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

Serge Darolles*
serge.darolles@dauphine.fr

Gaëlle Le Fol*
gaelle.le_fol@dauphine.fr

Yang Lu**
luyang000278@gmail.com

Théo Sun*
theosunran@gmail.com

June 2018

**Abstract**

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 capturing fund flows' clustering effects and over-dispersion. The model improves the forecast accuracy of future flows and provides a reliable risk indicator (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 with economic conditions, investor's behaviour stays the main significant determinant of flows' randomness. Our findings encourage fund manager to adopt an ALM approach.

**Keywords:** Financial Econometrics, Hedge Funds/Mutual Funds, Market Microstructure/Liquidity.

*Université Paris-Dauphine, PSL Research University, CNRS, DRM, Finance, 75016 Paris, France.
**THEMA - Université de Cergy-Pontoise, France, 33 Bd du Port 95011 Cergy-Pontoise cedex.

# 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 funds[1] 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 built 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

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

investors. Lastly, we study the liability risk under 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 transations 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. **?** 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, **?**). This problem is even more acute in illiquid fund classes. **?** highlight the fragility of corporate bond funds with low liquidity reserve. **?** 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. **?** 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 (**??**), fund risk level (**?**), market volatility level (**?**), fund companies'

marketing expense (?), etc. Furthermore, psychological factors might also affect fund flows. ? document the significant influence of fund manager's name and reputation on fund flows. ? 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 framework[2] and ignore the dynamic patterns of fund flows. ? document that fund investors exhibit a time varying behaviour, and fund flows evolve accordingly. Furthermore, ? 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. ? find that institutional investors behave differently from retail investors and exhibit a smart money effect. ? show that some investors can detect the mis-pricing of mutual funds and profit from this arbitrage opportunity. ? document that mutual funds investors from the same geographical zone might suffer from the correlated liquidity shocks, which generate correlated redemptions. ? examine the timing ability of fund investors. ? 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.

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

# 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.

## 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 guide[3] 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.

## 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

[3]See, "Journal officiel de l'Union européenne: Règlement (UE) 2017/1131 du Parlement Européen et du Conseil, du 14 juin 2017 sur les fonds monétaires"

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 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 categories[4] 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., **?**, **?**)[5], disaggregated monthly inflows/outflows (e.g., **?**, **?**), and recently daily or even ultra-frequency flows[6] (e.g., **?**, **?**). 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

[4]We 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".

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

[6]**?** refer the ultra-frequency the datas which record information of every transaction, in an order book or in other fields.

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 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 investors[7], 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, **?** 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

[7]The 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.

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. 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.

### 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 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 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., **?**; **?**). 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., **?**). 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 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.

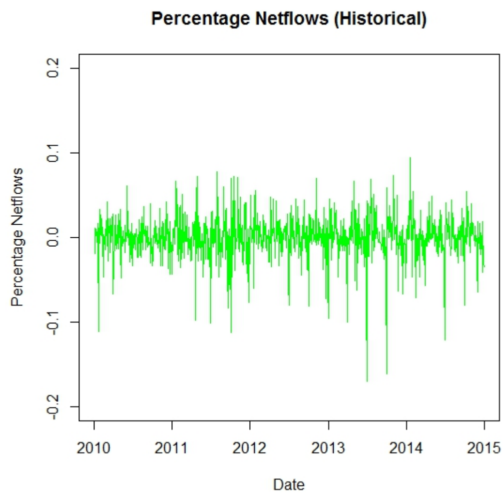
For the simplicity of presentation, we first take Fund 7 as an illustrative example before presenting the results for all other Funds later on. Fund 7 is a money market fund managed by Company B. As other funds in the same category, investors are very active and a significant number of flows occurs every 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 1). The next step is to see if we can gain from separating in and outflows as in Figure 2.

Table 2: Summary statistics for fund flows

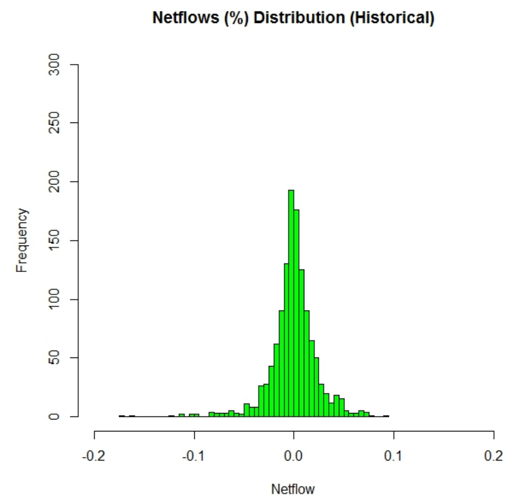
| Fund | Category | Period | N Days | Number | | | Freq | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | *Inflows* | *Outflows* | *All* | *Inflows* | *Outflows* | *All* |
| **Company A** | | | | **41 358** | **59 536** | **100 894** | | | |
| Funds 1 | EquityLargeCap | 2013-14 | 497 | 521 | 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** 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.

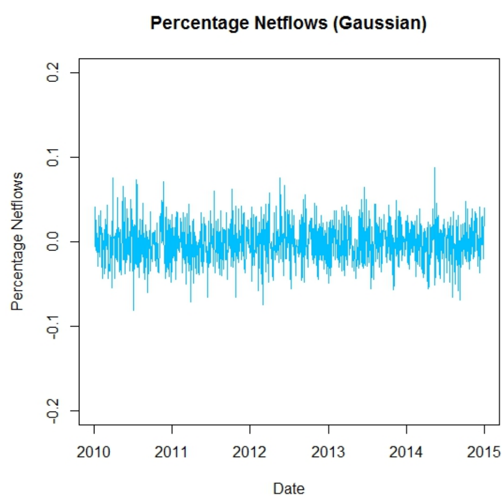
Figure 1: Inaccurate fits of traditional iid model



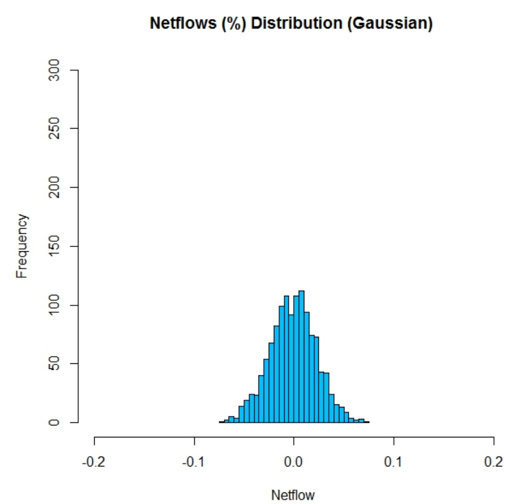

(a) Historical percentage netflow time series




(b) Distribution of historical percentage netflow time series



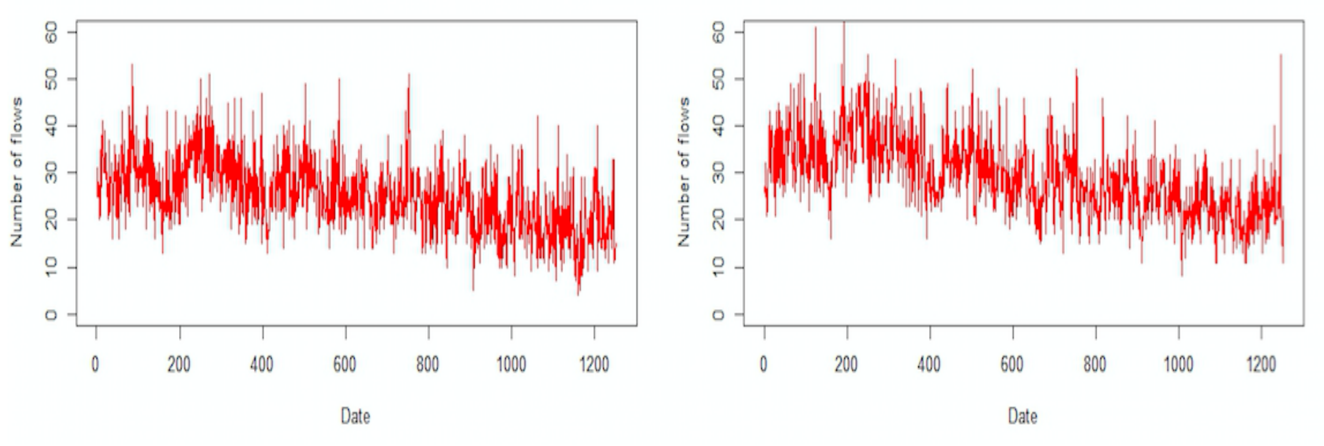

(c) Simulation: iid gaussian time series




(d) Distribution of iid gaussian simulated series

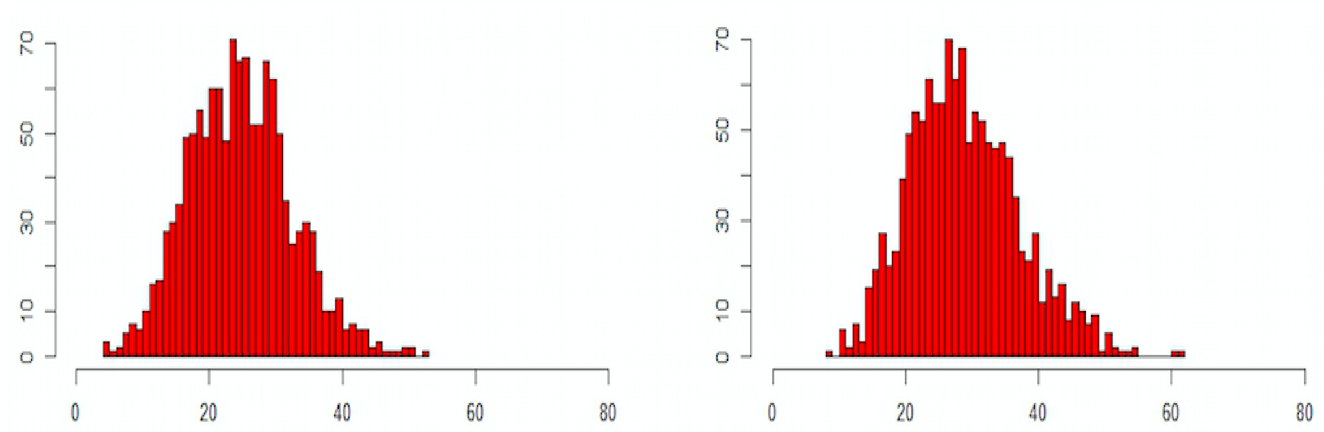
**Figure 1** shows the comparison between the historical data series and a Toy model. We first convert the net-flow series (inflows minus outflows) into percentage value by dividing them by fund's size. We present net-flows' time series for a 5 year-period, from 2010 to 2014, in Sub-Figure (a). Then 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: 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 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 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 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 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 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.

## 2.1 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 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. We handle overdispersion with a generalized Poisson model where 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 $\lambda$ represents the average number of flow arrivals; $s$ is the parameter which generates several 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 $\lambda$ and we get over-dispersion. On the contrary, if $s$ is larger than 1, the flow variance is smaller than the flow average, and we have under-dispersion. When $s$ equals to 1, the distribution reduces to a homogeneous Poisson distribution[8].

The estimation of the two parameters $\lambda$ 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

[8]See, **?** for details of the Double Poisson distribution. There are other choices to generate the over-dispersion property. See, **?** for other alternative choices.

redemptions cannot be captured using a static representation and we have to include a dynamic component.

### 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 parameter of the Poisson law $\lambda$ by a parameter $\lambda_t$ which varies in time as a function of the previous observation of the time series. The specification chosen is the following:

$$\lambda_t = \lambda_0 + \rho N_{t-1}, \tag{1}$$

where $\lambda_0$ is a constant intensity and the additional $\rho$ 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 $\lambda_t$, 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 $\rho$. Our estimation results show that for the vast majority of funds, the persistence parameters $\rho$ 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 $\rho$ 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.

## 2.4 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 $t$, $\lambda_t^{in}$ and the one describing the redemptions $\lambda_t^{out}$ satisfy the following two equations:

$$\lambda_t^{in} = \lambda_0^{in} + \rho^{in-in} N_{t-1}^{in} + \rho^{in-out} N_{t-1}^{out}, \tag{2}$$

$$\lambda_t^{out} = \lambda_0^{out} + \rho^{out-in} N_{t-1}^{in} + \rho^{out-out} N_{t-1}^{out}, \tag{3}$$

where the two parameters $\rho^{in-out}$ and $\rho^{out-in}$ 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 $\rho^{\cdot\cdot-\cdot\cdot}$ included in the most general version of the model. $\rho^{in-in}$ 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. $\rho^{in-out}$ 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, $\rho^{out-in}$ 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 $\rho$ parameters are zero and $\lambda_0^{in} = \lambda_0^{out}$, we get back to the simple (univariate) Poisson distribution. When $\rho^{in-out}$ and $\rho^{out-in}$ are null, we have the over-dispersed persistent Poisson representation with no cross effect.

In a matrix format,we have :

$$\begin{pmatrix} \lambda_t^{in} \\ \lambda_t^{out} \end{pmatrix} = \begin{pmatrix} \lambda_0^{in} \\ \lambda_0^{out} \end{pmatrix} + \begin{pmatrix} \rho^{in-in} & \rho^{in-out} \\ \rho^{out-in} & \rho^{out-out} \end{pmatrix} \times \begin{pmatrix} N_{t-1}^{in} \\ N_{t-1}^{out} \end{pmatrix} \tag{4}$$

# 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 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 3.2 and then on disaggregated flows in Subsection 3.3. The objective there is to exploit information from disaggregated data to check whether different investors types result into different investors behaviours.

## 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)-(3) when applied to funds belonging to different categories. We then estimate and compare estimation results for funds within and between categories. Table 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 4: Count model estimation for 12 selected funds

| **Fund & category** | **Company** | $\lambda_0^{in}$ | $\lambda_0^{out}$ | $\rho^{in-in}$ | $\rho^{in-out}$ | $\rho^{out-in}$ | $\rho^{out-out}$ | $S^{in}$ | $S^{out}$ |
|---|---|---|---|---|---|---|---|---|---|
| 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 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 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.

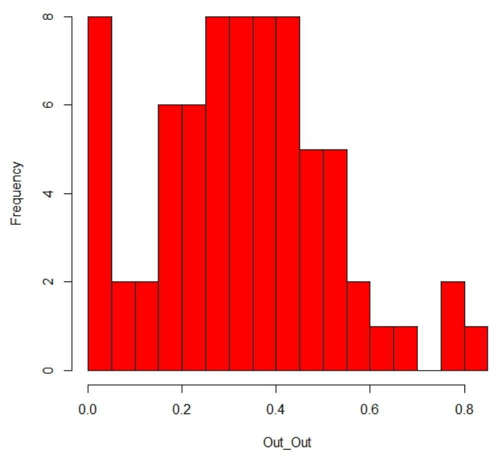
Turning now to dynamic effects, we can see that the autocorrelation of flows vary also quite a lot. On the one hand, the highest $\rho^{in-in}$ reaches 0.56 (Funds 10) and the highest $\rho^{out-out}$ coefficient is 0.65 (Funds 3). These results are directly related to the clustering behaviour in funds trading observed in Section 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 $\rho^{in-in}$ 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 $\lambda_0$. 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.

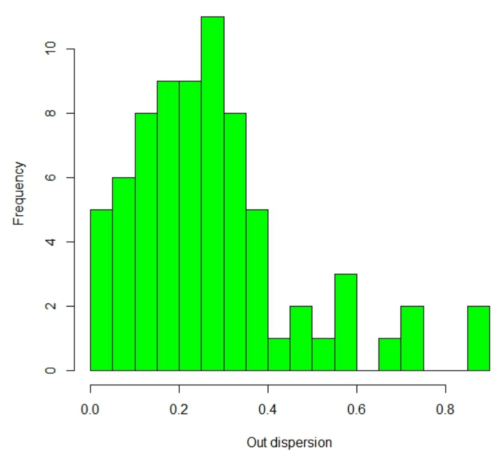
To further confirm these results on "liability risk", we estimate the parameters in a larger sample. We consider all large-cap equity and fixed incomes funds data of Company B in 2014, ending up with 51 equity funds and 23 fixed income funds in total. We estimate Equations (2)-(3) and present the distribution of $\rho^{out-out}$ and $S^{out}$ in Figure 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.

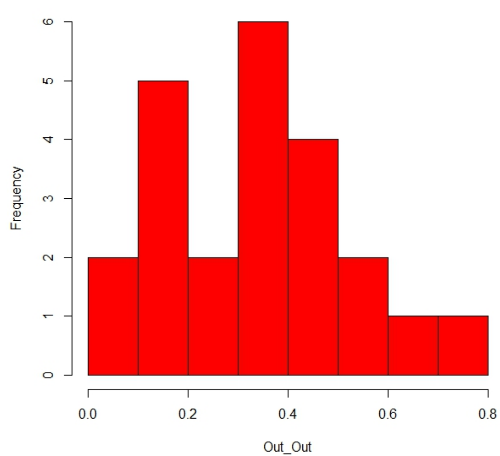
Figure 4: Distribution of Model's Parameters: Equity and Fixed Income Funds (an extended sample)



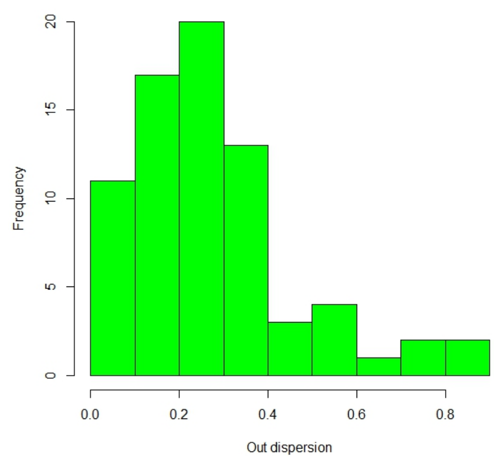

(a) Equity Out Out




(b) Equity Out Dispersion




(c) Fixed Income Out Out




(d) Fixed Income Out Dispersion

**Figure 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.

### 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. (4) page 18 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 $\rho^{in-out}$ and $\rho^{out-in}$ 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 $\rho$-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 $\lambda$ function.

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

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

$$\begin{pmatrix} \lambda_t^{in} \\ \lambda_t^{out} \end{pmatrix} = \begin{pmatrix} \lambda_0^{in} \\ \lambda_0^{out} \end{pmatrix} \tag{2D1}$$

$$\lambda_t = \lambda_0 \tag{1D}$$

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 5 [Resp. Table 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 5: Models' estimation (for the selected money market fund)

| **Specification** | $\lambda_0^{in}$ | $\lambda_0^{out}$ | $\rho^{in-in}$ | $\rho^{in-out}$ | $\rho^{out-in}$ | $\rho^{out-out}$ | $S^{in}$ | $S^{out}$ |
|---|---|---|---|---|---|---|---|---|
| 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 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 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 |

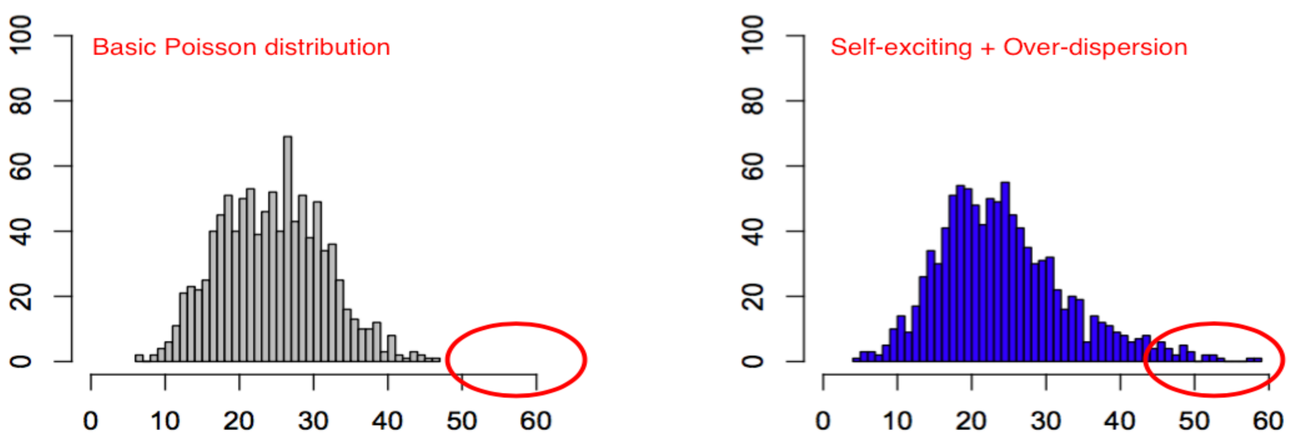
This table shows the summary of models' quality. We count the number of parameters (N) in each model (Specification) and calculate the Mean Square Errors for inflows, outflows. Then we sum two MSE up to obtain a quality ratio for the full model (MSE-All = MSE-in + MSE-out). Lastly we compute the AIC 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. (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 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 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.

## 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.

### 3.3.1 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, **?**). The second possible explanation comes from the existence of common factors which make fund flows correlated (e.g., **?**). 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

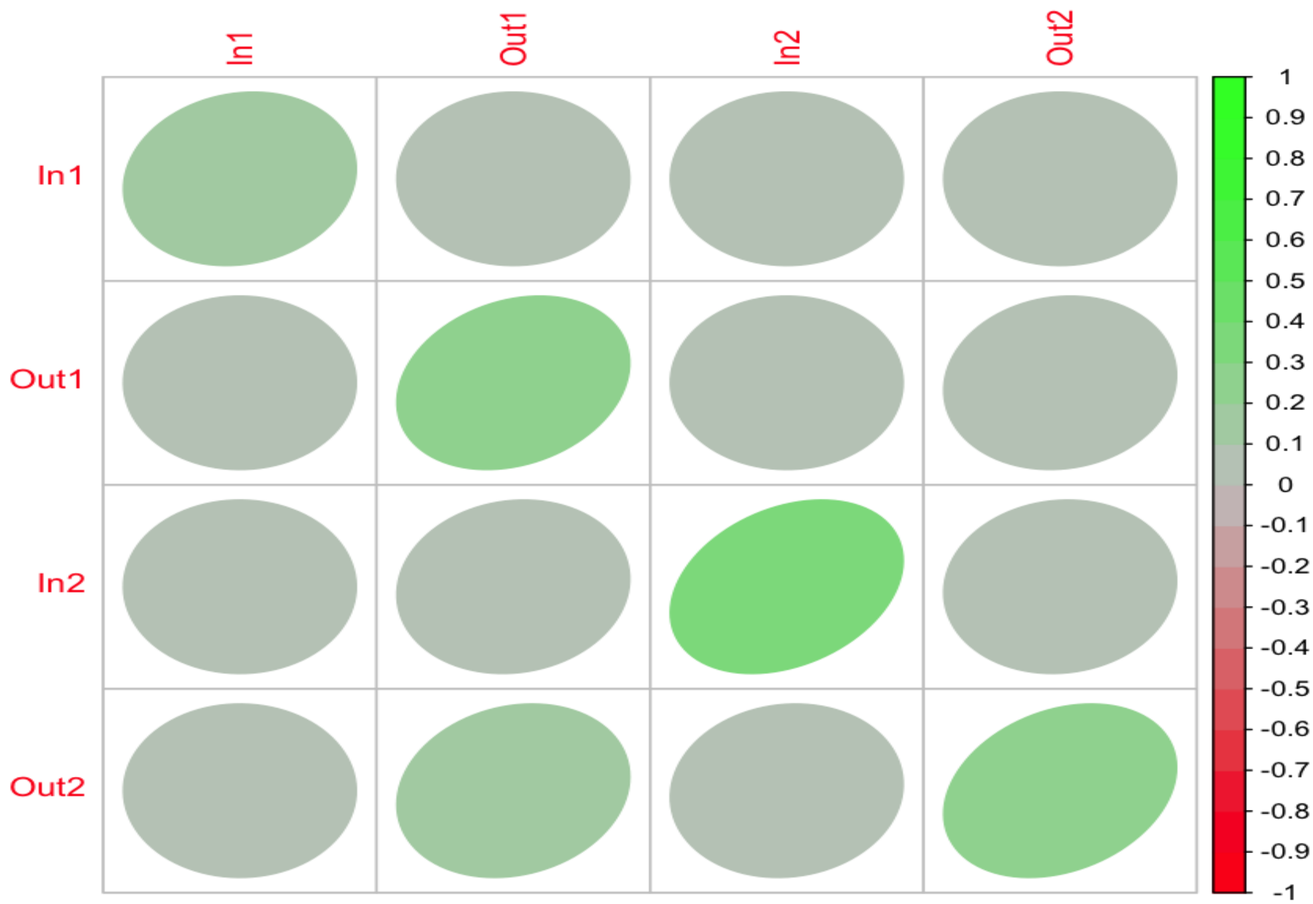
to the full model to highlight the lead-lag effect. The modified model becomes:

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

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 $\lambda_t^{in1}$ is composed by the baseline intensity $\lambda_0^{in1}$, 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 $4\times4 = 16$ correlation parameters. In the following, we use a covariance matrix representation to plot model's estimation in Figure 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 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 ($\rho$). The color indicates the value of the parameter: darker the green is, larger the coefficient is.

### 3.3.2 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 ($\lambda_0$) indicates the activism level of the sector; (2) the Self-Exciting ($\rho^{In-In}$ 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 $\rho^{In-All}$ 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 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. $\lambda_0$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. $\rho^{In-In}$ 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 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.

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

Table 7: Disaggregated sector model

| **Sector** | $\lambda_0^{in}$ | $\lambda_0^{out}$ | $\rho^{in-in}$ | $\rho^{in-all}$ | $\rho^{out-out}$ | $\rho^{out-all}$ | **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 ($\rho$) 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 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 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 leaded 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.

# 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., **?**). "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 "$\lambda_t^{Factor}$" 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. $\beta$ 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 $e^{\beta'_{In/Out}X}$. 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 $\lambda_0$. 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 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 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 |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 | |

**Table 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.

# 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 collect 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 high-dimensional statistic tools might be useful to analyse the even larger flow data. The other way is to examine the individual investor account. An investigation on investors' personal account might bring clearer evidence on their preference and habit on fund investing.