the evolved economic conditions would also have an impact. Among various economic variables, we choose to test the effect of the return on fund investing since the return is the most direct information that investors receive. Thus our second question is how fund investors interpret previous returns.

According to the studies of [Ben-Rephael et al. \(2012\)](#) and [Kamstra et al. \(2017\)](#), fund investors might exhibit a time-varying behaviour. Especially, in the fourth quarter of the year, they are more likely to exhibit the loss aversion. This might be explained by several facts: (1) securities held by the mutual funds could show a seasonality in their returns, hence investors react accordingly to modify their behaviours. (2) fund clients wish to show a good realized return in the annual reporting. To this end, they might adjust their portfolio at the end of the year. (3) sentiment factors might affect investors' decisions and they lead a high loss aversion in the fourth quarter of the year. In line with these facts, we hold the opinion that the return-redemption relationship might also be time-varying. Therefore, we propose our second hypothesis as:

Hypothesis 2: *Investors do not react to the previous return in the uniformly way (H2).*

The relationship between the return and the redemption propensity reveals investor's rationality. According to the traditional fund investor theories such as [Berk and Green](#)

(2004) and [Brown and Wu \(2016\)](#), rational investors should interpret positively the previous return since the good return is a sign of the superior ability of the fund manager. Hence, they should exit the worse-performed fund and keep the better-performed funds. On the contrary, the behavioural finance theory indicates a different return-redemption relationship. The previous research documents that fund investors might exhibit the loss aversion, a commonly committed emotional error (see, [Bailey et al. 2011](#) and [Odean 1999](#))⁶. With this bias, investors exhibit an extremely high level of risk aversion and they feel bad to realize the loss by selling a bad-performed security. Therefore, if a security has delivered a bad performance, investors prefer to hold it than sell it. In contrast, when a security performs well, they sell it quickly to avoid that the performance worsens in the future. Therefore, the return-redemption relationship would exhibit a positive sign in this situation.

Combining with the evidence on investors' seasonal behaviour in [Ben-Rephael et al. \(2012\)](#); [Kamstra et al. \(2017\)](#), we assume that their loss aversion is most likely to appear in the last quarter of the year. Therefore, we are expected to obtain the positive return-redemption relation in the fourth quarter when investors are most likely to exhibit this behavioural bias. However, during other three quarters, we assume that investors behave rationally and the return-redemption relationship switches to negative.

Investors' reactions to dynamic factors introduce the "pre-mature redemption risk", i.e. fund clients shorten their surviving times under certain economic contexts and exit the fund earlier than the asset manager has expected. It implies that the fund's liquidity risk exposure evolves with different economic conditions. Over a period of bad economic circumstance, the fund manager needs to be prepared to clients' reactions. Accordingly, we finish our analysis by testing which group of fund investors brings the higher pre-mature redemption risk.

Our last test concerns to the distinction between the long-term investors and short-term investors. [Johnson \(2004\)](#) shows how to identify whether an investor is a long-term or short-term one, by observing his/her characteristics. In the line of his study, we take a further step to examine the differences between these two investor groups. We intend to identify which investors are more sensitive to economic conditions, because the more sensitive investors bring the higher liquidity risk. Therefore, the fund manager should make more efforts to monitor "sensitive" investors.

The differences between LT and ST investors are generated by their distinguished investing motivations. We argue that investors invest in mutual funds for two principal mo-

⁶In mutual fund literature, some authors also denote this behaviour bias as the "disposition effect" (see, e.g., [Chang et al. \(2016\)](#)).

tivations: 1, the fund manager has an ability to achieve a superior return; 2, mutual funds' open-end structure satisfies random liquidity needs of investors. Investors who focus on the value appreciation (motivation 1) should stay longer than others, since the long horizons allow them to accumulate profits. In contrast, short-term investors use mutual funds as cash management tool (motivation 2). These "liquidity traders" redeem in response of their individual liquidity needs, which are difficult to predict at the aggregated level. Therefore, we propose our third hypothesis:

Hypothesis 3: *Long-term investors' behaviours are more sensitive to economic conditions and more understandable. (H3).*

We proceed our examination by dividing the data into long- and short-term sub-samples. We then estimate the hazard rates of these two sub-samples and compare investors' sensitivities to different economic factors. If **(H3)** holds, we suppose to identify greater or more significant sensitivities to covariates in the long-term sub-sample. Moreover, we expect to find a higher R^2 in the long-term sub-sample ⁷.

We start with examining the full sample to answer **(H1)** and **(H2)** in Section 4. Then, we analyse long-term and short-term sub-samples separately in Section 5, since we are interested in the distinction between them. This sub-samples study address our third research question.

3.4 Why investors redeem their fund shares?

In this section, we estimate Model (2-2) with all covariates presented in Section 3. This test examines our hypotheses **(H1)** and **(H2)**, and we present estimated coefficients in Table 3.5. We study all individual account closures of the fund company between the January 1st 1997 and the June 30th 2017. Each observation of surviving time is computed as the period between a fund account's opening and closure. We compute the percentile of surviving times and we drop observations which exceed the 98th percentile, 3 722 days, and ones which are shorter than the second percentile, 20 days. These observations are considered as outliers. ⁸ We end up with 7 912 observations. Among them, we observe 5 584 account closures. Accounts which stay open at the end of observation period are censored. We present each coefficient together with its significance level and its standard error. According to Equation (2-4), coefficient's value is not directly interpretable. Instead, we need to transform them into the Hazard Ratio **(HR)** by taking their exponential values. In the bottom of the table,

⁷In the rest of the paper, we use LT to indicate long-term investors and ST for short-term investors.

⁸ However, even when we keep these observations in the sample, the major findings hold. We do not present this additional test.

we present some standard criteria for model's quality such as R-square and Wald statistic.

Table 3.5: Cox proportional hazard rate model: full sample estimation

	<i>Conditional account closure probability:</i>		
	$h_i(t) = \lim_{\Delta t \rightarrow 0} \frac{[(t \leq T_i < t + \Delta t) T_i \geq t]}{\Delta t}$		
	(1)	(2)	(3)
A: Performance variables:			
$R^{holding}$	-0.927*** (0.116)	-0.729*** (0.123)	-0.690*** (0.124)
R^{daily}	0.605 (0.928)	0.840 (1.097)	-0.271 (1.054)
R^{weekly}	-0.125 (0.940)	-0.321 (1.105)	0.747 (1.061)
$R^{monthly}$	0.282 (0.192)	-0.008 (0.202)	0.390* (0.217)
$R^{quarterly}$	-0.703*** (0.100)	-0.591*** (0.118)	-0.768*** (0.123)
B: Return controls:			
$Q4 \times R^{holding}$	0.933*** (0.115)	0.730*** (0.123)	0.694*** (0.123)
$Q4 \times R^{quarterly}$	1.241*** (0.171)	1.236*** (0.198)	1.774*** (0.286)
Q4	0.004 (0.035)	-0.018 (0.035)	-0.098*** (0.035)
$Dec \times R^{holding}$	-1.005*** (0.222)	-1.112*** (0.248)	-1.441*** (0.258)
$Dec \times R^{monthly}$	-0.859*** (0.247)	-0.673** (0.269)	-1.444*** (0.345)
Dec	-0.111*** (0.042)	-0.159*** (0.042)	-0.131*** (0.043)
Extreme-gain	-0.991*** (0.082)	-0.607*** (0.086)	-0.566*** (0.086)
Extreme-loss	-1.220*** (0.137)	-0.904*** (0.148)	-0.945*** (0.159)
Extreme-gain \times Q4	0.536*** (0.189)	0.753*** (0.198)	0.982*** (0.199)
Extreme-loss \times Q4	0.158 (0.253)	0.272 (0.270)	0.112 (0.285)

(continued)

The table above is the 1st part of Table 3.5. Other parts are in following pages.

continued-Table 3.5

	<i>Conditional account closure probability:</i>		
	$h_i(t) = \lim_{\Delta t \rightarrow 0} \frac{[(t \leq T_i < t + \Delta t) T_i \geq t]}{\Delta t}$		
	(1)	(2)	(3)
C: Time controls:			
Period 1	0.517*** (0.064)	0.065 (0.065)	-1.461*** (0.093)
Period 2	-0.500*** (0.046)	-0.211*** (0.047)	-0.997*** (0.056)
Period 3	-0.660*** (0.051)	-0.457*** (0.051)	-0.755*** (0.053)
Period 4	-0.424*** (0.048)	-0.292*** (0.049)	-0.382*** (0.049)
Crisis 2001	-0.218*** (0.062)	-0.211*** (0.062)	-1.100*** (0.071)
Crisis 2008	-0.422*** (0.092)	-0.147 (0.093)	-0.615*** (0.095)
Crisis 2011	-0.559*** (0.084)	-0.386*** (0.084)	-0.647*** (0.085)
D: Other controls:			
Size	-0.0002 (0.004)	0.001 (0.005)	0.001 (0.005)
Trade frequency	-0.0004*** (0.0001)	-0.002*** (0.0001)	-0.001*** (0.0001)
Variance	0.454*** (0.101)	0.387*** (0.115)	0.630*** (0.124)
Net previous flow			-0.548 (0.501)
E: Macroeconomic controls:			
Global activity			0.00002*** (0.00000)
Short term rate			0.278*** (0.010)
Short term spread			-0.849*** (0.156)
Market return			-0.0003*** (0.00003)

(continued)

The table above is the 2nd part of Table 3.5. The final part is in the next page.

continued-Table 3.5

<i>Conditional account closure probability:</i>			
	$h_i(t) = \lim_{\Delta t \rightarrow 0} \frac{[(t \leq T_i < t + \Delta t) T_i \geq t]}{\Delta t}$		
	(1)	(2)	(3)
G: Investor sector			
<i>Bank^{own}</i>		-0.217*** (0.068)	-0.271*** (0.068)
<i>Bank^{private}</i>		-0.121** (0.059)	-0.135** (0.059)
Insurance		-0.226*** (0.084)	-0.234*** (0.084)
Corporate		0.284*** (0.080)	0.246*** (0.080)
Mutual		0.235*** (0.035)	0.213*** (0.035)
FoF		0.480*** (0.043)	0.507*** (0.043)
F: Asset class			
Money Market		1.294*** (0.035)	1.111*** (0.037)
Fixed Income		0.265*** (0.048)	0.205*** (0.048)
Absolute performance		0.019 (0.067)	-0.108 (0.067)
Observations	7,912	7,912	7,912
R ²	0.153	0.361	0.420
Max. Possible R ²	1.000	1.000	1.000
Log Likelihood	-47,287.920	-46,172.880	-45,791.950
Wald Test	1,112.990*** (df = 25)	3,330.450*** (df = 34)	4,335.430*** (df = 39)
LR Test	1,313.454*** (df = 25)	3,543.534*** (df = 34)	4,305.386*** (df = 39)
Score (Logrank) Test	1,182.234*** (df = 25)	3,843.898*** (df = 34)	4,803.571*** (df = 39)

Note:

*p<0.1; **p<0.05; ***p<0.01

In this table we present the survival analysis of the full investor sample. We adopt a semi-parametric approach to estimate conditional probability of individual fund account closures. The sample covers all daily mutual funds of one French fund company, from the 1st January 1997 to the 30 June 2017. Each "surviving time", the individual investor's staying horizon in fund, is computed as the duration between the initial purchasing and the final exit. The final exit is the redemption which clears the fund account inventory. The accounts which are still open at the 30 June 2017 are censored. The baseline hazard $h_0(t)$ is piece-wise constant and it represents the redemption intensity when all covariates take their average value. We use standard approach to calculate modelling errors and coefficients' significance level. We put related covariates in the same group and present these groups in different panels. Panel A presents performance related covariates. Panel B shows control variables for return metrics. Panel C shows control variables for the time effects. Panel D has other control variables. Panel E highlights macro-economic factors. Panel F and G contribute to the fund and investor types.

For ease of presentation, we use different panels to classify covariates as in the previous section: Panel A presents performance related covariates, Panel B shows return control variables, Panel C exhibits time control variables, Panel D lists other control variables, Panel E highlights macroeconomic factors, Panel F and G display fund and investor types. We test 3 different specifications of the hazard rate. Each column in Table 5 lists the result of one specification. Specification (1) contains only the return measures and elementary control variables (Panel A, B, C and D). Specification (2) includes additionally investor sectors and fund asset classes (Panel F and G). Finally the last column is the full specification (3), where we add the macro-economic shocks (Panel E).

We observe improvements of model fitting from Specification (1) to (3). Specification (1)'s adjusted R-square is only 15%. Specification (3) improves this ratio by more than twice ($R^2 = 42\%$). Other statistical criteria, such as score test or Wald test, confirm the improvements. We achieve the best improvement when we include the investor and fund types. It implies that investor and fund characteristics explain a significant part of investor behaviours. In the rest of this section, we discuss only the third specification of the model.

3.4.1 Sector effects on surviving times

We begin our investigation by testing how investor's activity sector affect his/her redemption decision. We estimate the effects of six sectors dummies and we report their coefficients in Panel G of Table 3.5. We observe that coefficients of six sector dummies are all strongly significant. This result is consistent with our hypothesis **(H1)**. A major difference among those sectors is that they have different levels of liquidity needs, i.e. their business activities require cash outflows such as paying salaries, distributing dividends, or paying insurance claims. Furthermore, we find that sectors with more short-term liquidity needs have shorter surviving times and sectors without frequent liquidity needs have longer surviving times. It suggests that the liquidity need is a plausible explanation for these strong sector effects.

Banks, private banking clients and insurance companies are found to have long surviving times. The $Bank^{own}$ dummy has a strongly significant and negative beta of -0.271. According to Equation (2-4), being a bank investor decreases the redemption rate by $1 - e^{-0.271} = 23.74\%$. This result is consistent with preliminary data description in Section 4.2.1: bankers' surviving times are longer than average. Private banking clients have similar behaviours, they stay in mutual funds in relative long horizons: $Bank^{Private}$ dummy has a beta value of -0.135. The last long-term sector is the insurance company. The beta associated with the *Insurance* covariate is -0.23. These negative coefficients reveal that these

three sectors have a lower redemption risk than the average level.

The other three sectors, corporate firms, FoFs and mutual insurances, stay for shorter investing horizons. The beta of *Corporate* dummy is 0.246, which indicates that corporate firms' exit rate is $e^{0.246} - 1 = 0.27$ higher than the average level. The Mutual insurance companies have the similar surviving time, "Mutual" dummy's hazard ratio is $e^{0.213} = 1.23$. The last sector, FoF has shortest surviving time. "FoF" investors increase their closure rates by $e^{0.507} - 1 = 66.03\%$. As the full sample average surviving time is 330 days, our model predicts that a FoF's survival time is around $330 \times (1 - 0.66) = 112$ days, which is consistent with our preliminary survival analysis in Table 3.3 in Section 4.1.2. These three sectors all bring higher account closure risk than the benchmark level.

A highly possible interpretation of these significant sector effects is that different sectors have the different levels of liquidity needs. Banks and private banking clients do not have frequent short-term cash needs. Therefore, their capital in mutual funds could stay longer than others. In contrast, corporate firms consume many short-term liquidities in order to pay the salary, distribute dividends and repay short-term debts to their suppliers. These cash needs increase corporate firm's redemption rate. FoFs have the shortest surviving times. a plausible explanation is that FoFs receive daily redemptions of their own clients, which constraint their ability to invest in long horizons.

Moreover, we observe a notable distinction between two types of insurance companies. The normal insurance companies stay in funds longer than other investors, however, mutual insurance companies have shorter surviving times. We explain this difference by the fact that these two insurance companies issue different insurance products and hence have different levels of liquidity needs. Traditionally, a normal insurance company offers long-term products such as the life insurance. Hence it only needs to pay policy owners' claims in the long-term. On the contrary, the mutual insurance companies issue more shorter-term products, such as the healthcare insurance. Their liquidity needs are more "urgent", thus their surviving times in the mutual funds are shorten.

This test extends the study of [Johnson \(2004\)](#), where his data contain only retail investors and equity funds. However, our data possess a higher diversity. It is worthy to generalize Johnson's findings in our diversified sample for several reasons. First, Johnson's approach to forecast surviving times is difficult to implement for institutional funds since he use various retail-related co-variates, such as age, employment situation, and personal revenue (c.f. Table 1 and Table 4 in [Johnson \(2004\)](#)). We could not find these variables for institutional investors. Second, the literature suggests that institutional investors and retail investors have different behaviours in mutual funds (see, [Ben-Rephael et al. 2017](#)), especially during

the periods of massive redemptions (see, [Schmidt et al. 2016](#)). We highlight that the liquidity need is an important factor for investors' behaviour. This factor is related more to the institutional investors, which all have different liability constraints. That's why institutional investors often use ALM ("Asset Liability Management") approach to manage this liquidity constraint. Moreover, our analysis is more practical in use. Johnson's approach requires the manager to monitor his clients one by one and it would be time-consuming if a fund receives a large number of clients. In contrast, we only require that fund manager to classify his clients into different groups. This approach is more suitable than Johnson's for funds with the large client-base.

3.4.2 Time-varying rationality of fund clients

We continue our investigation by examining investors' rationality through the return-redemption relationship. A negative reaction to returns indicates that investors exhibit the loss aversion behavioural bias, whereas a positive relationship reveals that investors are rational. Previous studies (see, e.g., [Kamstra et al. \(2017\)](#)) document that fund investors have the time-varying behaviour. Accordingly, we assume that investors are more likely to exhibit the loss aversion in the last quarter of the year. Therefore, we isolate the fourth quarter of year from other periods in our test. We present our regression results in Panel A and Panel B of Table 3.5. We obtain a time-varying behaviour as we have expected in **(H2)**: investors are rational in first three quarters and the loss aversion presents in the fourth quarter.

In Panel A of Table 3.5, we first study the relationship between return variables and the closure risk during normal periods: from quarter 1 to quarter 3. The relationship is strongly significant and negative for the "individual holding returns". The covariate $R^{holding}$ has a hazard ratio of $e^{-0.69} = 0.501$. If there is a 1% increase in the holding return, the account closure rate would be decreased by $1\% \times 0.501 = 0.501\%$. It reveals that investors choose to stay longer when their individual holding return is better.

We then study impacts of "fund returns" on redemption rates. Unlike the individual holding return which are different among investors, the fund return is unique to all clients in the same fund. We test fund returns in different horizons, from daily to quarterly. We find that the effect of quarterly fund return, a coefficient of -0.768, is similar to what we find for $R^{holding}$. On the contrary, other shorter horizons' returns only have negligible effects on redemption rates: R^{daily} and R^{weekly} have insignificant betas and $R^{monthly}$'s beta is only weakly significant.

In the first three quarters of the year, there is a negative relationship between the

performance and the redemption probability. It suggests that investors behave rationally over these periods. They keep "winner" funds and sell "loser" funds. The relationship is identical either we use the holding return or the quarterly fund return. But short-term returns' coefficients are less significant. This evidence differs from Grinbatt and Keloharju (2001) where short-term returns have greater impacts on trade intensity than long-term returns in the equity market. A plausible explanation for this difference is that investors have different investing motivations in these two markets. In the equity market, investors focus more on the short-term price appreciation. Hence the short-term returns have larger impacts on trade decisions. However, in the mutual fund market, investors evaluate their investments in longer horizons. Therefore, only long-term return affect investors' decisions.

Panel B presents the return-redemption relation when we add control variables $Q4$, $Extreme - gain$, $Extreme - loss$, and Dec . The first three covariates aim to identify the loss aversion and we add the December month dummy to control for the tax effect. We find the strong evidence on the loss aversion bias during the last quarter. Both $Q4 \times R^{holding}$ and $Q4 \times R^{quarterly}$ have positive influences on the closure risk⁹. In the last quarter of the year, a 1 % better individual holding return increases client's exit rate by $1 - e^{0.694 \times 0.01} = 0.69\%$. For the fund return, the $Q4 \times R^{quarterly}$ has a hazard ratio of $e^{1.774} = 5.89$ which is also positive. These positive coefficients show that the rational investors begin to exhibit the loss aversion bias at the end of the year. Combing with the results in the previous paragraph, we confirm the **(H2)** which predicts a time-varying behaviour of investors.

Meanwhile, we do not observe the loss aversion in the December. During this month, the return-redemption relationship switch back to negative. $Dec \times R^{holding}$ and $Dec \times R^{quarterly}$ both have negative betas: -1.444 and -1.441. This result may be explained by the tax issues. In December people calculate the annual tax amount based on realized gains or loss. Investors avoid to realize profits because it triggers a large tax liability. Further, they might also sell funds at loss to decrease taxes¹⁰.

We find that an extreme fund return triggers also the loss aversion. When the fund performance is extremely good or bad, investors hesitate to sell. Hazard ratios for "Extreme-gain" and "Extreme-loss" are both negative (-0.566 and -0.945). It indicates that the redemption intensity decreases under extreme scenarios. However, investors behave differently in the last quarter of the year. "Extreme-gain \times Q4" has a strongly significant and positive beta of 0.982. A possible explanation about this coefficient is that once investors achieve the

⁹The covariate $Q4$ has a significant negative impact on the hazard rate. It reveals that investors redeem less during the fourth quarter.

¹⁰Selling a security at loss would offset the realized gains from other assets. This strategy is called "Tax harvest".

extremely good return they redeem immediately to take profit : they fear the return turns to bad. However, we obtain an insignificant beta of "Extreme-loss $\times Q4$ ". We interpret this "non-result" by the fact that investors do not know how to deal with extremely bad returns: neither redeem nor hold seems to be a good choice.

This investigation adds to [Johnson \(2004\)](#) by highlighting the dynamic side of the redemption risk. In Johnson's regression, he mainly uses the static covariates and he ignores the fact that investors sometimes shorten their surviving times when economic conditions evolve. Johnson includes only the daily return in his test and find a limited influence (cf. Table 5 in [Johnson \(2004\)](#)). We show that R^{daily} indeed has merely the negligible impact on the hazard rate, however, the impact increases strongly for returns measured in longer horizons, especially $R^{quarterly}$ and $R^{holding}$. Furthermore, it is highly possible that there are other dynamic factors which affect a investor's decision. Hence this introduces the notion of "pre-mature" redemption risk: investors exit the fund earlier than what we forecast, in response of various economic conditions.

3.4.3 Other control variables

We complete the test by adding several control variables. We find that previous findings about the liquidity needs and investor's rationality hold after controlling for economic factors such as: the time effect (Panel C), the asset class (Panel F), macroeconomic conditions (Panel E) and other transaction patterns (Panel D). Impacts of these covariates are straightforward hence we do not present all of them in the paper. We only explain details of two important groups of control variables: the asset class and the time effect. Panel F shows that the fund's asset class has strong predictive power on investor's surviving times. Besides, the test on time effects (Panel C) shows us the evolution of investor's behaviour during 20 years.

Panel F presents the coefficients of asset class. The effect of asset classes is consistent with their liquidity level. Investors in money market funds have much shorter surviving times than those in equity funds which have the benchmark surviving rate. The interpretation is straightforward: they use these funds to manage their cash needs. Hence their investing horizons are short. Fixed income investors stay in funds a little longer than the benchmark funds. We explain this observation by the fact that fixed income funds invest in relatively illiquid assets. Therefore, investors hold these funds in relatively longer horizons. At last, we see no significant impact of funds following an "absolute performance" strategy: investors stay in these funds as long as in the benchmark funds. This result might be explained by the

fact that the "absolute performance" is a fund class which contains diversified management strategies, in average, clients' surviving times in this category are as long as the benchmark level.

Panel C lists the coefficients for time effects. We separate the full sample into 4 periods: Period 1 is before 2001 crisis, Period 2 is between 2001 crisis and 2008 crisis, Period 3 is between 2008 crisis and 2011 crisis and Period 4 is after 2011 crisis but before the year 2015. We consider the period after 2015 as the benchmark and the model shows how investors traded differently before. We find negative sensitivities of the hazard rate to all 4 time dummies, which indicate that the hazard rate decreases and the surviving time lengthens. For example, the variable "Period 4" shows that, from 2008 to 2011, the redemption intensity is lower than the benchmark level, all else being equal. The exit risk decreases by $1 - e^{-0.382} = 31.75\%$. Furthermore, we find that, from period 4 to period 1, redemption rates become more smaller¹¹. These covariates show us a major tendency of investor's transaction during these 20 years: investors trade more frequently and their investing horizons have shorten. This evidence conform to the statistics in Table 1: we observe that investors are more and more active. They shorten their investing horizon since they are less "patient".

All three dummies for the crisis periods have significant negative impacts on the exit risk. Covariates *Crisis2001*, *Crisis2008* and *Crisis2011* have hazard ratios of 0.332 ($e^{-1.100}$), 0.540 ($e^{-0.615}$) and 0.523 ($e^{-0.647}$) respectively. Hazard ratios smaller than 1 indicate that investors redeem less during crisis. There are several plausible explanations for this result. First, many assets are undervalued during the crisis, thus investors consider it is not an optimal moment to redeem. An alternative explanation is that three crisis happen before our benchmark period (2015 to 2017), the investors were less active during those times.

In summary, this section examines how investors choose to redeem and exit the fund. Their redemption probabilities are sensitive to both static and dynamic covariates. Investors' reaction to dynamic covariates indicates that the surviving times might evolve after account opening. Those who shorten their surviving times bring an additional liquidity risk to the fund. We refer this risk as the "**pre-mature closure risk**": investors exit fund earlier than expected. The pre-mature closure forces the fund manager to sell illiquid assets with a liquidity cost. To better monitor this liquidity problem, we ask the question which investors bring the higher "pre-mature" closure risk and we address this question in the next section.

¹¹Beta values decrease: -0.382, -0.755, -0.997 and -1.461. → Hazard ratios decrease accordingly: 0.682, 0.470, 0.368 and 0.232. Smaller hazard ratios stand for lower redemption rates.

3.5 Which investors bring the higher "pre-mature" redemption risk

In this section, we intend to identify which group of investors bring the higher "pre-mature" closure risk. To this end, we split the full investors sample into two sub-samples with long-term (**LT**) investors on one hand and short-term (**ST**) investors on the other one. We then apply the same survival model (Model (1), Specification (3) of Table 3.5) to two sub-samples and we compare the sub-samples' sensitivities to dynamic covariates. According to our hypothesis (**H3**), we are expected to observe that LT clients have higher sensitivities to "value" factors such as investing returns and economic conditions. Moreover, (**H3**) also suggests that LT sub-sample test would reach a higher R-square.

We choose one year, which equals to 250 trading days, or 360 calendar days, as a natural threshold to classify LT and ST investors. We further consider this threshold is suitable only for moderately liquid fund class such as "equity", "diversified" and "absolute performance". For fund classes which manage relatively illiquid assets, such as fixed income, convertible debts and real asset, we increase this threshold by 100 days. On the contrary, for money market funds, since they invest in highly liquid securities, we reduce this threshold to 60 days. Investors who stay in funds longer than the threshold horizon are seen as LT investors, others are thought as ST investors¹².

Table 3.6 presents the hazard rate estimations for two sub-samples. We report only the covariates which show the divergences between LT and ST investors. The first column repeats the results of the full sample in Section 3.4 for comparison purpose. The middle column lists the covariate coefficients of the LT sub-sample and the last column shows the ST sensitivities. There are 4 282 observations in the LT sub-sample and 3 630 in ST observations. Accounts which still stay open after Mars 2017 are censored.

The R-square is 0.431 in LT sub-sample test. This ratio is twice large as the ST sub-sample one ,0.223. The difference in R-square confirms one implication of (**H3**): ST investors are more "noisy", the model fits badly the ST sub-sample. We explain this finding by the fact that shortly stayed clients redeem the fund share to satisfy their urgent cash needs, which are random and difficult to forecast at the fund level. So the model has less predictive power for the ST sub-sample.

Panel A shows the micro-economic¹³ covariates' sensitivities. We observe that LT in-

¹²There are alternative methods to classify investors. We check the robustness of our test by using two other criteria in Section 3.6.

¹³In Table 3.6, "micro-covariates" stand for the variables which relate to the fund. "Macro-covariates" are factors which link to the macroeconomic conditions.

vestors are more sensitive to these variables. LT sub-sample has a significant sensitivity of -0.580 (hazard ratio equals to $= e^{-0.580} = 0.559$) to the holding returns, whereas the ST investors have a weak and insignificant sensitivity. Two sub-samples have both negative significant sensitivities to the quarterly fund return. However, this return-redemption relationship is stronger in LT sub-sample. $R^{quarterly}$ has a hazard ratio of $e^{-1.48} = 0.227$, which is nearly one third of ST's hazard ratio ($e^{-0.362} = 0.692$)¹⁴. It suggests that LT investors are more reactive to the fund performance than ST investors. This evidence confirms our argument that LT investors are value-motivated and they make more efforts to monitor the fund performance.

¹⁴When the hazard ratio is lower than one, a lower hazard ratio represents a stronger multiplicative effect to the hazard rate.

Table 3.6: Cox proportional hazard rate model: sub-sample tests comparison

	<i>Conditional account closure probability:</i>		
	$h_i(t) = \lim_{\Delta t \rightarrow 0} \frac{[(t \leq T_i < t + \Delta t) T_i \geq t]}{\Delta t}$		
	(Full)	(LT)	(ST)
A: Micro covariates:			
$R^{holding}$	-0.690*** (0.124)	-0.580*** (0.073)	-0.009 (0.009)
R^{daily}	-0.271 (1.054)	-1.17 (1.239)	0.766 (1.817)
$R^{quarterly}$	-0.768*** (0.123)	-1.480*** (0.341)	-0.362** (0.130)
Dec	-0.131*** (0.043)	0.347*** (0.048)	-0.175 (0.005)
Variance	0.630*** (0.124)	-4.818*** (1.219)	0.228* (0.126)
B: Macro covariates:			
Short term rate	0.278*** (0.010)	0.478*** (0.017)	0.036* (0.019)
Market return	-0.0003*** (0.00003)	-0.00005*** (0.00004)	-0.0001** (0.00004)
C: behavioural bias:			
$R^{holding} \times Q4$	0.694*** (0.123)	0.498*** (0.132)	-0.279 (0.543)
Extreme-gain $\times Q4$	0.982*** (0.199)	0.808*** (0.224)	0.171 (0.493)
Observations	7,912	4,282	3,630
R^2	0.420	0.431	0.223

Note:

*p<0.1; **p<0.05; ***p<0.01

This table presents the comparison between LT and ST sub-sample tests. We estimate the semi-parametric hazard rates in two sub-samples by the same model as the previous section. We present only the coefficients which have significant divergences between 2 sub-samples and other covariates are not presented. As the previous section, there are 39 covariates in the test. The first column "Full" repeats the estimation of the full sample estimation for comparison purpose. The column "LT" shows the results of LT sub-sample, and the column "ST" presents ST group. Panel A shows microeconomic covariates. Panel B presents macroeconomic variables and Panel C reports the results for investors' behavioural bias.

We further find that LT investors redeem more in December. The covariate "Dec" has a sensitivity of 0.347 in LT sub-sample. In contrast, December has no significant effect on ST investors. This positive December sensitivities indicates that LT investors have more tax-motivated redemptions than ST ones.

Surprisingly, we observe that LT sub-sample has a negative reaction to the fund variance (-4.818). It implies that the increase in return variance decrease the redemption intensity. On the contrary, this sensitivity is positive in ST sub-sample. A plausible explanation is that LT investors exhibit a larger behavioural bias. When the fund return is less understandable, LT investors hesitate to sell their fund share. On the other side, short-term investors react more quickly to the increase of variance (positive coefficient 0.228). This reaction is more rational than LT's.

Panel C confirms our assumption on the behavioural bias. Covariates $R^{holding} \times Q4$ and $Extreme - gain \times Q4$ both show how LT and ST investors diverge in the fourth quarter. During this period, LT investors react positively to the holding return and the extreme-gain. 1% better holding return in Q4 makes LT investors to increase the rate of selling by $e^{0.498 \times 0.01} - 1 = 0.49\%$. A similar effect exists on the extreme-gain in Q4. The multiplicative effect of extreme good return is even larger: $e^{0.808} - 1 = 1.24$. However, we do not observe the loss aversion in the ST sub-sample. Lastly, Panel B shows that LT investors are also more sensitive to the macro-economic variables.

Together these results support our hypothesis ((**H3**)). The LT investors are value-focused. They monitor more the fund performance. Their behaviour is more explainable. In contrast, ST investors behave more randomly. The survival model loses the ability to understand their trades. Furthermore, we find that LT investors exhibit higher loss aversion bias. These results suggest that LT investors bring the higher "pre-mature" closure risk to the fund. When the return or economic conditions are bad, the fund manager loses LT investors at first.

This sub-sample test is in line with [Johnson \(2004\)](#). In his previous study, Johnson examined the consequence of investor behaviour. He calculate and compares the liquidity cost of LT and ST investors' redemption to identify who generate a higher liquidity cost to the fund manager. We make a further step to test which investors are more likely to redeem. Our finding joins in Johnson's discussion on mutual fund's liquidity insurance function. As banks, mutual funds provide the liquidity insurance by pooling investors together. This insurance requires a Diamond-Dybvig equilibrium: investors' liquidity demand should be equal to their liquidity offer (see, [Diamond and Dybvig 1983](#)). Johnson highlights that the equilibrium would fail since there is an asymmetric liquidity consumption. LT investors

provide more than what they consume and ST investors consume more than what they offer. Our findings imply that the Diamond-Dybvig equilibrium could still be reached for two reasons: (1) there is a pre-mature redemption risk. Investors' liquidity consumption (surviving time) could not be perfectly predicted; (2) LT investors bring a higher pre-mature redemption risk. Therefore, LT investors cost more liquidity than what Johnson indicates.

3.6 Robustness checks

The methods to measure the covariates might have impacts on the model estimation. In this section, we modify the measures to check the robustness of our results. We focus more on tests of **(H1)** and **(H3)**. Concerning **(H2)**, since there is already a large number of studies which have tested different return measures (see: Barber et al. 2016; Berk and Binsbergen 2016; Ivkovich and Weisbenner 2008). We do not retest them in our investigation.

Previous tests for **(H1)** use the full sample which includes all fund types. The result might be biased since client sectors are not equally distributed in different funds. Here is a simplified example: suppose there are only 2 funds A and B and two investors sectors m and n. Fund A is liquid and Fund B is illiquid. Sector m has short-term cash needs and Sector n has not. Fund A contains only Sector m and Fund B contains only Sector n. If we test the full sample, we may find evidences to confirm **(H1)**: there a positive coefficient on sector m dummy which indicates a higher redemption risk. However, this evidence might be generated only by the fund asset class: fund A is more liquid than fund B. This bias motivates our test for asset class sub-samples.

We choose 5 most representative asset classes and we estimate Specification (3) of Model (1) in each sub-sample (c.f. Table 3.5, column 3). We ignore other fund types because they have less than 500 observations, hence the estimation is less reliable. Since we examine already in asset class sub-samples, we drop the asset class dummies from Specification (3). We present only investor sector dummies and other variables are hidden. We report each asset class in one column: equity, money market, fixed income, convertible debt ("coco") and absolute performance. If our findings regard to **(H1)** are robust, we are expected to observe the similar covariate sensitivities as the previous full sample test in Section 4.2.

Table 3.7 presents our robustness check for the first hypothesis. Signs of most coefficients are as same as the full sample test in Section 4.1: sectors with short-term liquidity needs shorten their surviving times and hence have larger redemption rate. The first column shows the coefficients of the equity sub-sample. We observe that first three sectors have negative impacts on the hazard rate and they are sectors without short-term liquidity needs.

"Firm" and "Mutual" shorten their surviving times and these 2 sectors have more short-term liquidity needs. "FoF" dummy has an insignificant sensitivity. Money market sub-sample provide similar results. "*Bank^{own}*" has a strong negative sensitivities -0.909. On the contrary, "Mutual" and "FoF" have both positive sensitivities. These coefficients are broadly consistent with what we find in Section 3.3. In the other 3 sub-samples, estimation coefficients are less significant. This insensitivity might be explained by the fact that the number of observations in these sub-samples is small. Overall, all significant sensitivities confirm our hypothesis (**H1**): fund clients with urgent cash needs have a higher rate of redemption ¹⁵.

Table 3.7: Survival analysis for asset class sub-samples

<i>Conditional account closure probability:</i>					
$h_i(t) = \lim_{\Delta t \rightarrow 0} \frac{[(t \leq T_i < t + \Delta t) T_i \geq t]}{\Delta t}$					
	Equity	Money market	Fixed income	Coco	Absolute
Investor sector					
<i>Bank^{own}</i>	-0.170** (0.080)	-0.909*** (0.339)	-0.209 (0.189)	-0.188 (0.201)	-0.155 (0.186)
<i>Bank^{private}</i>	-0.315*** (0.076)	0.374*** (0.137)	-0.669*** (0.211)	-0.292 (0.190)	-0.690* (0.369)
Insurance	-0.454*** (0.144)	-0.014 (0.143)	0.268 (0.231)	-0.167 (0.250)	-0.484 (0.393)
Firm	0.729*** (0.179)	0.148 (0.132)	0.620*** (0.158)	0.531* (0.318)	0.387 (0.278)
Mutual	0.302*** (0.059)	0.347*** (0.070)	0.390*** (0.086)	0.654*** (0.103)	0.242* (0.133)
FoF	0.027 (0.111)	0.464*** (0.076)	0.053 (0.138)	0.655*** (0.167)	0.346 (0.245)
Observations	3,052	2,149	1,271	918	600
R ²	0.362	0.188	0.313	0.444	0.404

Note:

*p<0.1; **p<0.05; ***p<0.01

In this table we present the semi-parametric hazard rate estimation for asset class sub-samples. We apply Specification (3) of Model (1) to the 5 asset class sub-samples. In this test, we drop the asset class dummies and the rest covariates of Specification (3) are kept. They are: performance variables, return controls, time controls, other controls, macroeconomic controls and investor sector dummies. We report only the investor sector dummies and other covariates are hidden. Each column lists the coefficients of one asset class: equity, money market, fixed income, convertible debt (coco) and absolute performance.

In Section 5.1 we divide LT and ST investors by a threshold value which adapts to fund classes. This choice of threshold might have an effect on the estimation results. To check the robustness of our findings of (**H3**), we use 2 alternative thresholds to separate LT and ST sub-samples. At first, we use the median surviving time as the threshold. With this threshold, the numbers of observations in each sub-sample are almost equal. Then we use another exogenous threshold: for equity funds, investors stay longer than 250 days are considered as LT and less than 200 days are ST. In this approach, we drop investors

¹⁵There is only one exception: the "*Bank^{private}*" has a positive impact on the money market sub-sample. We accept this deviation.

who have surviving times between 200 to 250 days. We adapt these values to other asset classes (60-30 for money market, 300-250 for fixed income). If the results in Section 5.1 are robust, the divergence between 2 sub-samples should be hold when we change the separation threshold.

Table 3.8 presents robustness tests for our third hypothesis. We confirm our findings about investors' heterogeneities. When we use 2 alternative thresholds to classify sub-samples, the divergences hold. Panel A shows, the results when we use the median threshold and Panel B reports the second criteria. We keep the same the hazard specification as in Section 4. However, we report only some key variables which exhibit difference between 2 sub-samples. Other covariates are hidden. Both 2 robustness tests confirm the **(H3)** by following evidences: (1) LT sub-sample has a higher R-square; (2) LT investors are more sensitive to dynamic covariates, especially the holding performance; (3) LT investors have a larger positive sensitivity to "Dec", which shows a larger number of tax-related transactions; (4) LT investors are more sensitive to the return variance. (5) LT investors are more sensitive to macro-economic variables; (6) LT investors exhibit the loss aversion bias at the end of the year, whereas ST ones do not;

In summary, liquidity need stays to be an important determinant for the survival rate when we test the model in asset class sub-samples. These tests control the impact of the asset liquidity. We confirm the significant effects of client sector dummies. Concerning to the LT-ST heterogeneities, we use different thresholds to divide the sample. The differences between LT and ST investors are held. This evidence confirms our third hypothesis.

Table 3.8: Long- and short-term investors' comparison with alternative criteria

Panel A							
LT							
$R^{holding}$	Dec	Variance	Short term rate	Rate spread	Market return	Extreme-gain \times Q4	$R^{holding} \times$ Q4
-0.405**	0.337***	-3.338***	0.476***	-0.943**	0.00005***	0.451***	-0.004
Observations	3 950	R^2	0.348				
ST							
$R^{holding}$	Dec	Variance	Short term rate	Rate spread	Market return	Extreme-gain \times Q4	$R^{holding} \times$ Q4
-0.013	-0.101*	0.122	0.004**	0.016	-0.0001*	-1.212	0.09*
Observations	3 962	R^2	0.14				
Panel B							
LT							
$R^{holding}$	Dec	Variance	Short term rate	Rate spread	Market return	Extreme-gain \times Q4	$R^{holding} \times$ Q4
-0.586***	0.333***	-2.879**	0.456***	-0.895**	0.00005***	0.611***	0.451***
Observations	4 298	R^2	-1.212				
ST							
$R^{holding}$	Dec	Variance	Short term rate	Rate spread	Market return	Extreme-gain \times Q4	$R^{holding} \times$ Q4
-0.268*	-0.145*	-0.211	0.030*	0.060	0.00004	0.767	0.611
Observations	3 280	R^2	0.697				
<i>Note:</i>		*p<0.1; **p<0.05; ***p<0.01					

In this table we use the semi-parametric approach to estimate account closure hazard rates in LT and ST sub-sample. We use different thresholds to divide the LT and ST investors to check the robustness of previous findings in Section 4.5. Each panel presents one threshold. Panel A presents results when we use median surviving time as separation criteria. In this test, the numbers of observation in each sub-sample are almost equal. In Panel B we use another exogenous threshold. Investors stay longer than 250 days are considered as LT and less than 200 days are ST. We keep the full specification (3) of Model (1). The tables only show some key covariates, other variables are hidden.

Conclusion

Our study shows why investors stay in different horizons in mutual funds. On one hand, we find evidence that fund client's in the same category have similar investing horizons. On the other hand, our tests highlight that investors might shorten these investing horizons when economic conditions deteriorate: they exit funds pre-maturely. Further, we find that long-term investors are more likely to exit in this pre-mature way. To obtain the results above, our investigation relies strongly on the individual transaction information of fund investors. This paper indicates that fund investors are not homogeneous and we suggest that the fund manager should monitor the redemption risk at the individual level.

According to our findings, the fund company faces an asymmetric liquidity cost. All investors pay similar management fee and receive the same right to redeem freely their fund shares. However, they consume fund's liquidity differently since they stay in funds for different horizons. Shortly-stayed investors cost relatively more liquidity than long-time-stayed investors. Furthermore, the fund company pays notable costs to maintain a large client-base: recording clients' profile, recruiting additional personnel for "client services". Fund companies also provide "new client bonus" to attract new investors. Therefore, receiving too much shortly-stayed investors are not profitable.

We propose several solutions for the problem above. The simplest solution is to charge different management fees or to offer distinct "new client bonus". Clients as a life insurance company would stay in fund for a long horizon. We should charge a lower AUM fee and offer the higher bonus to new entries. In contrast, a FoF investor might have a surviving time less than 2 quarters. The fund company could pay him a low "bonus". However, this solution is difficult to implement in practice, since investors require the equal rights in general.

Alternatively, we might consider the model's estimation as a benchmark and evaluate client's "quality". For example, the model forecasts an investing horizon of 110 days for a corporate firm. If corporate client A stays in fund for longer than 110 days, he is a "good" client. Hence we could award him by decreasing his AUM fee. Another corporate client B, if he exits the fund before 110 days, he is a costly client. We might charge him a higher subscription fee for the next time. This solution is more piratical than the previous one. Indeed, the government uses a similar policy to encourage long-term investments: decreasing the tax if the investment is held for a long period.

Besides, our examination of investors' heterogeneity helps to understand the fund run scenario, where the manager loses his clients in a sudden and massive way. We show that LT investors are more reactive to economic conditions and they bring the higher pre-mature

risk. This finding implies that during a crisis period, the fund manager loses more LT investors. Therefore, fund's average liability duration should be shortened during a run scenario. Accordingly, we suggest that the fund manager should sell long-term and illiquid securities to decrease the asset's duration too.

Finally, this survival analysis suggests regulators to introduce some liquidity-protection policies for mutual funds. We can "borrow" some liquidity provisions from hedge funds to enhance mutual fund's liquidity. For instance, the "lock-up" period prevents pre-mature redemption. The fund company could propose two types of AUM fee for investors: (1) a lower management fee for investors who can respect a "lock-up" period: do not redeem before a certain time; (2) a higher management fee for normal investors who can redeem freely. Alternatively, mutual funds could also adopt the "Gate Provision": setting a threshold for free redemption. If the investor wishes withdraw more than the threshold value, the fund company would charge a higher redemption load.

Chapter 4

Rebalance Activity Of Individual Mutual Fund Investors

This article studies how individual mutual fund investors rebalance their portfolio on a daily basis. We collect a unique data set recording the purchases and redemptions of all investors in a no-load French mutual fund company over a two-year period. We employ an "Ordered Response Model" to identify the determinants of investors' trade intensity. This method explains how investors choose among purchase, hold or redeem their fund shares. We consider not only the fund being traded but also other funds that the same investor holds in parallel. We find that investors adopt a "constant-mix" rebalance strategy to add positions to funds which have under-performed in the short-term horizon. Furthermore, investors have a tendency to switch to funds with higher liquidity and lower risk. Our findings stay robust after controlling for investors' behavioural bias and the majority of results are stable during different quarters of the year.

Key words: rebalance, mutual funds, institutional investors, liquidity, micro-transaction, constant-mix

Introduction

Until the end of 2016, mutual funds' total Asset Under Management (AUM) on a global scale has reached 40.4 trillion dollars, including 22% of total household's financial assets in the US¹. It makes the mutual fund one of the most popular investing vehicles. A primary reason for this popularity is the fact that the delegate fund manager can achieve a superior performance with his expertise. Since fund investors rely much on the fund manager, it raises a natural question: "do mutual fund investors need to be active in their fund investing?" Some literature suggests that investors should better follow a passive "buy-and-hold" strategy since frequent purchases/redemptions disturb the fund manager (e.g., [Akbas et al. 2015](#); [Bailey et al. 2011](#); [Darolles and Roussellet 2018](#)). On the contrary, others argue that the active reallocation of fund portfolio improves investors' overall return ([Dahlquist et al. 2017](#)). Further study of this activism is difficult since it requires the micro-transaction data of fund investors, which are not publicly available.

We record manually purchases and redemptions of all investors in one fund company over a two-year period. It allows us to monitor each individual trade decision on a daily basis, whereas traditional mutual fund databases, such as Morning-star Direct, LIPPER, CRSP Mutual Fund, provide only aggregated flow information at the fund level at a monthly frequency. We observe that fund clients in our sample purchase and redeem frequently. It seems that they do not follow the simple "buy-and-hold" strategy in fund investing. Instead, they continuously adjust their fund shares. Furthermore, more than 70% investors in our sample hold more than one fund. These multiple holdings often contain different asset classes. It implies that fund investors possess the multiple-assets "portfolio". Since a large part of investors in this sample are institutional ones who possess often the experiences and skills in investing, it is highly possible that they follow an asset allocation. Therefore, we ask the question whether investors treat their fund holdings as a whole portfolio or they only monitor funds in an isolating manner? Whether their asset allocation static or dynamic? And what factors drive their frequent transactions?

In this paper, we study the individual fund investor's rebalance activity via our hand-collected micro-transaction data set. We use the "Ordered Response Model" to analyse trade decisions among "Purchase", "Hold" and "Redemption". We first investigate investors' reaction to the fund performance. This test gains insights for which rebalance method they use. Then, we examine the mechanism of the rebalance activity by identifying how investors switch from one fund to the other. We compare each fund's return, risk and liquidity to

¹C.f. [ICI 2017](#)

other funds held by the same investor. Finally, we investigate whether investors' behaviour is stable over different periods of the year.

We establish several findings in this paper. We first calculate fund returns in different horizons and we demonstrate that investors have different reactions to short-term and to long-term fund returns. We consider a return measured by a horizon longer than 50 days as the long-term return, otherwise as the short-term one. We find that a better long-term return makes clients more possible to purchase. However, when we use short-term returns, this relationship is inverted: investors begin to purchase under-performed funds and redeem over-performed ones. We explain this behaviour by the fact that investors select "winner" funds to invest in a long-term perspective but they repurchase short-term's under-performed funds to keep a constant asset allocation. This finding is robust after we control for investors' behavioural bias. Secondly, we observe that investors have a global "portfolio" view to treat all their fund holdings together. An investor compares a fund return, risk and liquidity level to other funds that he holds in parallel, and these relative qualities impact his trade decision. Investors have a tendency to add positions to the funds with lower return, higher liquidity and lower risk. Finally, we confirm that the majority of investors' behaviours are stable during four quarters of the year.

Our article relates to three strands of literature. First, we continue the ongoing study on the micro-transaction data. Previous research seeks to identify the determinants of trade decision in different markets. [Grinbatt and Keloharju \(2001\)](#) focus on the equity market. They find that the historical price pattern has strong influence over investor's propensity to trade. [Ashcraft and Duffie \(2007\)](#) bring our attention to the "repo" market. They show how the over-night repurchase loans of US' federal funds are affected by various macro-economic conditions. [Iyer and Puri \(2012\)](#) investigate what factors trigger bank clients' withdrawals during a "bank run". Lastly, fewer works examine the mutual fund investor's trade behaviour (see, e.g., [Bailey et al. 2011](#); [Chang et al. 2016](#)). Fund investors are being found to exhibit the loss aversion behavioural bias. Although continuous efforts have been made to understand the micro-transactions, previous authors study in an isolate approach: they analyse security one by one. Our contribution is that we consider the full portfolio that investors possess and we analyse how securities (mutual funds in our case) held by one investor affect each other.

Second, our research brings new evidence to mutual fund's flow-return relationship. Previous literature intends to recognize how mutual fund flows react to the fund performance. Theoretically, there should be a positive reaction (see, [Berk and Green 2004](#)): a good return increases the inflow and a bad return increases the outflow. However, empirical studies

highlight that this relationship is more complicate in reality. People find that mutual funds in different asset classes, such as equity funds (Frazzini and Lamont 2008), corporate debt funds (Goldstein et al. 2017) and real estate funds (Fecht and Wedow 2014) have different shapes of flow-return relationship. Furthermore, Johnson (2010) indicates that this relation is not uniform over time and Johnson (2004) show it depends also on investors' characteristics. To our knowledge, we are the first to distinct fund flow's sensitivities to the long-term and short-term returns. We point out these two return measures relate to two different investing decisions: fund selection and portfolio rebalance. We show that the flow-return relationship is modified according to the horizon of the return measure.

Finally, we add to the continuous debate on how to evaluate open-end fund's performance. The simplest approach is using the factor model, such as the Fama-French 3-factor model, to adjust fund risk exposures (e.g., Barber et al. 2016; Brown and Wu 2016). However, for funds following special strategies or investing in exotic assets, we should consider additional risk factors such as: skewness risk, fat-tail risk or option risk (Agarwal et al. 2015). Berk and Binsbergen (2016) compare the performances of different factor-models and indicate that the market risk-adjusted model has the highest predictability on fund flows. It implies that fund investors are most likely to use this method to evaluate fund returns. We propose another alternative to assess a fund's quality by benchmarking its return on a unique peer group: all funds being held by the same client. This method would not be ease to implement since it requires insensitive data and calculations, however, it suggests that we should consider effects of clients on a fund. The same fund might be good for some investors but bad for others, depending on what portfolio investors hold and what is their allocation strategy.

The rest of this paper is structured as follows. Section 1 describes the data-base and presents the summary statistics for the fund investor trade intensity. Section 2 introduces our empirical method, the "Ordered Response Model". Section 3 and 4 studies how investors rebalance their portfolio. Then we discuss and conclude in the end.

4.1 Data description & summary statistics

4.1.1 Data-base description

We collect a novel dataset which records investors' micro-transactions from one French non-load fund company. The company charges a standard management fee which is proportional to one's fund share. We call it as "AUM fee". Additionally, the fund manager earns an

incentive bonus if his performance exceeds the benchmark. However, there is no load to pay when investors purchase or redeem, the transactions are entirely free. The fund company was created in 1988 and it has managed more than 100 funds ever since. In the current year, its total AUM reaches 60 billion EUR. It is affiliated to 2 insurance companies. These two parent companies are important sources of its capital inflows: 20 % of the global AUM. However, two parent companies leave the fund managers totally independent to make their own investing decisions.

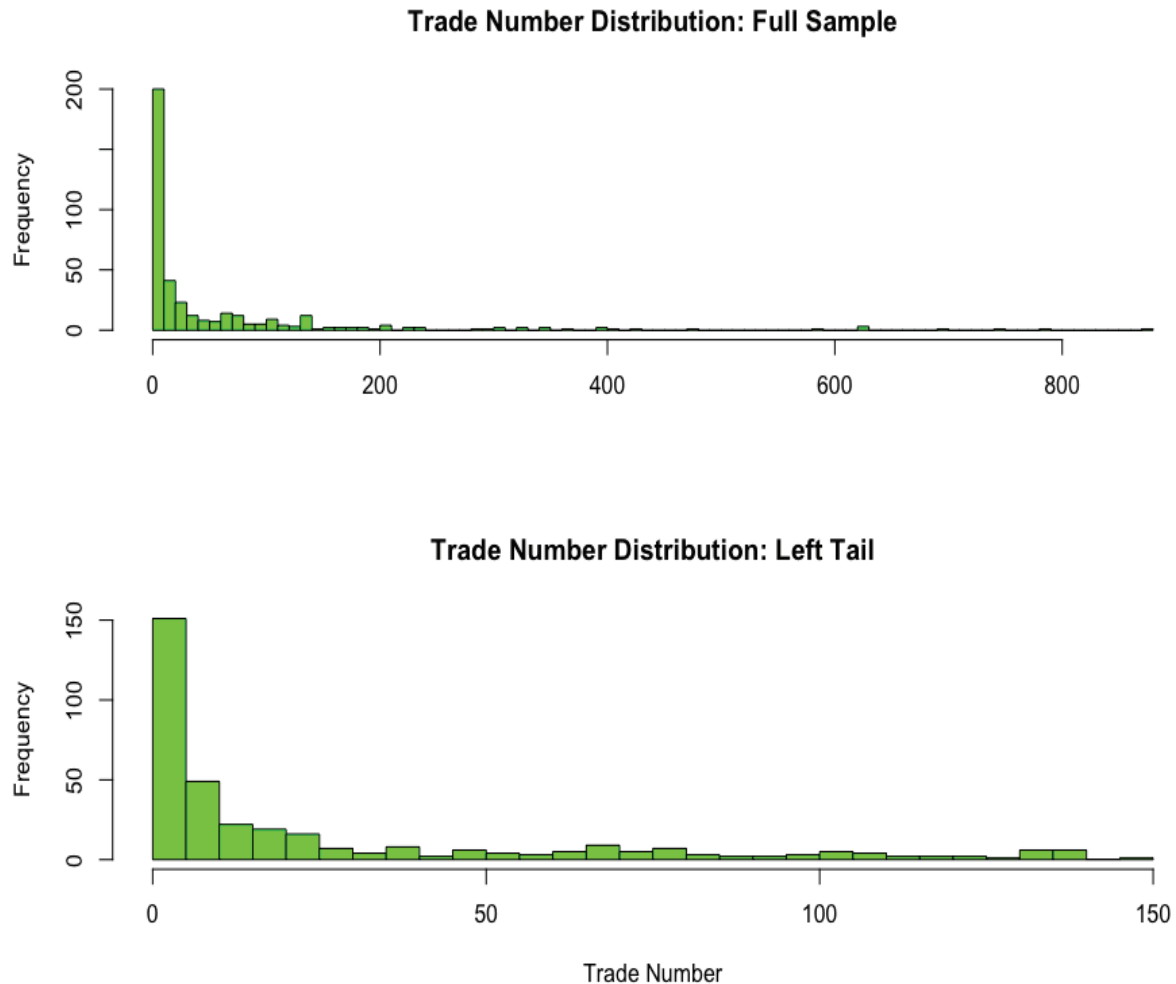
Our study focuses a two-year period, from 2015 to 2016. No new fund is created during this period and the market environment is stable. In our raw data file, transactions are recorded on a daily basis with elementary information such as: (1) Date of transaction; (2) Transaction price; (3) Client identity; (4) Client category; (5) fund ISIN number; (6) fund's asset class; (7) Amount of transaction. This companies' client base is diversified since investors belong to a large range of different activity sectors such as corporate firms, family offices, banks, insurance companies or pension funds. Most of investors are based in France and all transactions are executed in euro. There are 558 investor accounts in total. Investors hold a large range of fund classes, including equity, real estate, diversified, money market, fixed income, convertible debts and absolute performance. The sample contains 82 fund-ISINs with each of them represents one fund share class. In general, one mutual fund contains two to four share classes which charge different fees to different investors. We use "Datastream" as a secondary data source to download macroeconomic variables such as the MSCI index.

4.1.2 Summary statistics

We count the numbers of total trades, either purchase or redemption, executed by each investor during the sample period and present their distribution in Figure 4.1. From the left tail of this distribution, we find that a large part of investors trade more than twice. Executing two trades implies that the investor follows the "buy-and-hold" approach: under this strategy, investors buy a fund and do not execute any intermediary trade until the final redemption. When the trade number increases, the observation frequency decreases dramatically. On the right side of the distribution, only a few investors make more than 500 trades. It suggests that most investors do not trade at each day. Combining these observations, we deduce that majority of investors neither trade funds at a extremely high frequency, such as trading on a daily basis, nor just follow the simple buy-and-hold strategy to trade only twice. This behaviour differs from previous studies, which assume that the

"buy-and-hold" strategy is adopted (see, [Ivkovich and Weisbenner 2008](#); [Johnson 2004](#)).

Figure 4.1: Distribution of trade number of individual fund investors



This figure shows the trade number distribution of all individual investors in our sample. We count the number of trades executed by each investor over a two-year period, from 2015 to 2016. The X-axis is the number of the trades executed and Y-axis is the observation frequency. We present only the aggregated number, where purchases and redemptions are mixed. The graphic on the top presents distribution of the full sample. The bottom graphic is the zoom of the left tail of the distribution.

There are several possible explanations for the relatively high trade intensities observed at above. One is that investors use mutual funds as the cash management tools and trade in the response of their liquidity needs. This explanation is suitable only for the money market funds since these funds invest in highly liquid short-term securities. However, in our sample, money market funds represent only for less than 10 % of the full sample: 8 funds out of 82 in total. For fixed income funds or real estate funds, this explanation is less convincing since these funds invest in relatively illiquid assets so investors would not use

them to manage cash. The second possibility is that fund investors have the target asset allocation and they rebalance their fund shares dynamically toward it. This hypothesis is consistent with the fact that many investors in our sample are institutional investors, who are more "intelligent" than retail ones and are likely to follow some sophistic investment strategies (see, [Schmidt et al. 2016](#)). Indeed, previously used data-sets contain only retail investors which are less skilful in general (see, [Bailey et al. 2011](#)). That is a plausible reason why their investors follow the sample "buy-and-hold" method. Therefore, our data offer an opportunity to investigate the more "complicate" trade behaviours. The institutional investors in our sample are highly possible to adopt sophistic investing and trading strategies than retail investors in the previous literature. Before moving to next step, we first need to examine investors' holding situations and their trade intensities.

Table 4.1 presents the holding situation of individual investors. Our sample contains 558 investor from a diversity of activity sectors, including both retail and institutional categories. In Panel A we show the number of funds held by each investor. There are 185 clients invest only in 1 fund. We drop them out of our test since they have no possibility to rebalance. It leaves $558 - 185 = 373$ investors who hold the multiple funds portfolio. Then, we examine whether investors' fund holding covers different asset classes. The fund company follows the French financial regulator AMF's classification to categorize funds into eight classes: real estate, equity, diversified, money market, convertible assets, absolute return, fixed income and others. Panel B shows the number of asset classes each client invests. There are 231 investors who hold only one asset class and the rest of clients invest all in multiple asset classes. It is possible that many investors constructs their fund portfolio with a target asset allocation. A limit of AMF fund classification is that it does not consider the liquidity of the fund portfolio. For instance, a large-cap equity fund is more liquid than a small-cap equity fund, but they are all labelled as equity funds. Therefore, we create a liquidity level to attribute to each fund. We decide a fund's liquidity level according to its recommended investing horizon. This information is given by the fund manager based on the portfolio's liquidity level and turnover. The fund manager discloses this information in fund's annual report to all investors. In our sample, the recommended investing horizons vary from 6 months to 72 months. We consider: (1) the funds with a recommended investing horizon longer than 36 months as level 1, which is the least liquid level; (2) funds with horizons between 12 months and 36 months are classified as level 2; (3) funds with a recommended horizon shorter than 12 months have the liquidity level 3, which is the most liquid category. Panel C indicates that 253 investors hold funds with the same liquidity level. 264 clients invest in funds with 2 different liquidity levels and 141 investors trade funds in all 3 liquidity

levels.

Table 4.1: Investor holding situation

Panel A: fund holding										
Number of Funds Hold	1	2	3	4	5	6	7	8	More	Total
Number of Clients	185	86	58	50	30	20	25	12	93	558
Panel B: asset class holding										
Number of Asset classes	1	2	3	4	5	6	7	8		
Number of Clients	231	104	75	69	49	22	7	1		
Panel C: liquidity holding										
Number of Liquidity Levels	1	2	3							
Number of Clients	253	164	141							

This table shows the holding situation of individual investors. In Panel A, we present how many funds that each investor holds. "Number of Funds Hold" is the number of funds an investor hold and "Number of Clients" indicates how many investors hold this number. Panel B presents how many fund classes that investors hold and Panel C shows that investors' fund holdings across how many liquidity levels. We class funds' liquidity levels from 1 to 3 according to a fund's recommended investing horizon.

For the 373 investors who possess multiple fund holdings, we study the statistics for their trade intensity. We consider each observation as a trade decision that an individual investor makes for one fund of his holdings at a single day. Each observation has four possible outcomes: (1) "Hold", investors do nothing; (2) "Purchase", investors add fund's position; (3) "Redemption", investors cut fund's position; (4) "Closure", investors exit the fund. Finally, We obtain 1 163 519 observations at this "investor-fund-date" level. We present their summary statistics in Table 4.2. The first raw provides statistics for the full sample. The purchase and redemption probabilities are 1.16% and 1.28%, respectively. In 0.12% of situations, investors exit the fund. We then show the sub-sample statistics for different investor categories in Panel A and different fund classes in Panel B. "Platform" represents the investors who purchase funds through a brokerage platform service. With a purchase intensity of 1.76% and a redemption intensity of 2.76%, they are the most active investors among all. In other activity sectors, investors behave more similarly. In contrast, trade intensities in different fund classes are more heterogeneous. Money market funds receive the most frequent transactions. Both purchase and redemption intensities have a high level around 4%. It suggests that each investors make $(4\% + 4\%) \times 250 = 20$ trades on a annually basis. Equity and diversified funds have also active investors with trade rates around 1%. Clients in other fund classes are less active. Since the probability of "closure" is lower than 1% in most of the time, we merge "Closure" and "Redemption" for the our test in the rest of our study. It helps us to avoid several numerical errors during the model estimation².

²Since the intensity of the "Closure" is too small, our trade model cannot identify the impacts of certain

Table 4.2: Summary statistics for individual trade intensities

Full sample					
	Hold	Closure	Purchase	Redemption	Total
	97.48%	0.12%	1.16%	1.28%	1 163 519
Panel A: client category					
Bank	99.68%	0.07%	0.11%	0.12%	86 355
Private banking	98.86%	0.05%	0.36%	0.71%	128 341
Insurance	98.15%	0.08%	0.83%	0.92%	48 946
Firm	97.94%	0.15%	0.75%	1.14%	28 433
Mutual	97.23%	0.15%	1.16%	0.99%	304 626
Platform	95.41%	0.05%	1.76%	2.76%	220 873
Panel B: fund class					
Real estate	99.06%	0.02%	0.58%	0.32%	32 010
Equity	97.79%	0.07%	0.86%	1.12%	466 311
Diversified	96.28%	0.07%	1.55%	2.08%	58 076
Money market	91.91%	0.43%	3.96%	3.69%	170 054
Fixed income	99.27%	0.10%	0.28%	0.32%	162 295
Absolute return	99.66%	0.06%	0.17%	0.08%	121 532

This table provides the summary statistics for trade intensities. Each observation represents one choice that one investor makes for one fund at one day. The investor make transaction choices among: hold, closure, purchase and redemption. The closure represents the investor clears his inventory and exit from the fund. We show the intensity for each choice and the total observation , which is day number times investors number times fund number, is presented at the last column. The first row presents the full sample statistics. Then we give the Panel A for investor categories and Panel B for fund classes.

dummy variables on the "Closure" choice and hence proceed the numerical error in model estimation.

4.2 Empirical approach

4.2.1 Individual trade model

The "Discrete Response Model" is an appropriate tool to analyse the micro-transaction data, since we are interested in a non-continuous outcome, the trade decision. For example, [Odean \(1999\)](#) uses the Logistic regression to investigate the binary decision of investors: "Trade" vs "Hold". In our paper, we make a natural extension to the binary Logit model: in the "trade" category, we make the distinction between the "buy" and the "sell". Therefore, the trade decision becomes multiple-dimensional, investors choose among "Purchase", "Redemption" and "Hold". We further consider that, for the fund company, we prefer "Purchase" than "Hold" and we prefer "Hold" than "Redemption". The underlying intuition is straightforward, the asset manager earns a fee proportional to the fund size. Therefore, a boost of purchase trades increases fund companies' earning. Therefore, three trade choices are ordered in the model. Accordingly, the literature labels this model as the "Order Response Model" (see, e.g., [Becker and Kennedy 1992](#)). This approach has been used in various fields, such as the education choice ([Peng et al. 2002](#)), food consumption ([Myrland et al. 2000](#)), forensic anthropology ([Konigsberg et al. 2008](#)). In finance, people have experience of using this approach in the credit rating evaluations (e.g., [Carroll and Niehaus 1998](#); [Cheung 1996](#); [Landon and Smith 2000](#)).

Under this approach, there are three observed outcomes: $y = \textit{Redemption, Hold, Purcahse}$. Investors decide their trade choice by a latent utility function:

$$U_i = C + \beta^1 x_i^1 + \beta^2 x_i^2 + \dots + \beta^k x_i^k + \epsilon_i \quad (2.1)$$

In the equation above, i represents a "investor-fund-date" level trade decision. U_i is the utility value that determines the trade choice. C is the vector of constants. If there are n choices, there would be $n - 1$ constants in the model. Thus in our case, there are two constants. $X_i = \{x_i^1, x_i^2, \dots, x_i^k\}$ is the vector of explanatory variables of length K and β s are factor sensitivities. ϵ_i is the error term. Equation (2.1) describes that a mutual fund provides to the investor a "dynamic" utility value which is determined by several economic variables (X_i).

We suppose there are two thresholds α_1 and α_2 with $\alpha_1 < \alpha_2$. They separate investors' utility into three areas. We neither observe the utility function, nor these thresholds. We observe only the trade choices according to the following rule:

$$y_i = \begin{cases} \textit{Redemption} & \text{if } U_i < \alpha_1 \\ \textit{Hold} & \text{if } U_i > \alpha_1 \ \& \ U_i < \alpha_2 \\ \textit{Purchase} & \text{if } U_i > \alpha_2 \end{cases} \quad (2.2)$$

Each observation of y_i ³ represents a trade choice of an individual investor, for a single fund at one day. For instance, if investor 1 holds two funds in his portfolio: fund A and fund B. On 21st October 2015, he purchased more fund A but did not do anything for fund B. Accordingly, we have two observations: a Purchase and a Hold. Each trade is decided by the latent function U . If the investor achieves a utility higher than α_2 , he would purchase the fund. On the contrary, if the fund brings to him a utility lower than α_1 , he would redeem the fund. In other situations, he holds the fund and does nothing. We present this choice rule graphically in Figure 4.2. The green area at right indicates when the utility value is large enough, the investor would purchase the fund. On the contrary, the red area in the left is where investors redeem in response of the low utility.

The interpretations of factor sensitivities are more complicated than a traditional continuous regression (e.g. OLS). In a traditional linear regression, the marginal effect equals to the beta of the independent variable. However, in this multiple choice model, each independent variable has multiple marginal effects. We compute marginal effects on the predicted probabilities of each trade choice. For an explanatory variable is x^k , we have:

$$\begin{aligned} \frac{\partial \mathbb{P}(y = \textit{Redemption} \mid X)}{\partial x^k} &= -\phi(X'\beta - \alpha_1)\beta^k \\ \frac{\partial \mathbb{P}(y = \textit{Hold} \mid X)}{\partial x^k} &= [\phi(X'\beta - \alpha_1) - \phi(X'\beta - \alpha_2)]\beta^k \\ \frac{\partial \mathbb{P}(y = \textit{Purchase} \mid X)}{\partial x^k} &= \phi(X'\beta - \alpha_2)\beta^k \end{aligned} \quad (2.3)$$

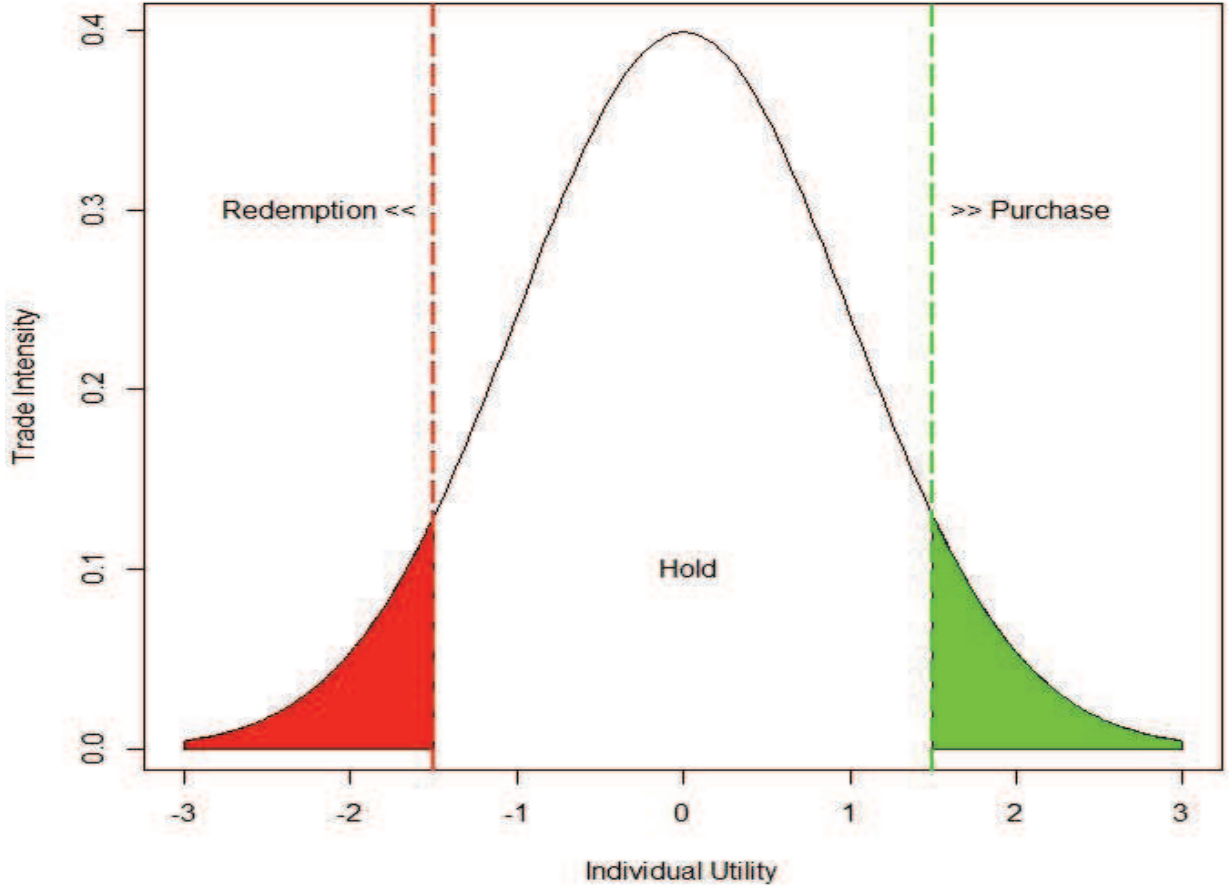
Therefore, the partial marginal effect of independent variable x^k on the predicted probability of:

- the "Purchase" decision has the same sign to β^k
- the "Redemption" decision has the opposite sign to β^k
- the "Hold" decision is not directly inferred from the sign of β^k

Luckily, we care less the Hold decision which is difficult since the marginal effect on this choice is ambiguous: it depends on the sign of $\phi(X'\beta - \alpha_1) - \phi(X'\beta - \alpha_2)$. Only Purchase and

³In the following of the paper, we drop sometimes the index i for the ease of the presentation.

Figure 4.2: Trade choice rule of ordered response model



In this figure, we show graphically the choice rule of the ordered response model. X-axis represents the utility value and y-axis is the trade intensity. Two dashed lines are thresholds to separate trade choices. The green area is where the investor choose to purchase and red area is where the investor redeem. The middle non-color area is where the investor holds their fund shares.

Redemption are our only concerns. In our investigation, a positive β indicates an increase of purchase intensity and a decrease of redemption intensity, and vice versa.

We estimate this non-linear model by maximizing its likelihood function. we set $z_{im} = \mathbf{1}(y_i = m)$, where m indicates three trade choices, i stands for an observation. Observation i's likelihood contribution is:

$$\begin{aligned}
 l_i &= \prod_m \mathbb{P}(y_i = m | x)^{z_{im}} \\
 &= \prod_m [\Phi(\alpha_m - x'_i \beta) - \Phi(\alpha_{m-1} - x'_i \beta)]^{z_{im}}
 \end{aligned} \tag{2.4}$$

We multiply all individual likelihood contributions and then we take the log-value of this

product to obtain the global log-likelihood function:

$$\log L(\alpha, \beta) = \sum_{i=1}^n \sum_m z_{im} \ln[\Phi(\alpha_m - x'_i \beta) - \Phi(\alpha_{m-1} - x'_i \beta)] \quad (2.5)$$

Then this optimization can be solved numerically.

4.2.2 Different rebalance methods

The standard portfolio management process requires a target asset allocation, which decides how an investor distributes his capital among different asset classes and it controls portfolio's risk exposures. In practice, the asset allocation is a dynamic process since the value of each security evolves. For example, if an investor wishes to have 40% equity securities in his portfolio, when equities' value appreciates, he would have a higher equity weight than that the target requires. Thereafter, his portfolio would have a higher equity risk exposure than what he desires. Accordingly, investors need to "rebalance" their portfolio, according to the performance of each asset. There are three traditional rebalance methods: "Buy-and-Hold", "Constant-Mix" and "CPPI".

The simplest method is the "Buy-and-Hold". Under this approach, after allocating the initial capital, investors do not make any further trade. It is consistent with trade patterns documented in the previous literature (see, e.g., [Johnson 2004](#)). It is a passive strategy, investors do not need to monitor his portfolio. Hence it is a suitable method for unskilled investors, such as retail ones (see, e.g., [Bailey et al. 2011](#)). However, investors lose the control of portfolio's risk exposure. If investors adopt this method, they should be insensitive to fund's return. Considering the high portion of institutional investors in our sample and high trade intensities observed (cf. Table 4.2, Section 4.1.2). This approach is unlikely to be used.

The second approach is referred to as "Constant-Mix". This approach intends to maintain a constant asset allocation. To this end, investors need to cut positions of securities which have appreciated and add positions of securities which have underperformed. For example, if investors have a target equity allocation of 40%, when the equity securities rise 5% in value, they would sell 5% equities to keep the target weight. This method ensures the constant weight of each asset and hence the portfolio's risk exposure is ensured. It performs the best in an oscillating market but it would underperform in a trending market. If mutual fund investors use the "Constant-Mix" strategy, we are expected to observe the negative sensitivities of return variables: positive return increases the redemption intensity

and decreases the purchase intensity.

The last rebalance strategy is the "CPPI": Constant Proportion Portfolio Insurance (see, [Perold 1986](#)). This strategy determines one asset's exposure E_j in the portfolio by formula $E_j = M_j \times (A_t - F_t)$, where M_j is a multiplier determined by the investor for each asset, A_t is the total portfolio value and F_t is a "floor" to ensure the minimum portfolio value. This strategy requires that investors purchase more the appreciated assets. It performs well in a trending market but it is not suitable for an oscillating market. If fund investors follow this strategy, they should buy "winner" funds and sell "loser" funds. It implies that a better return would increase the purchase intensity and decreases the redemption intensity.

In summary, we would identify investor's rebalance strategy by examining their reactions to fund returns. If they purchase more when the return is better, they have adopted the "CPPI" method; If they invert this relationship, the "Constant-Mix" approach is used; If fund returns do not affect the trade decision, investors have followed the "Buy-and-Hold" approach.

4.3 First evidence on rebalance activity

A. Description of the regression

We start our investigation in a "single fund" context, where investor's other fund holdings in parallel are not included in our test yet. In this section, we identify determinants of trade decisions for each transaction. We distinguish investors' sensitivity to short-term fund returns to their sensitivity to long-term fund returns. In our study, we consider the return calculated in a horizon longer than 100 days as the long-term return, a horizon between 5 days and 100 days as the mid-term return and returns within one week (5 days) as the short-term returns. We make this distinction since two types of returns relate to two different decisions: (1) Long-term returns affect investors' fund selection. When investors evaluate fund's quality, they would require a long track record, such as a quarter or a year, to assess the fund manager's ability and then they decide whether invest in this fund. (2) On the contrary, short-term returns relate more to the rebalance activity. When investors make daily portfolio adjustments, they "balance" the allocation among their holdings based on the actual performance in short horizons, such as a day or a week.

We identify the rebalance strategy by coefficients of short-term returns. Among three commonly used rebalance approaches (c.f. Section 4.2.2), "buy-and-hold" requires no reaction to returns. However, if investors adopt a "constant-mix" strategy, they would re-

purchase/resell under/better-performed funds. On the contrary, if "CPPI" method is used, investors should buy "high" and sell "low". In our test, we find evidence that fund investors follow the "constant-mix" strategy to add positions to under-performed funds and to cut positions to better-performed funds. However, in the long-term perspective, the purchase-return relationship is inverted: investors select more funds which have out-performed in the horizon longer than 100 days. The result stays robust either after we control for investors' behavioural bias or when we use different return measures.

We apply the ordered response model (Equation 2.1 and 2.2) to regress individual trade choices on three groups of explanatory variables and we present the estimated coefficients in Table 4.3. The first group (Panel A) contains short-term fund returns. We calculate fund returns for 5 previous days before each observation: $R[0-1]$ is the return of the day before the observation, $R[1-2]$ is of two days before and so on. Besides, in order to control for the tax effect, we isolate the trades executed during the December (" \times Dec"). In Panel B we present the mid/long-term return measures. We calculate the average return from 6 to 20 previous days (" $R[6-21]$ ") before each observation, from 21 to 50 previous days (" $R[21-50]$ "), from 51 to 100 previous days (" $R[51-100]$ ") and from 101 to 200 previous days (" $R[101-200]$ "). We then control for the tax effect in December in the same way. The last group includes various of control variables (Panel C). The first six variables in Panel C are all dummies variables. They relate the actual daily return ($R[0-1]$) to the historical performance patterns: whether actual return exceeds the 30 previous days' average return? whether it is positive? whether it is the maximum return during 30 past days? whether it is the minimum return during 30 past days? whether it is an extreme gain/loss? We define the extreme gain as the return which exceeds 10% and the extreme loss as returns lower than -10%. The "10%" and "-10%" extreme return criteria conform approximately to the 5th and 95th percentile value of daily returns' distribution in our sample. "ST variance" is the fund return variance during 50 previous days and the "LT variance" is calculated over 100 past days. "Dec" indicates if the trade happens in December. We control additionally for the previous aggregated net fund flows of the whole fund company (" $Flow^{company}$ ") and of the fund (" $Flow^{fund}$ "). We use MSCI index' daily return to proxy the market performance ("Market"). Finally, we include effects of the vintage year⁴, fund class and investor type. We label this regression as the "**Basic Trade Model**".

⁴The vintage year is the year when the fund is created. The previous literature finds strong impacts of vintage year on the fund performance

Table 4.3: Basic Trade Model (for rebalance analysis)

<i>Ordered Response Model: Redemption < Hold < Purchase</i>			
Estimated coefficients:			
	(1) Raw Model	(2) Average Model	(3) Market Model
Panel A: Short-Term Returns			
R[0-1]	-2.746*** (0.015)		-2.494*** (0.0076)
R[1-2]	-2.843*** (0.024)		-2.655*** (0.0073)
R[2-3]	-1.236*** (0.035)		-1.029*** (0.0076)
R[3-4]	-1.244*** (0.040)		-1.032*** (0.0077)
R[4-5]	-1.797*** (0.048)		-1.584*** (0.0076)
<i>R^{weekly}</i>		-1.919*** (0.018)	
R[0-1] × Dec	-1.091*** (0.012)		-5.241*** (0.023)
R[1-2] × Dec	8.151*** (0.052)		4.013*** (0.024)
R[2-3] × Dec	4.957*** (0.067)		0.736*** (0.023)
R[3-4] × Dec	4.693*** (0.039)		0.982*** (0.024)
R[4-5] × Dec	3.254*** (0.049)		-0.490*** (0.0237)
<i>R^{weekly}</i> × Dec		4.250*** (0.018)	
Panel B: Mid/Long-Term Returns			
R[6-20]	-1.197*** (0.039)	-1.163*** (0.039)	-1.153*** (0.0082)
R[21-50]	-0.100*** (0.41)	-0.096*** (0.041)	-0.072*** (0.0165)
R[51-100]	0.227*** (0.069)	0.239*** (0.068)	0.247*** (0.0213)
R[101-200]	0.163*** (0.063)	0.164*** (0.063)	0.177*** (0.0243)
× Dec	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>
Panel C: Control Variables			
Above 30 days' average	-0.171*** (0.014)	-0.169*** (0.013)	-0.169*** (0.0061)
Max during 30 days	-0.114*** (0.002)	-0.114*** (0.002)	-0.110*** (0.0006)
Min during 30 days	0.847*** (0.0003)	0.838*** (0.0003)	0.861*** (0.0008)
Positive	0.104*** (0.014)	0.094*** (0.013)	0.103*** (0.0057)

Continued on next page

Table 4.3 – Continued from previous page

	Estimated coefficients:		
	(1) Raw Model	(2) Average Model	(3) Market Model
Extreme gain	1.351*** (0.00005)	1.327*** (0.00006)	1.316*** (0.00001)
Extreme loss	-0.531*** (0.0003)	-0.419*** (0.0004)	-0.541*** (0.00001)
ST variance	-100.309*** (0.0028)	-142.703*** (0.0028)	-114.977*** (0.0009)
LT variance	-249.485*** (0.0022)	-244.542*** (0.0022)	-244.043*** (0.0002)
Dec	-0.067*** (0.024)	-0.069*** (0.024)	-1.067*** (0.0072)
$Flow^{Company}$	-0.105*** (0.012)	-0.104*** (0.011)	-0.107*** (0.009)
$Flow^{Fund}$	0.737*** (0.001)	0.734*** (0.001)	0.767*** (0.0001)
Market	-0.0004*** (0.00001)	-0.0004*** (0.00001)	-8.794*** (0.000007)
Vintage year	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>
Fund class	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>
Investor type	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>
Observations	1,163,519	1,163,519	1,163,519
AIC	252000.59	251997.97	251952.62
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01		

Table 4.3 reports coefficients of the Ordered Response Model, estimated by maximizing its likelihood function. The model explains individual daily trade choices among "Redemption", "Hold" and "Purchase" of all investors in one fund company over two years. We suppose there is a latent utility function with two threshold values, the left-threshold and the right-threshold. When an investor obtains an utility higher than the right-threshold, he makes the purchase decision. On the contrary, when the utility is lower than the left-threshold, the investor redeems his fund share. In other cases investor choose to hold. The sample contains 587 investors and 98 funds share-classes (ISIN number). In sum, there are 1 163 519 observations, from 1/1 2015 to the 30/12 2016. Each observation represents a trade decision made by one investor for one fund at one day. Empirical frequencies of three trade choices are 1.16% for "purchase", 97.48% for "hold" and 1.40% for "redemption". We put related variables together in one panel. Panel A shows the short-term return variables. Panel B presents mid/long-term returns. Panel C lists diverse control variables. We also include variables control for: the vintage year, the asset class and the investor type. In this regression, we adopt three specifications which use different alternatives to calculate returns. In "**Raw Model**", we use the simplest raw returns. In "**Average Model**", we use the weekly average return to replace 5 separate daily returns. In "**Market Model**", we calculate returns in excess of market performance. We label this regression as the "**Basic Trade Model**".

We use three alternatives to measure the fund returns (in all nine horizons). The first is the simple raw return. It is calculated as the log-difference of fund NAVs between two dates. We present it in the first column (**Raw Model**). Then in the second column (**Average Return**) we use the weekly average return (" R^{weekly} ") to replace first 5 days return ("R[0-1]" to "R[4-5]"). We intend to control for short-term returns' "cyclicality": some asset classes are susceptible to exhibit "cyclical" patterns in the short-term returns. For instance, money

market securities have this property since their performance is influenced by short-term interest rates which are often "cyclical". In the last column (**Market Model**) we calculate the fund excess return related to the market return.

B. "Constant-mix" approach of rebalance

In Panel A, we observe that short-term returns decrease the intensity of the purchase. This relationship stays the same in all three specifications. In "**Raw Model**", all five previous daily return have negative coefficients. They indicate a decrease of the latent utility value (c.f. 2.3). For example, a 1% increase in the previous daily raw return (" $R[0-1]$ ", first column) decreases investor's utility by 2.746%. It makes the trade density curve in Figure 2 shifting to the left. Therefore, investors reluctant to purchase. When we use "**Market Model**" and "**Average Model**", the negative relationship holds.

We interpret this result by the fact that investors adopt the "constant-mix" strategy to rebalance their portfolio of funds. Under this strategy, investors intend to keep the same weights different assets (funds) when their value evolves. Consequently, they would buy funds whose value has depreciated. On the contrary, they cut positions of funds whose value has appreciated. Since the rebalance decision is a short-term trading behaviour, we hold the opinion that investors rebalance according to the short-term fund returns.

However, the negative return-trade relationship is modified in December. In this month, with few exceptions, the best part of short-term return variables have strong positive influences on the purchase decision. Thus a better return decrease the propensity of redemptions. For example, " $R^{weekly} \times Dec$ " has a beta of 4.250, which has an opposite sign to " R^{weekly} "'s coefficient. It tells us that investors sell less better-performed funds in December. The most possible explanation for this positive return-trade relationship observed in December is the tax issue. In the last month of the year, investors avoid to trigger a large tax bill. One way to decrease the tax payment is to hold the out-performed funds which would realize the high profit if they are sold. An alternative approach is to sell securities at loss to offset realized earnings on other securities. Both these two methods require a "hold-high-and-sell-low" behaviour which we observe.

Panel B presents how investors react to mid- and long-term returns. Until the horizon of 50th previous day, the negative return-trade relationship still holds. In all three specifications, $R[6 - 20]$ and $R[21 - 50]$ both have negative coefficients as the short-term returns. However, long-term returns' coefficients become to positive. " $R[101-200]$ " calculated under "Market Model" has a coefficient of 0.177. It suggests that investors purchase more

better-performed funds in a long-term perspective. A very plausible explanation of this relationship is that long-term fund performance, especially quarterly (around 50-100 days) and annually (around 100-200 days) returns, are indicators for the fund manager's quality. Accordingly, investors select funds which have out-performed in long-term horizons.

An alternative explanation for the short-term negative return-trade relationship is the well documented "disposition effect" (Chang et al. 2016). It is a behaviour generated by the loss aversion bias: investors feel extremely bad to realized loss, hence they keep under-performed funds for the long time. Moreover, since they fear that a good return would not persist in the future, they would sell "winner funds" immediately. This effect creates also negative coefficients estimated above. Technically, it is difficult to separate the impacts of the "disposition effect" and the impacts of the rebalance activity. However, in our examination, there are two proofs which support more our argument about "rebalance". The first one is that long-term returns have positive impacts on fund purchase. It is odd that investors only exhibit "short-term" loss aversion, while they behave differently for long-term returns. Secondly, we add several variables to examine investors' reaction to previous price patterns. It gives us the evidence whether investors trade in a rational way or they exhibit the behavioural bias such as the loss aversion.

We present our control variables in Panel C. We first investigate investors' reactions to various historical return patterns, it reveals whether investors are rational. Then, we control different impacts of economic conditions such as the market performance ("Market"), the previous fund flows (" $Flow^{Fund}$ " and " $Flow^{Company}$ ") and the fund return variance (LT/ST "Variance"). Two flow variable represent aggregated net flows at two different levels. " $Flow^{Fund}$ " is the aggregated net flows of the observation fund in the previous day and " $Flow^{Company}$ " is the aggregated net flows of all funds in the company in the previous day. Since these control variables are same in all three specifications, we interpret only the result in the first column. We first observe from a sensitivity of -1.171 that investors purchase less a fund when its return is above its 30 days' average. Moreover, if the fund return raises to the 30-days maximum, it makes investors more possible to sell; in contrast, if the return is at 30-days' low, investors have a tendency to purchase the fund. These three coefficients are consistent with the "constant-mix" rebalance strategy: investors buy "low" and sell "high". Then we find investors are rational with following evidence: (a) the "Positive" dummy has a coefficient of 0.104, it indicates that investors purchase more funds with positive returns. It contradicts to the "disposition effect" which predicts that investors would sell "winner funds (/positive funds)" immediately; (b) we further study how investors trade under extreme returns. When a return is extremely good, we detect that investors

are more likely to purchase (a positive coefficient of 1.351); (c) in contrast, an extreme loss decreases investor's utility by 0.531. These two coefficients have exactly inverted signs to what the "disposition effect" requires. With this behavioural bias, investors fear that the extremely good return becoming bad, hence the coefficient of "Extreme gain" would be negative. On the contrary, irrational investors are reluctant to realize the loss, so "Extreme loss" would have a negative coefficient. Combining all findings above, investors' behaviour in our sample is rather rational. Therefore, we confirm that the negative return-purchase relationship in short-term is more possibly due to the rebalance strategy.

Together, we find that short-term and mid-term fund returns have negative impacts on investors' trade decisions. This result supports the "constant-mix" rebalance strategy but also the "disposition effect". After analysing (1) what are investors' interpretations to fund's long-term returns and (2) how investors react to the extreme returns; we are more convinced by the explanation of the rebalance activity than the explanation of the disposition effect. To continue our investigation, we ask a further question: what is the exact mechanism of the rebalance activity? We need to study how investor switch from one fund to another. We answer this question in the next section.

4.4 How investors switch among funds

A. Description of regression

In this section, we move from the previous "single fund" context to a "multiple funds" context: for each trade decision, we not only consider the fund which the investor trades, but also other funds the investor holds in parallel. For instance, an investor holds three mutual funds A, B and C. When we analyse his trade decisions on A, we take also the performance of B and C into account. We call the fund A as the "observation fund", other two funds as "parallel funds". In this test, we intend to identify the exact mechanism of rebalance: how investors switch from one fund to another. We search for three types of the "fund switch". First, we study the "return switch". It highlights how investors rebalance their fund holdings with different returns. Naturally, we expect a consistent result with the previous section, the "single fund" context. Secondly, we examine the "liquidity switch". We intend to identify the order of transactions when an investor holds funds with different liquidity levels. For example, either the investor redeems the liquid fund or the illiquid fund at first. Finally, we check the "risk switch". In this test, we evaluate the risk of "observation fund" relatively to investor's "parallel holding", then we monitor how fund's relative risk

impacts the trade decision. We present the estimation in Table 4.4.

We keep a similar model specification as in the last section, however, we include several additional variables concerning the "parallel holding". As the "Basic Trade Model", we start to test how investors react to fund returns. We keep the nine return measures: five short-term returns, 4 mid-/long- term returns. However, we modified the calculation to take the parallel holdings into account. For each observation date, we first compute the average return of the "parallel funds", then we calculate the excess return of the "observation fund" to this average. We note these returns as "parallel returns (R^*)". Then, we use the same control variables as the previous section. Finally, we add variables to check the three "fund switch" behaviours. The first two dummies test the "return switch": "Sell better R" is a dummy to highlight whether the investor has redeemed a better-performed fund in the previous week before the observation date. Inversely, "Sell worse R" show whether the investor has redeemed an under-performed fund. After, we construct several dummy variables for the "liquidity switch". "Most/Least liquid" indicates if the observation fund is the most/least liquid fund among investor's holding. "Sell other liquid" shows whether the investor has sold a more liquid fund. "Sell other illiquid" highlights whether the investor has redeemed less liquid funds. Finally, we use two variables to test the "risk switch". The variable "Most risky" equals to one when the observation fund's (LT) return variance is the highest among investor's holding. "Excess variance" is the variance of observation fund minus the average variance of parallel funds.

Table 4.4: Fund Switch Model (for rebalance analysis)

<i>Ordered Response Model: Redemption < Hold < Purchase</i>	
Estimated coefficients:	
Panel A: short-term returns	
R*[0-1]	-2.074*** (0.015)
R*[1-2]	-1.615*** (0.0035)
R*[2-3]	-1.007*** (0.0040)
R*[3-4]	-0.994*** (0.0036)
R*[4-5]	-1.204*** (0.0044)
× Dec	Yes
Panel B: mid/long-term returns	
R*[6-20]	-1.146*** (0.039)
R*[21-50]	-0.089***

	(0.041)
R* [51-100]	0.240***
	(0.066)
R* [101-200]	0.164***
	(0.063)
× Dec	Yes
Other control variables	Yes
Panel C: funds switches	
Return switch:	
Sell better R	0.011***
	(0.031)
Sell worse R	-0.021***
	(0.0032)
Liquidity switch :	
Most liquid	0.010***
	(0.027)
Least liquid	-0.005***
	(0.093)
Sell other liquid	-0.102***
	(0.037)
Sell other illiquid	0.031***
	(0.010)
Risk switch :	
Most risky	0.012
	(0.018)
Excess variance	-507.153***
	(0.0015)
Observations	1,163,519
AIC	251709.29
<hr/>	
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 4.4 reports coefficients of the Ordered Response Model, estimated by maximizing its likelihood function. The model explains individual daily trade choices among "Redemption", "Hold" and "Purchase" of all investors in one fund company over two years. We suppose there is a latent utility function with two threshold values: the left-threshold and the right-threshold. When an investor obtains an utility higher than the right-threshold, he makes the purchase decision. On the contrary, when the utility is lower than the left-threshold, the investor redeems his fund share. In other cases investor choose to hold. The sample contains 587 investor accounts and 98 funds share-classes (ISIN). In sum, there are 1 163 519 observations, from the 1st January 2015 to the 30th December 2016. Each observation represents a trade decision made by one investor for one fund at one day. Empirical frequencies of three trade choices are 1.16% for "purchase", 97.48% for "hold" and 1.40% for "redemption". Comparing to previous "**Basic Trade Model**", this regression includes the "parallel holdings" in addition. In this model, we calculate the average holding returns of each investors and we calculate fund excess returns to these individual averages. The model further studies how investors switch among different funds. We put related variables together in one panel. Panel A shows the coefficients of the short-term return variables. Panel B presents the coefficients of the mid/long-term returns. In Panel C, we include variables to test the "fund switch" behaviour. We keep all the control variables in the "**Basic Trade Model**". We label this regression as "**Fund Switch Model**".

B. Evidence on fund switches

Panel A and B show the coefficients of the return variables. We obtain the consistent results as the "Basic Trade Model": investors sell funds with higher short-term and mid-

term returns but this return-transaction relationship is inverted for the long-term returns. From $R^*[0-1]$ to $R^*[21-50]$, all coefficients are negative. In contrast, investors have positive sensitivities 0.240 and 0.164 for $R^*[51-100]$ and $R^*[101-200]$ respectively. All returns in this section are calculated as the observation fund's return in excess of the "parallel holding". The same fund may have different "parallel returns" for different investors which depends on each investor's holding situation. The short-term negative purchase-return relationship reveals that investors compare performances of all holdings and reallocate capital from depreciated funds to appreciated funds. Therefore, our analysis in the "single fund context" and the "multiple fund context" both give the evidence on the "constant-mix" rebalance. Meanwhile, we use the same control variables as the previous section and their coefficients are similar as the previous test. Hence we leave these control variables unreported.

Panel C displays our test for the "fund switch" behaviour. We begin with the "return switch". We find that, in the week before the observation date, if an investor has sold a better-performed fund, his utility of purchasing the observation fund would increase by 0.011. Thus he switches from a better-performed fund to another under-performed fund. This behaviour is consistent with the "Constant-Mix" rebalance strategy. However, if he has already sold a "worse" fund, the utility value would decrease by 0.021, his purchase intensity falls down. The possible underlying intuition is as follows: when an investor is forced to sell an under-performed fund, it is highly possible that he has a liquidity "shock". Since his cash need is urgent, he sells all his holdings, even the worse-performed securities. Consequently, he reduces his purchase intensity. We could also speculate that when investors need to sell his assets for the liquidity reason, he first sell better-performed funds.

We further investigate the "liquidity switch". The "Most liquid" Dummy has a positive coefficient of 0.010 and "Least Liquid" has a negative coefficient of -0.005. These two variables indicate that, when investors face two funds, a liquid one and an illiquid one, to choose, they prefer to purchase the more liquid fund at first. Then, we examine the impacts of investors' previous trades in a similar way as the "return switch" in the last paragraph. We observe that when an investor has sold an illiquid fund in the week before the observation date, he is less likely to purchase after. The plausible explanation is that investors tend to sell the more liquid fund at first when liquidity problem arrives. When he is forced to redeem his holdings to satisfy his cash needs, he starts to sell the most liquid fund. Thereafter, the possibility to purchase a new fund is reduced in presence of the liquidity shocks. Alternatively, we can interpret the coefficient in the point of view of the redemption: when an investor has the liquidity need and he has already redeemed one fund, it is highly possible that he would redeem another fund. Moreover, we find that when the investor has

sold a more illiquid fund previously, he is more likely to buy the observation fund since the utility increases by 0.031. It shows a "Jump-to-Liquidity" behaviour. Investors switch from illiquid funds to liquid funds in order to prepare future cash needs.

Lastly, we study how relative variance impacts investor's trade decisions. "Most Risky" dummy has no significant impact on the purchase possibility. The fact that a fund has the highest variance among one's holdings does not affect investor's trade choice. However, we find that "Excess variance" has a negative coefficient. It suggests that if a fund's performance variance increases relatively to investor's parallel holdings, investor's intention to purchase would decrease. This finding suggests that investors switch from high-risk funds to low-risk funds.

C. Is investor's behaviour time-varying?

The previous research (e.g, [Ben-Rephael et al. 2012](#)) documents that fund clients may exhibit the time-varying behaviour. Their trade pattern might be modified over the different periods of the year for several reasons. First, investors' risk aversion might evolve. [Kamstra et al. \(2017\)](#) show that investors have tendency to avoid the high level of the risk in the end of the year. Second, investors have different liquidity needs in the beginning and in the end of the year. For instance, an institutional investor needs to pay the dividends to its investors in the end of the year. Hence it would redeem more over this period. A second example is a bank whose clients withdraw more deposits in the end of the year, perhaps for preparing for the holiday consumption. Hence the bank investor in mutual funds redeem more accordingly. Lastly, some investors may "window-dress" their portfolio for reporting purpose in some periods of a year. All these scenarios make investors modify their trade behaviour. Accordingly, we separate the data into four sub-samples, one for each quarter. We seek to verify the robustness of our previous findings by examining whether the rebalance behaviour stays stable during the year.

We present four quarters' sub-sample tests in Table 4.5. We apply the same specification, the "Fund Switch Model", to four quarter sub-samples. Each column of the table lists the estimation for one quarters, from Q1 to Q4. Most of coefficients stay stable during the year, therefore we do not present these variables. We list only the coefficients that have significant divergences in four quarters.

First, we find that investors' reactions to the long-term returns change significantly in the fourth quarter. In the first three quarters, $R^*[101-200]$ has positive coefficients. The positivity indicates that investors select well-performed funds in the long run. However,

Table 4.5: Fund Switch Model in quarterly sub-samples (for rebalance analysis)

<i>Ordered Response Model: Redemption < Hold < Purchase</i>				
Estimated coefficients (time varying):				
	Q1	Q2	Q3	Q4
R* [101-200]	0.729*** (0.030)	1.395*** (0.013)	0.653*** (0.024)	-0.032*** (0.0051)
Max during 30 days	0.328*** (0.001)	-0.491*** (0.001)	-0.093*** (0.001)	-0.064*** (0.001)
$Flow^{Fund}$	0.144*** (0.001)	-0.429*** (0.001)	-0.339*** (0.002)	1.307*** (0.0001)
Most liquid	0.064*** (0.015)	0.045*** (0.011)	-0.013*** (0.0084)	-0.036*** (0.028)
Least liquid	0.010 (0.016)	-0.017*** (0.012)	-0.051*** (0.012)	0.026*** (0.010)
Sell other illiquid	0.030*** (0.005)	0.152*** (0.002)	-0.119*** (0.006)	-0.010*** (0.003)
Best R* [0-1]	2.557*** (0.012)	-1.329*** (0.0097)	-1.279*** (0.0027)	-3.456*** (0.0037)
Observations	289,433	282,251	299,162	292,673
AIC	63118.05	63290.05	61259.63	63871.09

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 4.5 presents our robustness checks for four quarterly sub-samples. For each sub-sample, we apply the "Fund Switch Model" and we intend to examine previous whether estimated coefficients stay constant during four quarters. We list the coefficients which have the significant divergence over the different periods. Other stable coefficients are not presented.

this relation switches to negative in the end of the year. A possible interpretation is that investors sell more better-performed funds to realize profits. Then we find investors' reaction to "Max during 30 days" is positive in the first quarter, whereas this coefficient is negative in other periods. It seems that, in the beginning of the year, investors prefer the fund which has "peaked" the monthly high. The variable "Best R*[0-1]" exhibits a similar pattern. We observe that investors react negatively to "Best R*[0-1]" from Q2 to Q4. However, their sensitivity to this variable becomes positive in the first quarter. These two factors show that "Constant-Mix" rebalance deviates slightly in the first quarter. " $Flow^{Fund}$ " has a positive

coefficient in the Q1 and Q4 but it turns to negative in Q2 and Q3. It highlights that investors react more to others' trade at the beginning and the end of the year.

Investors' attitude towards to the most liquid fund changes over time too. In the first half of the year (Q1 and Q2), "Most liquid" has a positive coefficient as in the "Fund Switch Model" (cf. Table 4.4). However, in the second half of the year (Q3 and Q4), this coefficient becomes negative. A possible explanation to this change is that in the second half of the year, investors need to set a larger cash reserve to prepare the liquidity needs. They sell the most liquid funds at first. For the "Least liquid" dummy, we observe a different effect in the last quarter: investors' sensitivity is negative in Q2 (-0.017) and Q3 (-0.051) but becomes positive in Q4 (0.026). A positive coefficient indicates either investors purchase more or redeem less. In this case, we hold the opinion that investors redeem less the funds which are the least liquid ones among parallel holdings in the Q4 because they sell more liquid funds to prepare the cash needs. Furthermore, "Sell other illiquid" has the positive coefficient in Q1 and Q2 but it switches to negative in Q3 and Q4. The positive coefficient reveals the "jump to liquidity" behaviour. However, this behaviour vanishes in the second half of the year. This result might also be explained by the fact that, during the second half of the year, investors sell liquid funds at first to prevent cash consumption. If they have already sold illiquid assets, it shows that they have already "liquidity shocks", therefore they purchase less.

In summary, when we include the "parallel holding" information in our study, we find clear evidence that investors hold a global view of the full portfolio to make trade decisions, and funds held by the same investor are analysed together. We first confirm our hypothesis on the "Constant-Mix" rebalance strategy: investors buy low and sell high. Further, we examine how investors switch their fund portfolio. We observe that investors tend to "jump" to under-performed, more liquid and lower-risk funds. Most of their behaviours stay stable during the four quarters. However, some of their behaviours are modified during some periods, especially in the first and last quarter of the year.

4.5 Discussion and conclusion

Our paper presents a comprehensive analysis on the rebalance activities of mutual fund clients. We document that investors do not follow the simple "buy-and-hold" strategy in the fund investing, but rebalance their portfolio on a daily basis. They purchase short-term under-performed funds to maintain a constant capital allocation. However, we need to distinguish the short-term rebalance trades to the long-term fund selection. Moreover,

investors do not monitor funds' performances in the isolated manner. Funds owned by the same client have influences on each other. Finally, we find that the majority of the rebalance activities are stable during different quarters of the year.

Our investor sample contains a diversified range of investors, including several institutional categories such as banks, corporate firms and FoFs. It differs from previous studies on micro-transaction data which cover mostly only retail investors. It is a possible reason why we observe relative sophisticated behaviours such as the capital allocation and rebalance activities. The "intelligence" of fund investors makes us asking a fundamental question of the delegated asset management: whether investors need only passive diversified management style or fund companies should provide the more active and technical strategies? It is an old debate and we see successful examples in both approaches. Triumphant passive managers such as John Bogle offer only index fund services. They contest the claim that the active management could not deliver the superior return and all asset managers should offer fully diversified portfolio. However, there are also uncountable examples of successful active fund managers, such as James Simons. Investors often confuse which type of asset managers is most suitable for them. Our research helps investors to answer this question.

Our study documents that fund investors adopt relative complicated rebalance strategies. It implies that these investors decide already their asset allocation. Therefore, the diversification is done at the investor level. It seems that these investors do not need the asset manager provides a diversified service. Besides, mutual funds often charge a non-negligible fees which are too expensive for passive strategies. Investors have cheaper substitutions in the passive investing market, such as ETF or index futures. Based on the fact that investors have already a target allocation, mutual funds should offer more "technical" management strategy: (1) Offer manager's expertise in some asset classes. Mutual funds need deliver a better return than what investors can achieve themselves. (2) Provide the unique risk exposure. Funds with a unique style often have different risk exposures. For instance, quantitative and/or ISR (Investments in Social Responsibility) funds provide isolate risk exposure to investors, they help their clients to better diversify their investment and reach a higher efficient frontier. Overall, in this highly competitive market, mutual fund managers should offer their strong ability to investors. A simple passive investing service is not suitable for the professional institutional investors.

Conclusion générale

Mes travaux de thèse contribuent à la modélisation du passif des fonds ouverts. Ce côté passif est une source non-négligeable du risque de liquidité, mais reste souvent ignoré par les études académiques et les praticiens du marché. De nombreux fonds bien gérés ont échoué à cause de sorties massives des clients. Ces scénarios de "fund run" impactent non seulement les sociétés de gestion et leurs investisseurs, mais aussi génèrent des conséquences néfastes au niveau macroéconomique. Nous avons vu que, pendant les "fund run" de 2006 et 2008, le prix d'une classe d'actifs entière avait été impacté. En 2011, nous avons observé que les "fund runs" ont influencé les activités de financement des entreprises. En outre, un "run" peut aussi transférer le problème de liquidité entre des actifs différents. Un actif initialement de bonne qualité peut être contaminé pendant un "run". A l'égard de ce problème, les régulateurs européens ont introduit de nouvelles réglementations sur la gestion de la liquidité du passif du fonds. Mes études répondent aux trois dimensions de cette préoccupation: collecter la base de données du passif des fonds, connaître le comportement des investisseurs et prévoir le risque de sorties massives.

Dans le premier chapitre, nous révisons d'abord la littérature sur le sujet de risque de liquidité des fonds ouverts. Cette étude nous a découvert les différentes bases de données utilisées par des études précédentes. Nous observons que les bases de données traditionnelles ne fournissent pas d'information sur les investisseurs individuels. Ainsi, leurs comportements sont rarement analysés. Afin de compléter la littérature, nous avons collecté une nouvelle base de données au passif des fonds mutuels. Cette base nous permet d'observer les transactions individuelles des investisseurs et de calculer différentes mesures des flux. Au cours de ce travail, nous reconnaissons certaines difficultés par rapport à la qualité de données. Beaucoup de sociétés de gestion ne tient pas à jour un rapport complet des transactions individuelles. En conséquence, il est impossible d'utiliser les outils statistiques que nous avons développé pour ces sociétés pour gérer le risque client. Les sociétés qui ont commencé à archiver les données "clients" de bonne qualité sont en avance sur le marché.

Le deuxième chapitre de cette thèse propose une nouvelle modélisation des flux des

fonds. Nous nous interrogeons sur l'aspect "time series" des flux, alors que, la littérature étudie plutôt l'aspect "cross-section". Au lieu d'identifier les fonds qui vont recevoir plus de flux comme la littérature a étudié, nous nous focalisons sur les propriétés statistiques de chaque série de flux. Nous développons un modèle couvrant deux faits stylisés des flux: la "self-exciting" et la "over-dispersion". Notre étude montre que ces deux propriétés de flux génèrent des scénarios de sorties extrêmes. Nous observons également la contagion des chocs de liquidité entre des investisseurs dans un même fonds. Cela suggère qu'une partie de risque de liquidité provient du côté passif du fonds. De ce perspective, cet article de recherche diffère des études précédentes qui ne considèrent que l'actif du fonds. Ce nouveau modèle aide les gérants de fonds à obtenir une prévision fiable des flux futurs. Ce travail fait partie du projet "Modélisation et Gestion du Passif du Fonds (MGPF)" coopéré par l'Université Paris-Dauphine et plusieurs partenaires industriels comme Groupama AM, OFI AM et Scaled Risk. Ce projet a pour objectif de répondre aux besoins réglementaires et développer les outils informatiques permettant de gérer le risque de passif.

Nous continuons d'étudier l'horizon d'investissement des investisseurs dans les fonds mutuels dans le chapitre suivant. Cette étude cherche à expliquer les différences parmi des investisseurs diverses. Nous identifions un effet significatif des caractéristiques des investisseurs, comme leur secteur d'activité par exemple. Ensuite, nous documentons que les horizons d'investissement évoluent suite aux différentes conditions économiques. Par exemple, les investisseurs peuvent raccourcir leurs horizons d'investissement et sortir du fonds suite à une mauvaise performance. Nous introduisons ainsi le risque de sortie pré-maturée: ses clients sortent avant "l'échéance" à cause de la mauvaise circonstance économique. Enfin, nous trouvons que ce risque est plus élevé chez les investisseurs de long-terme. Cette étude indique que l'hétérogénéité des investisseurs a des impacts sur le risque de liquidité. Ainsi, nous conseillons aux gérants de fonds de surveiller le risque de sortie au niveau individuel. Cela confirme l'utilité de notre base de données qui enregistre les comportements individuels. D'ailleurs, cette analyse couvre des fonds investi dans les différentes classes d'actifs et elle montre que ce risque de sortie existe dans tous types de fonds ouverts.

La dernière partie de cette thèse met en évidence des activité de rééquilibrage des investisseurs. Nous trouvons que les investisseurs détiennent souvent un portefeuille contenant plusieurs fonds et le rééquilibrent par une approche "constant-mix". Les investisseurs ajustent leurs portefeuilles selon le rendement, le niveau de liquidité et le niveau de risque de chaque fonds. Cette étude explique le degré extraordinaire d'activités de transactions dans les fonds par l'activité de rééquilibrage. En outre, nous mettons en exergue deux nouveaux aspects du comportement des investisseurs. Premièrement, les investisseurs institutionnels

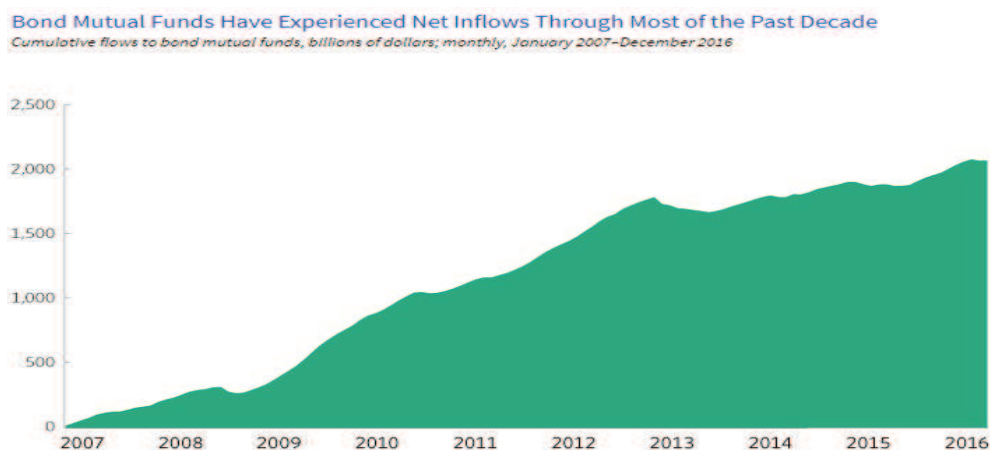
évaluent leurs investissements de façon globale. Deuxièmement, notre étude suggère que les investisseurs de fonds ont une allocation d'actifs et ont probablement déjà diversifié leurs portefeuilles. Pour ces clients, un service de gestion passive est ainsi insuffisant. La société de gestion doit absolument proposer des services plus actifs et sophistiqués à ces clients institutionnels.

La nouvelle base de données joue un rôle crucial dans nos travaux. Afin de gérer le risque de passif, il est nécessaire de collecter les données "clients". Au cours de ces travaux, nous avons observé que le gérant et les vendeurs se situent dans deux équipes différentes et ne communiquent pas. Ainsi, le côté actif et le côté passif du fonds sont séparés par une frontière: le gérant ne connaît pas le passif du fonds. Nous conseillons fortement de lever cette frontière et créer une coordination entre le gérant, qui connaît l'actif du fonds, et les vendeurs, qui connaissent le passif du fonds. Le timing de l'achèvement de cette thèse coïncide avec la nouvelle réglementation sur le risque de passif qui a été mise à place en janvier 2018. Suite à cette réglementation, toutes les sociétés de gestion européennes vont commencer à construire une base de données ressemble à la nôtre. Avec trois sociétés de gestion de taille moyenne, notre échantillon contient déjà plusieurs millions de transactions. La future base de données future sera beaucoup plus volumineuse. Nous pourrions ainsi commencer à préparer des techniques de "big data" d'étudier les nouvelles données.

Annexe générale

Annexe A: l'évolution des flux cumulés des fonds "fixed income" de la dernière décennie

Figure 4.3: Les souscriptions cumulées dans les fonds "fixed income" (2007-2016)

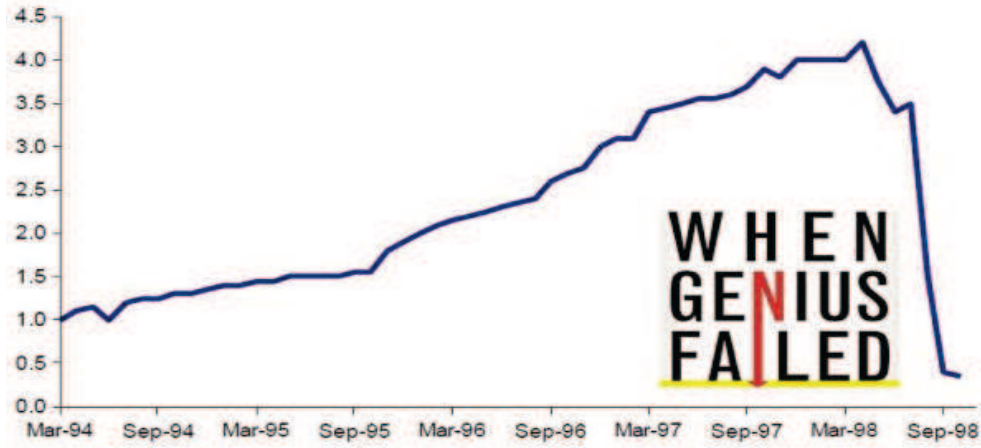


Ce graphique présente les souscriptions cumulées dans les fonds "fixed income" aux États-Unis, de janvier 2007 à décembre 2016, en fréquence mensuelle. L'unité de chiffre dans l'axe des ordonnées est de milliard dollars. Source: "ICI 2017 INVESTMENT COMPANY FACT BOOK, chapter 2."

Dans la Figure 4.3, on présente les flux cumulés investis dans les fonds "fixed income" de janvier 2007 à Décembre 2016 à une fréquence mensuelle aux États-Unis. Ce graphique montre une remarquable tendance croissante d'investir dans les fonds "fixed income". Cette étude est effectuée par l'ICI.

Annexe B: l'évolution de l'encours de LTCM et sa chute au septembre 1998

Figure 4.4: Évolution de l'encours du fonds LTCM (1994.3- 1998.9)



Ce graphique présente l'évolution de l'encours du fonds LTCM de mars 1994 à septembre 1998. L'unité de chiffre est de milliard euros. Source: "zerohedge.com"

Annexe C: différentes classes d'actifs gérées par les fonds mutuels ouverts

Cette figure nous montre la distribution du nombre de fonds dans chaque catégorie en France à la fin 2016. On observe que, les fonds actions représentent la plus grande part du marché. Ensuite, les fonds "fixed income" et "allocation" représentent un nombre important du total. Les fonds monétaires semblent peu nombreux mais représentent une grande part des AUMs. Un autre point remarquable est que les fonds alternatives ont également une part non-négligeable, soit 10% du marché. Cela implique que les sociétés de gestion commencent à utiliser la structure ouverte pour gérer des actifs non-traditionnels.

Figure 4.5: Différentes classes d'actifs gérées par des fonds mutuels ouverts

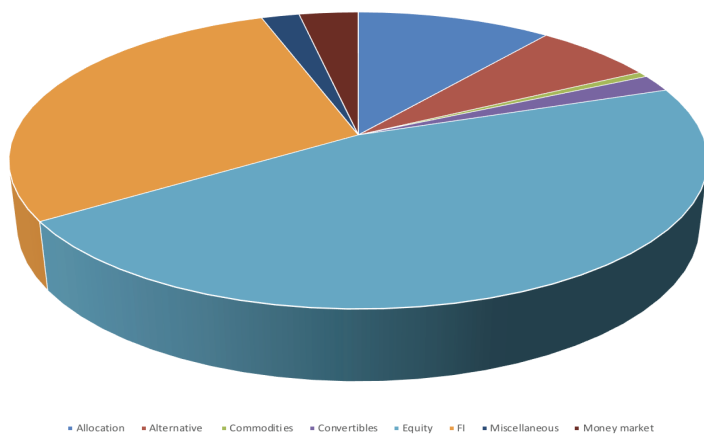


Figure 4.5 montre la distribution de nombre des fonds français dans chaque catégorie à la fin 2015. Cette classification est créée par Morningstar.

Annexe D: démonstration du modèle Berk-Green⁵

Le modèle Berk-Green a pour objective de lier deux comportements principaux que la littérature précédente avait eu de difficultés de se réconcilier: les flux du fonds sont sensibles à la performance et il n'y a pas de persistance dans la performance.

Le modèle suppose que la capacité d'un gérant est représentée par sa sur-performance par rapport au rendement de référence ("benchmark"). Le modèle suit un équilibre partiel: les gérants n'influencent pas la performance du "benchmark" et on n'observe pas la source de cette capacité. Cette dernière est inconnue à gérant lui-même et à ses clients. Ses derniers apprennent cette capacité avec l'histoire des rendements observés. Le rendement brut de la période t est $R_t = \alpha + \epsilon_t$. Cependant, les investisseurs ne reçoivent que le rendement net qui égale à R_t moins les coûts et les frais. Le paramètre α est la source de capacité et le terme d'erreur ϵ_t suit une distribution gaussienne avec la moyenne 0 et la variance σ^2 et il est distribué indépendamment du temps. Nous introduisons un nouveau paramètre $\omega = \frac{1}{\sigma^2}$ qui indique la précision de cette incertitude. L'apprentissage des investisseurs est la source de la relation entre la performance et le rendement.

L'AUM du fonds à la période t est q_t et il y a un coût C de gérer ce portefeuille qui est une fonction de q_t : $C = C(q_t)$. Pour tous $q > 0$, on a $C(q) \geq 0$, $C'(q) > 0$ et $C''(q) > 0$. Cette hypothèse est expliquée par le fait que un plus grand AUM est associé à un coût de gestion (e.g., frais de transaction, impact de prix, contrainte de capacité) plus élevé. Ainsi, la taille du fonds diminue sa performance. En outre, nous supposons qu'il y a un frais de gestion, f , qui est exprimé en proportion de q_t . Après avoir enlevé ces deux coûts, l'investisseur reçoit un rendement net:

$$TP_{t+1} = q_t R_{t+1} - C(q_t) - q_t f$$

Nous utilisons r_t pour représenter le taux de sur-performance à la période t . Nous avons:

$$\begin{aligned} r_{t+1} &= \frac{TP_{t+1}}{q_t} \\ &= R_{t+1} - \frac{C(q_t)}{q_t} - f \\ &= R_{t+1} - c(q_t) \end{aligned} \tag{1-A}$$

⁵Dans cette annexe, on présente la partie principale du modèle Berk-Green. Cette partie du modèle se trouve dans la section 1 et 2 du [Berk and Green \(2004\)](#). On utilise les mêmes représentations des paramètres et les numéros des équations (avec "-A" en plus) comme dans leur article, afin de garder la cohérence.

où

$$c(q_t) \equiv \frac{C(q_t)}{q_t} + f \quad (2-A)$$

Ainsi, r_t est le rendement empiriquement observé et $c(q_t)$ est le coût unitaire associé à la gestion active du fonds.

Le modèle suppose que les investisseurs sont rationnels, ainsi, ils mettent à jour leurs estimation sous une approche bayésienne. Quand un fonds est créé, le prior de l'investisseur sur la capacité du gérant est de moyenne Φ_0 et de variance η^2 . La précision de cette distribution est de $\gamma = \frac{1}{\eta^2}$. La moyenne de α a posteriori est:

$$\Phi_t \equiv \mathbb{E}(R_{t+1} | R_1 \dots R_t)$$

Nous supposons que les souscriptions sont infiniment élastiques à la sur-performance positive. Ainsi, dans chaque équilibre sous la compétition parfaite, le rendement marginal sur le dernier dollar investi doit être 0:

$$\mathbb{E}_t(r_{t+1}) = 0 \quad (3-A)$$

La condition (3-A) implique qu'il n'y a pas de persistance dans les rendements et la moyenne inconditionnelle des sur-performances est de 0. Nous prenons l'espérance sur les deux côtés de (1-A) et fixe l'espérance de rendement reçu à 0 comme dans l'équation (3-A), cela donne:

$$\begin{aligned} \Phi_t &= c(q_t) \\ &= \frac{C(q_t)}{q_t} + f \end{aligned} \quad (4-A)$$

Si Φ_t change, q_t change en conséquence afin de garantir l'égalité à l'équation (4-A). Cette relation indique comment les flux (Δq_t) évoluent en fonction de "l'estimation" de la capacité du gérant (Φ_t). En appliquant le théorème 1 dans [DeGroot \(1970\)](#), nous obtenons que, pour chaque fonds pendant $t - 1$ et t , l'évolution de Φ_t et les flux entrants/sortants, en fonction de la performance passée, sont donnés par la solution de:

$$\Phi_t = \Phi_{t-1} + \frac{\omega}{\gamma + t\omega} r_t \quad (5-A)$$

et

$$c(q_t) = c(q_{t-1}) + \frac{\omega}{\gamma + t\omega} r_t \quad (6-A)$$

Notons:

$$c'(q_t) = [C'(q_t) - \frac{C(q)}{q}] > 0$$

Nous savons par le théorème des accroissements finis, pour chaque q , il y a un point \bar{q} , $0 \leq \bar{q} \leq q$, tel que $[C(q) - C(0)]/q = C'(\bar{q})$. La convexité stricte de $C(\cdot)$ et l'hypothèse $C(0) = 0$ implique $C'(q) > C'(\bar{q}) = C(q)/q$. Ainsi, une bonne performance (positive r_t) sera suivi par les "inflows" ($q_t > q_{t-1}$), et vice versa. Le principal résultat ainsi est obtenu, le modèle montre que même si sans la persistance de la performance, les investisseurs suivent toujours les fonds avec les meilleures performances passées.

Annexe E: démonstration du modèle Chen-Goldstein-Jiang⁶

Le modèle Chen-Goldstein-Jiang explique dans un contexte où un fonds illiquide a réalisé une mauvaise performance, les "payoff complementarities" génèrent une motivation aux investisseurs de sortir du fonds. Cela entraînerait le scénario de "fund run".

Le modèle est discrète avec deux périodes 1 et 2. Chaque investisseur provenant d'un continuum $(0, 1)$ détient une part dans un fonds mutuel. Le montant d'investissement global est normalisé à 1. A la date 1, le rendement du fonds est R_1 et il est observable au tout le monde. Dans cette date, une fraction des investisseurs, \bar{N} , choisissent à rester ou sortir du fonds et une fraction N des investisseurs choisissent à sortir. On suppose que les investisseurs vont obtenir un rendement normalisé, 1, à la date 2 après être sorti du fonds à la date 1. Donc les investisseurs sortants réalisent un rendement total sur deux périodes R_1 .

Nous imposons une externalité négative aux investisseurs restants: le gérant paie une prime d'illiquidité λ quand il vend ses actifs afin de répondre les rachats à la date 1. Le paramètre $\lambda > 0$ indique le niveau d'illiquidité du fonds. Sans prenant compte des "inflows", les investisseurs restants réalisent un rendement pendant deux périodes:

$$\frac{1 - (1 + \lambda)N}{1 - N} R_1 R_2(\theta) \quad (5-B)$$

⁶Dans cette annexe, on présente la partie principale du modèle Chen-Goldstein-Jiang. Cette partie du modèle se trouve dans l'annexe A.1. du [Chen et al. \(2010\)](#). On utilise les mêmes représentations des paramètres et les numéros des équations (avec "-B" en plus) comme dans leur article, afin de garder la cohérence.

où $R_2(\theta)$ est la performance du fonds pendant la deuxième période, qui est une fonction croissante de paramètre θ . Ce dernier représente les fondamentaux du fonds, qui couvrent la capacité du gérant à délivrer une bonne performance.

Ensuite, on intègre le fait que la bonne performance peut attirer les flux entrants à la date 1. On dénote $I(R_1)$ est les "inflows" qui est une fonction croissante de R_1 . Grâce à ces flux entrants, le gérant vend moins d'actifs à la date 1. Par conséquent, on remplace N de l'équation (5-B) par $\max\{0, (N - I(R_1))\}$. Cela signifie que, si les "inflows" sont plus grands que les "outflows", le gérant n'a plus besoin de liquider ses actifs à la date 1. Le rendement des investisseurs restants dévient:

$$\frac{1 - (1 + \lambda)\max\{0, (N - I(R_1))\}}{1 - \max\{0, (N - I(R_1))\}} R_1 R_2(\theta) \quad (6-B)$$

Ainsi, les investisseurs choisissent entre sortir à la date 1 et obtenir R_1 et attendre jusqu'à la date 2 et obtenir le rendement dans (6-B). Clairement, le rendement des investisseurs restants a une relation positive avec le paramètre $\theta \rightarrow$ les fondamentaux du fonds. En revanche, il décroît avec le nombre des investisseurs qui sortent, N . Ce dernier détermine l'équilibre du modèle.

Puisque le rendement de chaque investisseur dépend aussi des décisions des autres à cause de la externalité négative du rachat, le modèle pourrait posséder plusieurs équilibres. Afin de les identifier, on définit deux seuils des fondamentaux: $\underline{\theta}$ et $\bar{\theta}(R_1)$. $\underline{\theta}$ est le niveau de fondamentaux tel que les investisseurs vont racheter à la date 1, n'importe quelles décisions que les autres ont fait. Ainsi,

$$R_2(\underline{\theta}) = 1 \quad (7-B)$$

Symétriquement, $\bar{\theta}(R_1)$ est le niveau de fondamentaux tel que les investisseurs vont rester, n'importe quelles décisions que les autres ont fait. De manière pareil, on obtient:

$$R_2(\bar{\theta}) = \frac{1 - \max\{0, (\bar{N} - I(R_1))\}}{1 - (1 + \lambda)\max\{0, (\bar{N} - I(R_1))\}} \quad (8-B)$$

Nous définissons ensuite \bar{R}_1 tel que $I(\bar{R}_1) = \bar{N}$, on a:

$$\bar{\theta}(R_1) > \underline{\theta} \text{ si } R_1 < \bar{R}_1$$

$$\bar{\theta}(R_1) = \underline{\theta} \text{ si } R_1 \geq \bar{R}_1$$

(9-B)

Sous l'équilibre, tous les investisseurs rachètent à la date 1 quand $\theta < \underline{\theta}$, par contre, tout le monde reste jusqu'à la date 2 quand $\theta > \bar{\theta}(R_1)$. Quand θ se situe entre ces deux seuils, il existe deux équilibres: dans un équilibre tout le monde reste et dans un autre tout le monde sort.

En appliquant les techniques de jeux globaux, on peut trouver un seuil θ^* . Si les investisseurs observent un niveau des fondamentaux à l'inférieur de cette valeur, ils vont tous sortir à la date 1; Si le niveau des fondamentaux est supérieur à cette valeur, tout le monde reste. Ainsi, θ^* est le niveau des fondamentaux tel qu'un investisseur est indifférent de rester ou de sortir. θ^* satisfait l'équation suivante:

$$R_2(\theta^*) = \frac{1}{\int_{-0}^1 \frac{1-(1+\lambda)\max\{0,(\alpha\bar{N}-I(R_1))\}}{1-\max\{0,(\alpha\bar{N}-I(R_1))\}} d\alpha} \quad (12-B)$$

Cette équation donne le résultat principal du modèle: quand la performance est mauvaise, i.e. $R_1 < \bar{R}_1$, le seuil des fondamentaux θ^* augmente avec λ et décroît avec R_1 . Cela signifie que les fonds illiquides (qui ont un plus grand λ) souffrent plus de sorties.

Bibliographie générale

- Agarwal, V., Daniel, N., and Naik, N. (2004). Flows, performance, and managerial incentives in hedge funds. *SSRN Working Paper*.
- Agarwal, V., Green, C., and Ren, H. (2015). Alpha or beta in the eye of the beholder: What drives hedge fund flows. *SSRN Working Paper*.
- Aiken, A., Clifford, C., and Ellis, J. (2015). Hedge funds and discretionary liquidity restrictions. *Journal of Financial Economics*.
- Akbas, F., Armstrong, W., Sorescu, S., and Subrahmanyam, A. (2015). Smart money, dumb money, and capital market anomalies. *Journal of Financial Economics*, 118(2):355 – 382.
- Aragon, G., Ergun, T., Getmansky, M., and Girardi, G. (2017). Hedge fund liquidity management. *SSRN Working Paper*.
- Aragon, G. and Strahan, P. (2012). Hedge funds as liquidity providers: Evidence from the lehman bankruptcy. *Journal of Financial Economics*.
- Ashcraft, A. and Duffie, D. (2007). System illiquidity in the federal fund market. *American Economic Review*.
- Aydogdu, M. and Wellman, J. (2011). The effects of advertising on mutual fund flows: Results from a new database. *Financial Management*.
- Bailey, W., Kumar, A., and Ng, D. (2011). Behavioral biases of mutual fund investors. *Journal of Financial Economics*, 102(1):1 – 27.
- Banegas, A., Gillen, B., Timmermann, A., and Wermers, R. (2013). The cross section of conditional mutual fund performance in european stock markets. *Journal of Financial Economics*.
- Baquero, G. and Verbeek, M. (2005). A portrait of hedge fund investors: Flows, performance and smart money. *SSRN Working Paper*.

- Barber, B., Huang, X., and Odean, T. (2016). Which factors matter to investors? evidence from mutual fund flows. *Review of Financial Studies*, 29(10):2600–2642.
- Becker, W. and Kennedy, P. (1992). A graphical exposition of the ordered probit. *Econometric Theory*.
- Ben-Rephael, A., Da, Z., and Israelsen, R. (2017). It depends on where you search: Institutional investor attention and underreaction to news. *Review of Financial Studies*.
- Ben-Rephael, A., Kandel, S., and Wohl, A. (2012). Measuring investor sentiment with mutual fund flows. *Journal of Financial Economics*, 104(2):363 – 382. Special Issue on Investor Sentiment.
- Bergstresser, D. and Poterba, J. (2002). Do after-tax returns affect mutual fund inflows? *Journal of Financial Economics*, 63(3):381 – 414.
- Berk, J. and Binsbergen, J. V. (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics*, 119(1):1 – 23.
- Berk, J. and Green, R. (2004). Mutual fund flows and performance in rational markets. *Journal of Political Economy*.
- Berk, J. and Tonks, I. (2007). Return persistence and fund flows in the worst performing mutual funds. *SSRN Working Paper*.
- Blackrock (2016). Understanding liquidity fees and redemption gates. *Blackrock white papers*.
- Bollen, N. and Busse, J. (2004). Short-term persistence in mutual fund performance. *Review of Financial Studies*.
- Brown, D. and Wu, Y. (2016). Mutual fund flows and cross-fund learning within families. *Journal of Finance*, 71(1):383–424.
- Brunnermeier, M. and Pedersen, L. H. (2009). Market liquidity and funding liquidity. *Review of Financial Studies*, 22(6):2201–2238.
- Busse, J. (1999). Volatility timing in mutual funds: Evidence from daily returns. *Review of Financial Studies*, 12(5):1009–1041.

- Cao, C., Chang, E., and Wang, Y. (2008). An empirical analysis of the dynamic relationship between mutual fund flow and market return volatility. *Journal of Banking and Finance*, 32(10).
- Cao, C. and Petrasek, L. (2014). Liquidity risk and institutional ownership. *Journal of Financial Markets*, 21:76 – 97.
- Cao, C., Simin, T., and Wang, Y. (2013). Do mutual fund managers time market liquidity? *Journal of Financial Markets*, 16(2):279 – 307.
- Carhart, M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82.
- Carroll, T. and Niehaus, G. (1998). Pension plan funding and corporate debt ratings. *Journal of Risk and Insurance*.
- Cashman, G., Deli, D., Nardari, F., and Villupuram, S. (2008). Understanding the non-linear relation between mutual fund performance and flows. *SSRN Working Paper*.
- Chalmers, J., Edelen, R., and Kadlec, G. (2001). On the perils of financial intermediaries setting security prices: The mutual fund wild card option. *Journal of Finance*, 56(6):2209–2236.
- Chang, T., Solomon, D., and Westerfield, M. (2016). Looking for someone to blame: Delegation, cognitive dissonance, and the disposition effect. *Journal of Finance*.
- Chaudhuri, R., Ivkovic, Z., Pollet, J., and Trzcinka, C. (2016). What a difference a ph.d. makes: More than three little letters. *SSRN Working Paper*.
- Chen, Q., Goldstein, I., and Jiang, W. (2010). Payoff complementarities and financial fragility: Evidence from mutual fund outflows. *Journal of Financial Economics*, 97(2):239–262.
- Chen, Y. and Nan, Q. (2017). The behavior of investor flows in corporate bond mutual funds. *Management Science*.
- Chernenko, S. and Sunderam, A. (2014). Frictions in shadow banking: Evidence from the lending behavior of money market mutual funds. *Review of Financial Studies*.
- Chernenko, S. and Sunderam, A. (2016). Liquidity transformation in asset management: Evidence from the cash holdings of mutual funds. *SSRN Working Paper*.

- Cheung, S. (1996). Provincial credit ratings in Canada: An ordered probit analysis. *SSRN Working Paper*.
- Chevalier, J. and Ellison, G. (1999). Career concerns of mutual fund managers. *Quarterly Journal of Economics*, 114(2):389–432.
- Christoffersen, S. and Simutin, M. (2017). On the demand for high-beta stocks: Evidence from mutual funds on the demand for high-beta stocks: Evidence from mutual funds. *Review of Financial Studies*.
- Christoffersen, S. and Xu, H. (2017). Investor attrition and fund flows in mutual funds. *Journal of Financial and Quantitative Analysis*.
- Copper, M., Gulen, H., and Rau, R. (2006). Changing names with style: Mutual fund name changes and their effects on fund flows. *Journal of Finance*.
- Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*.
- Cox, R., Kimmel, R., and Wang, G. (2017). Proportional hazards model of bank failure: Evidence from USA. *Journal of Economic and Financial Studies*.
- Cremers, M. and Pareek, A. (2016). Patient capital outperformance: The investment skill of high active share managers who trade infrequently. *Journal of Financial Economics*.
- Cuthbertson, K., Nitzsche, D., and Sullivan, N. (2016). A review of behavioural and management effects in mutual fund performance. *International Review of Financial Analysis*.
- Dahlquist, M., Martinez, J. V., and Söderlind, P. (2017). Individual investor activity and performance. *Review of Financial Studies*, 30(3):866–899.
- Darolles, S., Ggliardini, P., and Gouriéroux, C. (2013). Survival of hedge funds: Frailty versus contagion. *SSRN Working Paper*.
- Darolles, S. and Roussellet, G. (2018). Hedge fund portfolio management with illiquid assets. *Unpublished Working Paper*.
- DeGroot, M. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.
- Desmettre, S. and Deege, M. (2016). Modeling redemption risks of mutual funds using extreme value theory. *Journal of Risk*.

- Diamond, D. and Dybvig, P. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*.
- Ding, B., Getmansky, M., Liang, B., and Wermers, R. (2007). Hedge fund flows and contagion in financial markets. *SSRN Working Paper*.
- Dong, X., Feng, S., and Sadka, R. (2017). Liquidity risk and mutual fund performance. *Management Science*.
- Duffie, D., Eckner, A., Horel, G., and Saita, L. (2009). Frailty correlated default. *Journal of Finance*.
- Dungey, M., Milunovich, G., SusanThorp, and Yang, M. (2015). Endogenous crisis dating and contagion using smooth transition structural garch. *Journal of Banking and Finance*.
- Edelen, R. (1999). Investor flows and the assessed performance of open-end mutual funds. *Journal of Financial Economics*, 53(3):439 – 466.
- Edelen, R., Ince, O., and Kadlec, G. (2016). Institutional investors and stock return anomalies. *Journal of Financial Economics*.
- Edwards, F. (1999). Hedge funds and the collapse of long-term capital management. *Journal of Economic Perspectives*.
- Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, 81(395):709–721.
- Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica*, 68(1):1–22.
- Fang, M., Lin, B., and Yang, G. (2015). Who drives fund performance, manager or investor? *SSRN Working Paper*.
- Fecht, F. and Wedow, M. (2014). The dark and the bright side of liquidity risks: Evidence from open-end real estate funds in germany. *Journal of Financial Intermediation*, 23(3):376 – 399.
- Foster, D. and Warren, G. (2016). Interviews with institutional investors: The how and why of active investing. *Journal of Behavioral Finance*.
- Franzoni, F. and Schmalz, M. (2017). Fund flows and market states. *Review of Financial Studies*.

- Frazzini, A. and Lamont, O. (2008). Dumb money: Mutual fund flows and the cross-section of stock returns. *SSRN Working Paper*.
- Friesen, G. C. and Sapp, T. R. (2007). Mutual fund flows and investor returns: An empirical examination of fund investor timing ability. *Journal of Banking and Finance*, 31(9):2796 – 2816.
- Getmansky, M., Liang, B., Schwarz, C., and Wermers, R. (2015). Share restrictions and investor flows in the hedge fund industry. *SSRN Working Paper*.
- Goetzmann, W., Ingersoll, J., and Ross, S. (2003). High water marks and hedge fund management contracts. *Journal of Finance*.
- Goldstein, I., Jiang, H., and Ng, D. (2017). Investor flows and fragility in corporate bond funds. *Journal of Financial Economics*.
- Grinbatt, M. and Keloharju, M. (2001). What makes investors trade? *Journal of Finance*.
- Guercio, D. D. and Tkac, P. (2002). The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds. *Journal of Financial and Quantitative Analysis*, 37(4):523–557.
- Guercio, D. D. and Tkac, P. (2008). Star power: The effect of morningstar ratings on mutual fund flow. *Journal of Financial and Quantitative Analysis*, 43(4):907–936.
- Hau, H. and Lai, S. (2016). The role of equity funds in the financial crisis propagation. *Review of Finance*.
- Hill, J. (2010). On tail index estimation for dependent, heterogeneous data. *Econometric Theory*.
- ICI (2017). 2017 investment company factbook. *Investment Company Institute Report*.
- Ippolito, R. (1992). Consumer reaction to measures of poor quality: Evidence from the mutual fund industry. *Journal of Law and Economics*, 35(1):45–70.
- Ivkovich, Z. and Weisbenner, S. (2008). Individual investor mutual fund flows. *Journal of Financial Economics*.
- Iyer, R. and Puri, M. (2012). Understanding bank runs: The importance of depositor-bank relationships and networks. *American Economic Review*, 102(4):1414–45.

- James, C. and Karceski, J. (2006). Investor monitoring and differences in mutual fund performance. *Journal of Banking and Finance*, 30(10):2787 – 2808.
- Jank, S. and Wedow, M. (2010). Purchase and redemption decisions of mutual fund investors and the role of fund families. *SSRN Working Paper*.
- Johnson, W. (2004). Predictable investment horizons and wealth transfers among mutual fund shareholders. *Journal of Finance*, 59(5):1979–2012.
- Johnson, W. (2007). Who monitors the mutual fund manager, new or old shareholders? *SSRN Working Paper*.
- Johnson, W. (2010). Do investors trade uniformly through time? *Journal of Empirical Finance*, 17(4):645 – 658.
- Jordan, B. and Riley, T. (2015). Volatility and mutual fund manager skill. *Journal of Financial Economics*.
- Kacperczyk, M. and Schnabl, P. (2013). How safe are money market funds? *Quarterly Journal of Economics*.
- Kamstra, M., Kramer, L., Levi, M., and Wermers, R. (2017). Seasonal asset allocation: Evidence from mutual fund flows. *Journal of Financial and Quantitative Analysis*.
- Kempf, A. and Ruenzi, S. (2008). Tournaments in mutual-fund families. *Review of Financial Studies*, 21(2):1013–1036.
- Keswani, A. and Stolin, D. (2008). Which money is smart? mutual fund buys and sells of individual and institutional investors. *Journal of Finance*, 63(1):85–118.
- Khandania, A. and Lo, A. (2011). What happened to the equants in august 2007? evidence from factors and transactions data. *Journal of Financial Markets*.
- Konigsberg, L., Herrmann, N., Wescott, D., and Kimmerle, E. (2008). Estimation and evidence in forensic anthropology: Age-at-death. *Forensic Sciences*.
- Kumar, A., Niessen-Ruenzi, A., and Spalt, O. (2015). What’s in a name? mutual fund flows when managers have foreign-sounding names*. *Review of Financial Studies*.
- Landon, S. and Smith, C. (2000). Government debt spillovers and creditworthiness in a federation. *Canadian Journal of Economics*.

- Leung, D. and Kwong, M. (2018). The flow-performance relationship in emerging market bond funds. *SSRN Working Paper*.
- Levy, A. and Liberman, O. (2015). Active flows and passive returns. *Review of Finance*.
- Liang, B. and Park, H. (2010). Predicting hedge fund failure: A comparison of risk measures. *Journal of Financial and Quantitative Analysis*.
- Liao, L., Zhangb, X., and Zhang, Y. (2017). Mutual fund managers' timing abilities. *Pacific-Basin Finance Journal*.
- Liesenfeld, Nolte, and Pohlmeier (2006). Modelling financial transaction price movements: adynamicintegercountdatamodel. *Recent Developments in High Financial Econometrics*.
- Lou, D. (2014). Attracting investor attention through advertising. *Review of Financial Studies*.
- Lowenstein, R. (2000). *When Genius Failed: The Rise and Fall of Long-Term Capital Management*. Random House.
- Ma, L. (2013). Mutual fund flows and performance: A survey of empirical findings. *SSRN Working Paper*.
- Myrland, O., Trondsena, T., Johnstonb, R., and Lund, E. (2000). Determinants of seafood consumption in norway: lifestyle, revealed preferences, and barriers to consumption. *Food Quality and Preference*.
- Nanda, V., Narayanan, and Warther, V. (2000). Liquidity, investment ability, and mutual fund structure. *Journal of Financial Economics*.
- Odean, T. (1999). Do investors trade too much? *American Economic Review*, 89(5):1279–1298.
- Pastor, L., Stambough, R., and Taylor, L. (2017). Do funds make more when they trade more? *Journal of Finance*.
- Paula, J. and Croce, R. D. (2016). Annual survey of large pension funds and public pension reserve funds. *ODCE Public Report*.
- Peng, C.-Y. J., So, T.-S. H., Stage, F., and John, E. (2002). The use and interpretation of logistic regression in higher education journals: 1988-1999. *Research in Higher Education*.

- Perold, A. (1986). *Constant Proportion Portfolio insurance*. Harvard Business School.
- Puy, D. (2016). Mutual funds flows and the geography of contagion. *Journal of International Money and Finance*.
- Rohleder, M., Schulte, D., and Wilkens, M. (2017). Management of flow risk in mutual funds. *Review on Quantitative Finance and Accounting*.
- Sadka, R. (2010). Liquidity risk and the cross-section of hedge-fund returns. *Journal of Financial Economics*, 98(1):54 – 71.
- Schmidt, L., Timmermann, A., and Wermers, R. (2016). Runs on money market mutual funds. *American Economic Review*, 106(9):2625–57.
- Shek, J., Shim, I., and Shin, H. S. (2017). Investor redemptions and fund manager sales of emerging market bonds: how are they related? *Review of Finance*.
- Sherman, M. G. (2012). The life cycle of hedge funds: Fund flows, size, competition, and performance. *Quarterly Journal of Finance*.
- Sirri, E. and Tufano, P. (1998). Costly search and mutual fund flows. *Journal of Finance*.
- Solomon, D., Soltés, E., and Sosyura, D. (2014). Winners in the spotlight: Media coverage of fund holdings as a driver of flows. *Journal of Financial Economics*.
- Spiegel, M. and Zhang, H. (2013). Mutual fund risk and market share-adjusted fund flows. *Journal of Financial Economics*, 108(2):506 – 528.
- Tchamyou, V. and Asongu, S. (2017). Conditional market timing in the mutual fund industry. *Research in International Business and Finance*.
- Teo, M. (2011). The liquidity risk of liquid hedge funds. *Journal of Financial Economics*.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*.
- Zhiguo, H. and Manela, A. (2016). Information acquisition in rumor-based bank runs. *Journal of Finance*, 71(3):1113–1158.
- Zhu, F. (2012). Modeling time series of counts with com-poisson ingarch models. *Mathematical and Computer Modelling*.

Résumé

Cette thèse étudie le comportement des investisseurs des fonds mutuels ouverts et ses implications au risque de liquidité. Ces travaux de recherche ont pour objectif d'aider les gérants de fonds à éviter le scénario de "fund run" où ils perdent leurs clients de manière soudaine. La première étape de cette étude est de collecter une nouvelle base de données qui enregistre les "micro-transactions" des investisseurs. Cela nous permet d'analyser leurs comportements au niveau individuel et d'effectuer ces trois articles de recherche autour de ce sujet. Dans le premier article, nous développons un modèle de comptage auto-excitant qui capture des faits stylisés des séries des flux du fonds. De là, nous montrons un risque lié au passif du fonds qui est différent de celui lié à l'actif déjà documenté dans la littérature. Nous identifions également une contagion des chocs de liquidité entre les différents clients dans un même fonds. Dans le chapitre suivant, nous étudions les horizons d'investissement des clients individuels. Ces horizons sont fortement liés aux caractéristiques des investisseurs et aux conditions économiques. Nous montrons également que les gérants de fonds subissent un risque de sortie prématurée relatif au raccourcissement des horizons d'investissement de ses clients. Nous observons ensuite une hétérogénéité entre les investisseurs: ceux de long-terme comportent différemment que ceux de court-terme. Enfin, dans le dernier chapitre, nous nous intéressons aux activités de rééquilibrage. Nous trouvons que de nombreux investisseurs détiennent un portefeuille contenant plusieurs fonds et le rééquilibrent afin de garder la même allocation d'actifs.

Mots Clés

Liquidité de financement, fonds ouverts, panique financière, micro-transactions

Abstract

This thesis studies the behaviour of investors in open-end mutual funds and its implications to the liquidity risk. We seek to help the fund managers to avoid the "fund run" scenarios where they lose their clients in a sudden way. We begin our works by collecting a unique data set which records the micro-transactions of fund investors. It allows us to monitor investors' behaviour at the individual level and to accomplish three research articles around this topic. In the first article, we develop a self-exciting counting process to model the stylized facts of fund flows. Therefrom, we highlight a novel risk linked to the fund liability which is different than the asset-related risk documented by the previous literature. We also identify a liquidity contagion among investors in a same fund. In the next chapter, we study the dispersion in the investing horizons of individual fund clients. These horizons are strongly determined by investors' characteristics and economic conditions. We show that the fund managers suffer a pre-mature redemption risk, i.e. clients shorten their investing horizons and redeem pre-maturely. Especially, we observe a heterogeneity among investors: long-term ones bring a higher pre-mature redemption risk. In the last chapter, we are interested in the rebalance behaviour. We find that numerous investors hold a multi-funds portfolio and rebalance it to keep the target asset allocation.

Keywords

Funding liquidity, open-end funds, fund run, micro-transactions