IMT School for Advanced Studies, Lucca
Lucca, Italy

Essays on financial stability: old and new risk sources

PhD Program in Economics, Network, Business Analytics
XXXIV Cycle

By

Ilaria Gianstefani

2023

The dissertation of Ilaria Gianstefani is approved.

PhD Program Coordinator: Massimo Riccaboni, IMT School for Advanced Studies Lucca

Advisor: Prof. Irene Crimaldi, IMT School for Advanced Studies Lucca

Co-Advisor: Prof. Roberto Renò, ESSEC Business School

The dissertation of Ilaria Gianstefani has been reviewed by:

,

,

IMT School for Advanced Studies Lucca
2023

Contents

Introduction

European Central Bank (ECB) defines financial stability[1] as "a condition in which the financial system – which comprises financial intermediaries, markets and market infrastructures – can withstand shocks and unravel financial imbalances. This mitigates the prospect of disruptions in the financial intermediation process that are severe enough to impact real economic activity adversely." Practically speaking, stability is a balance among the agents participating in the financial environment: market participants weave relationships, creating dependencies and interconnections. The risks and vulnerabilities affecting one agent can impact many others, generating a cascade effect that propagates and might throw the system out of balance. Hence it is essential to identify all the potential sources of risk in the spirit that if we can recognize the form and assess the severity, we can cope with specific risks and prevent the system from unbalancing.

The most common risks threatening the financial system's stability are (see Vernimmen et al. (2014)):

- Market risk is exposure to unfavorable trends in product prices, interest rates, exchange rates, raw material prices, or stock prices;
- Credit (or counterparty risk) risk is the risk of loss on an outstanding debt that is not paid on time;
- Liquidity risk stems from the lack of marketability of an invest-

[1] The financial stability and macroprudential policy of ECB is available on the website at the link www.ecb.europa.eu/ecb/tasks/stability/html/index.en.html

ment that cannot be bought or sold quickly enough to prevent or minimize a loss; [2]

- Operational risk summarizes the uncertainties and hazards a company faces when it attempts to do its day-to-day business activities within a given field or industry;

- Political, regulatory, and legal risks impact investment returns and could suffer due to political changes or instability in a country. Instability affecting investment returns could stem from government, legislative bodies, foreign policymakers, or military control changes.

The last two decades have presented new challenges for financial stability, affecting it with very heterogeneous risk sources. Other sources of risk have developed due to social, economic, and technological changes, adding to those already known. Moreover, as recent episodes have highlighted, even events that are not strictly financially economic related can drastically impact financial stability: wars, pandemics, climate change, and technological innovations have shown how the financial system is vulnerable on unforeseen fronts.

This thesis focuses on analyzing, modeling, and assessing some of these risks and contributes to expanding knowledge on risk management that affects financial stability with the ultimate goal of providing insights to improve the security and operability of the financial system. In an always-changing world, it is essential to identify new potential sources of risk that might affect financial stability and improve the tools at our disposal to assess known ones. Specifically, this thesis deals with three of the risks mentioned above: the liquidity risk in the stock market, the market risk in the bond market, and finally, the analysis of one of the brand-new risks affecting the market, namely, the risk generated by the retail investors' coordination, also called noise trader risk.

[2] https://www.investopedia.com/terms/l/liquidityrisk.asp

Liquidity broadly refers to the ease with which an asset, or security, can be converted into ready cash without affecting its market price. Financial market microstructure models depict risk-averse investors with preferences for liquid assets: the liquidity premium is negatively priced as investors prefer assets that can be easily and quickly converted into cash at their fair market value. Investing in illiquid assets requires compensation for the risk of allocating money to assets that may not be able to be sold for an extended period at a fair value. Many empirical works confirm the result (see, for example, Tarun Chordia, Avanidhar Subrahmanyam, and V. R. Anshuman (2001) and references therein), which proves the negative relationship between the level of liquidity (usually proxied by the dollar volume) and the asset returns. On the contrary, what is still not completely understood in financial economics is whether liquidity uncertainty (usually measured as the volatility of dollar volume) grants a positive or negative premium. Theoretically speaking, two branches of literature present opposing views. On one side, Acharya and L.H. Pedersen (2005) claims the liquidity volatility premium to be positive as compensation for an agent holding stocks with uncertain transaction costs. On the other side, Joao Pereira and Harold Zhang (2010) justifies the negativity of the sign, seeing in the liquidity variability (fluctuations) an opportunity for the investor who times the trades according to the liquidity level. The liquidity uncertainty premium sign is also controversial at a practical level, and this thesis's first chapter goes deep into this. Inspired by the puzzle presented in the paper by Tarun Chordia, Avanidhar Subrahmanyam, and V. R. Anshuman (2001), which finds a negative premium for liquidity risk, in the first chapter of the thesis we replicate the analysis on a different dataset and propose new proxies for liquidity uncertainty based on high-frequency data. Our findings are extremely promising. We confirm the negative premium for liquidity risk when using a low-frequency computed proxy for liquidity uncertainty on a long window period. When we proxy liquidity uncertainty with one-minute trading data over the most recent months, we uncover a positive and significant relationship between liquidity uncertainty and returns. We show that the relationship between liquidity uncertainty and returns is not robust to how volume volatility is computed.

The second chapter of the thesis has its roots in a recent episode that has shed light on an unsuspected source of financial vulnerability: the GameStop frenzy generating the short squeeze at the end of January 2021. The newspapers claimed it was a battle around GameStop (GME) shares between the big players of the financial markets (reads, hedge funds) and an army of unsophisticated investors who wanted to beat the bigs. The hedge funds were short-selling[3] GME shares. The retail investors noticed it and started buying the stock massively, driving the price up and generating instability and substantial losses for those shorting the stocks. To provide some cover for the market-wide price impact due to the retail action, the hedge funds started closing their short positions (i.e., buying back the stocks), triggering a loop that contributed to the price increase. Here again, liquidity plays a key role: the higher the illiquidity of the implicated asset, the greater the vulnerability of the funds dealing with that asset. The financial system reveals its fragility, and the injuries arrived from apparent harmless players, the retail investors. Always relegated as a residual category in all market microstructure models, they demonstrate how not marginal they are when many single small investors pool together, lead, and coordinated by some "fanatics". The noise trade risk[4] is due to the decisions made by so-called noise traders - unskilled, uninformed, or novice retail traders that participate in the market and are mainly trend-following, emotional, and undisciplined. These traders can create price volatility and make irrational decisions or mistakes that can affect prices to the detriment of professional or well-informed traders. The first framework considering the retail traders and the associated risk they generate is the one proposed by L. Pedersen (2021). The two major innovations that favor this new trend are the fintech (r-) evolution and the development of virtual platforms (social media) pooling people together. Fintech has developed a

[3]A financial strategy that allowed them to make money when the stock performs poorly.
[4]The definition of noise trade risk can be found at the link https://www.investopedia.com/terms/n/noisetraderrisk

lot during the last few years, favoring the diffusion of low-cost, low-capital-requirement, user-friendly trading apps that simplify access to financial markets. The core of the second chapter investigates the crucial role of social media (specifically Reddit) in permitting such effective co-ordination to happen. Analyzing the row social media data, basically the posts and comments thread, we uncover how the retail movement put in place by unsophisticated investors was accurately coordinated by some users in the network. We design a social-media-data-fed alert system to detect potential advance turmoil that retail investors can generate in the financial system. The alert tool is based on unusual activities (in terms of volume) and advanced network analysis to spot the users coordinating the social movement.

Finally, the third chapter of the thesis deals with a very well-known risk in the financial literature: the market risk; when applied to the bond market, is also named interest rate risk. More precisely, the reported analysis, developed during an internship in the International Finance division in the directorate of Financial Stability at the Central Bank of Ireland (CBI), aims to assess the impact of an interest rate shock on the Irish domiciled bond funds. Since the global financial crisis back in 2008, the economic structure has changed significantly, with a notable expansion of the non-banking financial sector (also referred to as non-banking financial institutions or NBFI). Ireland was not an exception; the Irish NBFI sector has almost tripled over the past decade. This evolution has required the development of new regulations and tools to assess and monitor the new risk sources. This study represents a chuck of risk identification and assessment tool that CBI use to judge the risk associated with the bond investment funds critically. In particular, we run a series of stress test exercises to test the Irish-domiciled bond funds' shock-absorbing capacity when an exogenous shock on the market manifests. The activity highlights that the less resilient fund categories contain instruments with a long maturity, hence much more sensitive to market fluctuations.

The thesis ends with a section of conclusions, where the main findings and the potential evolution of the work are summarized.

Chapter 1

The liquidity uncertainty premium puzzle

The puzzling negative relation between liquidity uncertainty and asset returns, originally put forward by Chordia, Subrahmanyam, and Anshuman (2001) and confirmed by the subsequent empirical literature up to date, is neither robust to the aggregation period, nor to the observation frequency used to compute the volatility of trading volume. When using one-minute trading data over the most recent months, the relation is positive, in line with investors' aversion to liquidity uncertainty. However, portfolio strategies based on liquidity uncertainty do not appear to be profitable. This chapter is based on the work 'The liquidity uncertainty premium puzzle' in collaboration with Maria Flora and Roberto Renò (Flora, Gianstefani, and Reno', 2021).

1.1 Framework and scope

Liquidity is a wide and general concept that can be summarized as the ability to trade a certain quantity of an asset quickly, at the desired cost and without impacting too much on the security's price.
It is hard to find a single proxy that captures all the characterizing as-

pects of liquidity: throughout the years, researchers have proposed several measures, depending on data availability and the dimensions of liquidity to stress. We end up with a variety of studies which explore the relationship between the asset returns and the level and variability of liquidity measured with different approches.
Many empirical studies focus on how the level of liquidity impacts the stock returns throughout a cross-sectional analysis.
One of the first proposed proxy was the illiquidity cost, measured as the bid-ask spread: introduced by Amihud and Mendelson (1986), the model specifies an increasing and concave function linking the illiquidity cost and the expected returns.
M. Brennan and A. Subrahmanyam (1996) use a measure derived from the market microstructure literature, the average of the marginal cost of trading, to study the illiquidity-return relation: they confirm a positive effect of illiquidity on expected returns.
M. Brennan, T. Chordia, and A. Subrahmanyam (1998) proposed as measure for liquidity the dollar trading volume of the stock. They regress the risk-adjusted returns (using the three Fama-French factors) on the stock's dollar volume and other characteristics finding a significant negative impact of dollar volume on the stock returns.
An other well-known proxy for liquidity is the stock turnover rate, first introduced by Datar, Naik, and Radcliffe (1998) and defined as the number of shares traded as a fraction of the number of shares outstanding. Employing a Fama Macbeth (1973) setup, they evidence that the stock returns are a decreasing function of the turnover rates. This relation persists after controlling for the firm size, book to market ratio and the firm beta.
The most widly used measure for the illiquidity level is the one proposed by Amihud (2002a) who defines a measure called $ILLIQ = |R|/(P * VOL)$, where R is the daily return, P the closing price of the day and VOL the number ot shares traded during the day. $ILLIQ$ is the average ratio of the daily absolute return to the dollar trading volume on that day and it represents the daily stock price reaction to a dollar of trading volume. A monthly indicator of $ILLIQ$ is then computed and regressed in

a cross-sectional setup against the monthly returns. Coherently with the economic theory, $ILLIQ$ has a positive impact on the stock returns.
It is clear from the studies mentioned above that liquidity has an impact on the asset prices. Hence, it is reasonable that a variation in liquidity should affect the asset price: in fact, the liquidity is not a static quantity, it varies over time and its fluctuations are strongly impactful for the time series of excess returns.
This chapter deals with a puzzle put forward in Chordia, Subrahmanyam, and Anshuman (2001) (henceforth, CSA). Using a classical factorial approach rooted in the Arbitrage Pricing Theory to study the impact of liquidity on asset pricing, they reveal a sound negative relation between volume and excess returns, and a puzzling negative relation between volatility of volume and excess returns.
It is commonly accepted that investors require higher expected returns on assets with lower liquidity levels, establishing a sound negative relationship between the dollar volume and the excess returns. An illiquid asset entails the risk of not being traded in a short time at the fair value; on the contrary a security with high liquidity level does not face this kind of risk, requiring a lower compensation. Both the theoretical and empirical literature witness a negative liquidity premium.
Another fact about liquidity is that it varies over time: this introduce uncertainty to the investors who do not know the transaction cost they will incur in the future when they need to trade the asset. Fluctuations in liquidity might be convenient for an investor who time her trade, but they can also be a source of risk for a risk-averse agent who dislikes liquidity uncertainty. Whether the volatility of liquidity premium is positive or negative is a debated issue, both empirically and from an market microstructure point of view.
The theoretical literature is divided between researchers that justify a positive premium for the volatility uncertainty and academics that, conversely, are in favour of a negative premium for volatility of liquidity risk.
To the extent of our knowledge, all the papers dealing with the relation between liquidity volatility premium side with one of the two contrast-

ing theories: or they take the side of positive premium or the negative one; none of them embrace the hypothesis that these two theories are not conflicting and they can co-exist, as we sketch theoretically in the below and we deeply analyse in the empirical parts of the chapter.
Stocks have infinite maturities $t \in [0, T]$ and different levels of liquidity. The higher the liquidity level, the lower transaction costs are.
Assume a risk averse agent on the financial market, her investment choices are based on:

- The holding period: depending on the selected investment horizon $t \in [0, T]$, she can be a short-term investor or a long-term investor;

- The frequency of trading (intended as the average number of trades in a month, γ): she can trade less than once a month (not-high-frequency trader) or with regularity within a month (high-frequency investor);

- The degree of risk aversion as a function of the investment horizon (the longer the investment horizon, the lower the risk aversion) and the frequency of trading (the higher the frequency, the greater the aversion).

For simplicity, only two prototypes of investor operate on the market: a fraction $\alpha \in [0, 1]$ of them are long-term and not-high-frequency investors and the remaining percentage $(1 - \alpha)$ are agents with a short investment horizon and high-frequency activity on the market.
Short-term, high-frequency investors usually operate on the financial market for hedging purposes and have a strong preference for liquid stocks: the need to close open positions and place many trades (within a day) make them to prefer an asset with low transaction costs. They also prefer instruments with low liquidity uncertainty: the immediate liquidity needs and the impossibility to wait for periods of high liquidity force them to select stocks with low liquidity risk. According to this framework (embraced by Acharya and L.H. Pedersen (2005)), the security's return should include a positive premium to compensate the agents for the liquidity risk.

On the other hand, long-term and not-high-frequency traders usually reveal a lower degree of risk aversion and in equilibrium tend to invest in less liquid assets. Investors of this type can better withstand higher transaction costs because they incur in them infrequently and they can amortize them over longer periods. In addition, they can take advantage of the time-varying liquidity (they are keen on investing in securities with higher degree of liquidity risk) by timing their trades according to the state of liquidity. These agents do not require to be compensated with an extra premium for the liquidity uncertainty, but on the contrary, they support the evidence (described in Joao Pereira and Harold Zhang (2010)) that stocks with higher volatility in liquidity have lower returns.
The two theories, apparently in contrast, are complementary because they co-exist in explaining how different types of investor react to volatility of liquidity. For not-high-frequency investors with a long-term horizon, it makes sense to analyse their reaction at a monthly frequency (i.e., consider the liquidity level constant for a month): they do not have pressure in liquidating their positions, hence they can wait for the proper moment to place an order. Given their conduct on the market, it is reasonable to compute a measure of volatility in liquidity with data sampled every month; computing a measure of volatility with daily or one-minute data introduce noise that this type of agents do not even consider in their choice/ maximization of utility function. On the contrary, for high-frequency agents with a short-term horizon, it makes no sense to consider the level of liquidity constant during the month because within a single month they trade more than once, hence they care about the daily/intraday volatility of liquidity.
Our approach contributes to this debate in two directions. First, we propose new measures of volatility of liquidity. Volatility is a latent variable, and volatility of liquidity is no exception. This means that, to capture liquidity risk, we need proxies based on trading dollar volume at different time frequencies and on various window length.
The proxy used by CSA is the dispersion (measured by the coeffcient of variation, the ratio of the standard deviation of the variable to the average value of the variable, to soften the impact of outliers) of the monthly

trading dollar volume in the last three years. Of course, this measure may be problematic because of low precision (since only 36 data points are used), and because it smooths the dynamics of the volatility of volume (which may vary at a higher frequency than 3 years). Nowadays, transaction data run at a much higher frequency, and allow for superior resolution of dispersion measures. High-frequency data are rampaging in the volatility literature (see, e.g., Aït-Sahalia and Jacod (2015) and references therein), but much less employed to measure the volatility of trading volume, which is a natural measure of liquidity uncertainty. Using up to one-minute volume data to estimate monthly liquidity uncertainty (again, with the coefficient of variation, but now first computed at daily frequencies,and then aggregated to monthly measures), we uncover a significant and robust positive relation between liquidity uncertainty and excess returns.
Second, we investigate whether it is possible to exploit the presence of a premium for liquidity uncertainty in out-of-sample trading strategies based on sorting over liquidity uncertainty proxies, and double-sorting on volume and its uncertainty. Our results, which are net of time and stock varying transaction costs, clearly indicate the presence of a premium for holding illiquid stocks. However, the simple strategy based on sorting stocks with trading volume cannot be surpassed by double-sorting based on liquidity uncertainty. This indicates that the premium for liquidity risk, if existing, is economically extremely weak.

1.2 Literature review

The way in which liquidity uncertainty affects asset returns is still not completely understood in financial economics. Several papers deal with the topic, both proposing theoretical market micro-structure frameworks and empirical econometric analysis, but still many contrasting theories and results dwell in the literature.

1.2.1 Theoretical literature

There is not a clear stance on what should the relation between excess return and liquidity uncertainty be: as mentioned above, economic theory predicts both a negative and positive premium of liquidity risk.
In the most standard model in which the representative agent is averse to liquidity risk (Acharya and L.H. Pedersen (2005)), the relation should be positive, since the aggregate investor needs compensation for holding stocks with uncertain liquidity. The authors propose a liquidity-adjusted capital asset pricing model where risk-averse agents, in an overlapping generations economy, trade securities whose liquidity varies over time. The model shows that, since liquidity is persistent, a positive shock to illiquidity predicts high future illiquidity, which increases the required returns lowering contemporaneous prices. The required return positively depend on the expected security's illiquidity (the illiquidity cost are intended as the transaction costs) and the covariances between its return/liquidity and market return/liquidity. The channels through which liquidity uncertainty affect stock prices are:

- The covariance between asset's illiquidity and market illiquidity, which positively affects the asset's return (investors want to be compensated for holding a security that becomes illiquid when the market in general becomes illiquid);

- The covariance between the security's return and the market illiquidity, which negatively affects asset's returns (investors are willing to accept a lower return on an asset with a high return in times of market illiquidity);

- The covariance between asset's illiquidity and market returns, which negatively affects asset's returns (investors accept a lower expected return on a security that is liquid in a down market).

Hence, the model predicts a positive impact of liquidity risk on expected returns, but due to the co-movement of firm-specific liquidity with market return and market liquidity that affects expected returns, while CSA

use idiosyncratic volume variability in their study, so their empirical results might not be at odds with the model intuition.
On the other hand, J. Pereira and H. Zhang (2010) argue that liquidity variability may be an opportunity for a rational risk-averse, utility-maximizing investor who can can profit from the liquidity fluctuations. The authors intend the liquidity as the averse price impact of trading. The core element of the model is that an agent can time her trades and adapt her strategy to the liquidity state, to take advantage of periods of both high liquidity (by trading large amounts) and low liquidity (by trading small amounts). By embracing this approach, investors can benefit from liquidity uncertainty and so increase their demand for stocks with high liquidity risk. In equilibrium, stocks with higher liquidity uncertainty command a lower return premium, in agreement with the empirical findings of CSA.
The theoretical explanation that CSA provide to rationalize their empirical findings are based on the clientele effects. The clientele hypothesis states that different policies attract different types of investors, and changes to the policies will cause a shift in demand for the company's stock by investors, impacting not only its share price but also the variability of the traded volume because a more heterogeneous clientele leads to a greater variability in trading activity.
As a matter of fact, the theoretical literature proposes two apparently contrasting theories on the liquidity risk premium but it does not provide further guidance on which of the effects should prevail.

1.2.2 Empirical literature

Despite the theoretical ambiguity about the sign of the premium for the volatility of liquidity, the empirical literature either confirms the results of CSA or finds an insignificant relation.
A single study (Akbas, Armstrong, and Petkova, 2011), to best of our knowledge, contradicts CSA. However, this paper uses the Amihud (2002b) measure of illiquidity instead of dispersion measures such as volume or

turnover as done by all other studies we considered. In any way, the liquidity risk is measured with a firm-specific proxy for variability in liquidity. At the contrary, the paper by Pastor and Stambaugh (2003) focuses on systematic liquidity risk in returns and finds that stocks whose returns are more exposed to market-wide liquidity fluctuations command higher expected returns.

Easley, Hvidkjaer, and O'Hara (2002), for the period 1983-1998, confirm the result of CSA by adopting their same methodology but using turnover (to proxy liquidity level), computed as the number of shares traded divided by the number of shares outstanding, and coefficient of variation of monthly turnover in year $t-3$ to $t-1$. Also Fu (2009) (in the period 1963-2006) confirms the result of CSA, using as proxy for liquidity risk, the coefficient of variation of the previous 36 months' turnover (instead of trading volume). The negative relation is also confirmed by other empirical works:

- Joao Pereira and Harold Zhang (2010), in the period 1966-2005, proxy the liquidity risk (for contemporaneous return in t) using the coefficient of variation of both the dollar trading volume and the share turnover computed from month $t-37$ to $t-2$ (instead CSA compute the volatility measures over $t-36$ to $t-1$);

- Bali et al. (2014), in the period 1963-2010, find a statistically significant negative relation between the expected returns and the standar deviation of the monthly share turnover over the past 12 months, while the impact of the coefficient of variation of Amihud (2002a) measure over the past 12 months is not significant;

- Andreou et al. (2018), in the period 1996-2005, find a negative relation between the coefficient of variation of the monthly turnover over the past 36 months beginning in the second-to-last month;

- Huang (2018), in the period 2004-2015, using the coefficient of variation of dollar volume calculated over the past 12 months, finds a negative impact on stock returns;

- Chung and Chuwonganant (2018), in the period 1990-2012, find a negative relation between the expected returns and the standard deviation of monthly turnover over the last 12 months.

Feng, Giglio, and Xiu (2020) propose a method to systematically evaluate the contribution to asset pricing of new risk factors in explaining stock returns. The model distinguish useful factors from useless/redundant as they are introduced in the literature and they find the volatility of liquidity, defined as the coefficient of variation of dollar trading volume, to be one of the few factors that are significant in explaining the cross section of expected returns and robust to their double-selection test.
However, Hou, Xue, and L. Zhang (2020) fail to replicate a significant CSA effect for both volatility of turnover and volatility of dollar volume in the long sample 1967-2016, using a different statistical methodology: they compute the stock's coefficient of variation for the daily share turnover/ dollar volume over the prior 6 months.
Green, Hand, and X. F. Zhang (2017), in the period from 1980 to 2014, also find an insignificant relation of the monthly standard deviation of daily dollar trading volume and share turnover. Finally, Barinov (2015), in the period 1966-2010, finds the CSA effect using their same statistical methodology, but he shows that the effect disappears when controlling for idiosyncratic volatility. He concludes that the volatility of liquidity co-moves with idiosyncratic volatility and thus captures its negative relation with expected returns.
Figure (1) presents all the cited empirical works that re-propose a study similar to CSA and contextualize them on a timeline denoting the time sample over which the authors conduct the analysis.
In our 1998-2018 sample, we instead confirm the CSA effect using a generalization of their statistical model; however, to the best of our knowledge, we are the first to prove that the negative impact of volume uncertainty on expected returns is a spurious by-product of the way volatility of volume is measured. Using our statistical model and a more accurate measure of volume dispersion, the impact of volatility uncertainty on expected returns turns out to be positive and significant. We differ from the analysis of Barinov (2015) in several aspects. First, we resort to a more

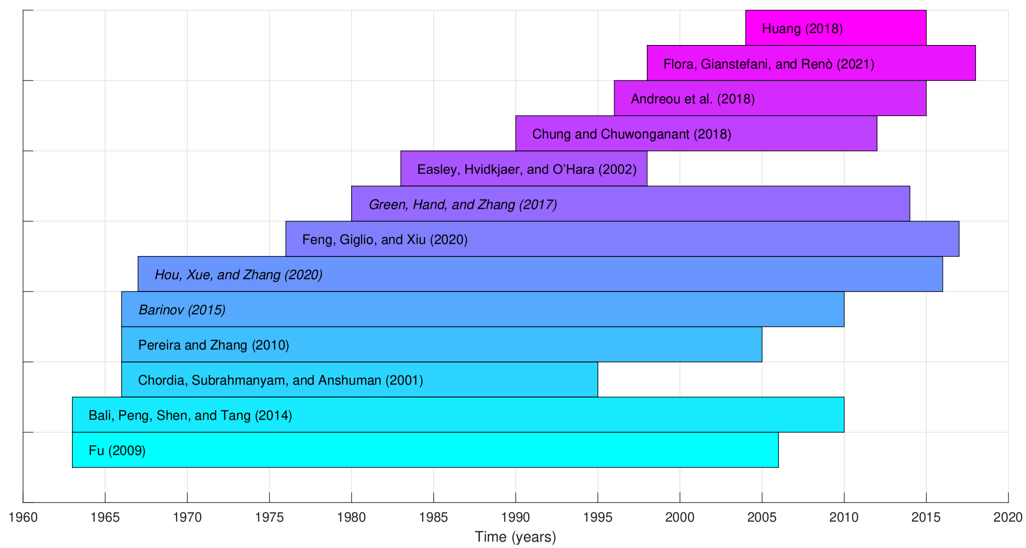


Figure 1: Empirical works that recreate a study similar to the one proposed by CSA and the time sample over which they run the analysis. References in italics, differently from the non-italic ones, fail to replicate the CSA effect.

precise measure of liquidity volatility, something he is calling for in the conclusions of his paper in his introduction, where he writes "*The zero relation, in contrast to the negative one, opens the gate to future studies of liquidity variability pricing, because the zero relation might arise because proxies for liquidity variability are imprecise*". We find this is indeed the case, and that the relation is positive when liquidity uncertainty is computed with higher frequency measures, on shorter window length and hence, closer to the contemporaneous return t. Second, we show that the reason for the puzzling result of CSA is mainly ignoring the time-varying nature of volume and its volatility, inspired by the results of Fu (2009) who shows that also idiosyncratic volatility bears a positive premium, not a negative one. Our results thus reconcile all of this literature by showing that, when accounting for the time-varying nature of state variables, both idiosyncratic volatility and variability of liquidity affect expected returns positively.

1.3 Methodology

To study the impact of liquidity level and of liquidity risk on asset prices, we use an approach that can be viewed as a generalization of CSA (which, in turn, used the methodology of M. J. Brennan, Tarun Chordia, and Avanidhar Subrahmanyam (1998)). After the description of the risk-adjustment procedure, we explain in details the core methodology of the analysis, the variables involved in it and the several models which can be estimated.

We denote by P_t^i the closing price of the i-th stock on day t, and we define

$$R_{t:t+h}^i = \frac{P_{t+h}^i - P_t^i}{P_t^i} \tag{1.1}$$

as the percentage return between days t and $t+h$.

1.3.1 Risk adjustment procedure

In the asset pricing literature it is common to risk adjust the returns by using factors which are known to partially describe the stock returns as a function of some general behaviours of the market. Specifically, Fama and French (1993) identify three factors that are able to explain the performance throughout the following factors: excess return of market portfolio ($R_{mkt} - R_f$), the outperformance of small vs. big companies (SMB) and the outperformance of high-book-to-market ratio vs. low book-to-market ratio companies (HML). We implement the analysis for every day of the dataset t. To risk adjust the returns computed on the interval from t to $t + h$ we need the Fama-French factors in the corresponding period of time. We download the daily version[1] of the variables and we aggregate them on the required window. At every time step t we slide the window one day ahead and we consequently re-aggregate the factors on the new window of width h. To adjust returns for the systematic part of factor risk, we use a standard Fama-MacBeth procedure, a method proposed by Fama and Macbeth (1973) to estimate parameters in a asset pricing context with panel data. The procedure involves an asset-by-asset time-series regression to estimate the factor loadings (β) for each stock (the sensibility of the stock to the risk factors) and a cross-sectional regression of expected returns on the estimated factor loadings to compute the risk premia (λ). It involves the following steps:

1. For each stock i-th we run a time-series regression to evaluate the exposure of the asset i to the risk factors $j = 1, \dots, N$; we estimate factor loadings $\beta_{j,i}$ from a linear factor model with N factors:

$$R^i_{t:t+h} = \alpha_i + \sum_{j=1}^{N} F_{j,t:t+h}\beta_{j,i} + \varepsilon^i_{t:t+h}\,, \tag{1.2}$$

where $R^i_{t:t+h}$ is the raw return over the period from t to $t + h$ with the exclusion of null returns and the returns that exceed 100%, $F_{j,t:t+h}$ is the value of the j-th factor over the period t to $t + h$, and

[1] We download the data from the website Kennet R. French, Data Library section.

$\varepsilon^i_{t:t+h}$ are independent forecast errors. To estimate the time-series regression model in (1.2), we employ only 60 observations, by sliding the window $\tau = t : t+h$ backwards until $t - 60h : t - 59h$ if available, otherwise we use all available observations with a minimum of 24. We label $\widehat{\beta}^{(\tau)}_{j,i}$ the estimated j-th factor loading obtained for each τ and for each stock i.

2. We then compute the risk premium $\lambda_{j,\tau}$ for each factor j by using the estimates of the factor loadings, $\widehat{\beta}^{(\tau)}_{j,i}$, and of the intercept, $\widehat{\alpha}^{(\tau)}_i$ in the cross-sectional regression

$$\widehat{\alpha}^{(\tau)}_i - r^f_\tau = \sum_{j=1}^{3} \lambda_{j,\tau} \widehat{\beta}^{(\tau)}_{j,i} + \epsilon_i \,, \tag{1.3}$$

where r^f_τ is the risk-free rate over the period τ computed by aggregation[2] of daily risk free rates over the period $t : t+h$. We exclude the outlier values for the $\widehat{\beta}^{(\tau)}_{j,i}$ by looking at their pooled distributions. For $\widehat{\beta}^{(\tau)}_{1,i}$ we consider outliers the values outside the interval $[-2, 4]$, for $\widehat{\beta}^{(\tau)}_{2,i}$ and $\widehat{\beta}^{(\tau)}_{3,i}$ we remove the values outside the interval $[-5, 6]$.

3. We use the estimated factor risk premia $\widehat{\lambda}_{j,\tau}$ to compute the risk adjusted returns as

$$R^*_{i,\tau} = R_{i,\tau} - \sum_{j=1}^{3} \tilde{F}_{j,\tau} \widehat{\beta}^{(\tau)}_{j,i} \,, \tag{1.4}$$

where $\tilde{F}_{j,\tau} \equiv \widehat{\lambda}_{j,\tau} + F_{j,\tau}$.

[2]By definition the risk free rate for the US market is the one-month T-bill rate; Fama-French propose on their website a daily version of the risk-free rate that can be aggregated over the required period, from t to $t+h$ as

$$r^f_\tau = \prod_{\tau=t}^{t+h} \left(\frac{1 + r^{f,daily}_\tau}{100} - 1 \right) * 100$$

1.3.2 The core methodology

Once obtained the risk-adjusted returns with the Fama-MacBeth (1973) procedure, we focus on the core of our analysis. The methodology consists in a cross-sectional regression of the risk-adjusted excess returns on a number K of security characteristics, $Z_{k,t}$, for each time step t:

$$\overline{R}^{i}_{t:t+h} = a_t + \sum_{k=1}^{K} b_{k,t} Z^{i}_{k,t} + \varepsilon^{i}_{t:t+h} \, , \tag{1.5}$$

where $\overline{R}^{i}_{t:t+h}$ denotes the *risk adjusted excess* return[3] from t to $t+h$, that is equal to $R^{*i}_{t:t+h}$ minus the 1-month T-bill rate properly standardized[4], and $\varepsilon_{t:t+h}$ is zero-mean random noise which is independent across stocks. The model (1.5) yields a series of $\hat{b}_{k,t}$ for each time t; by time-series averaging the $\hat{b}_{k,t}$, we obtain the final coefficients $\hat{b}_k$ for each security characteristic k. The significance of the estimated parameters is computed with the Newey-West procedure. In fact, employing overlapping windows $t : t+h$ allows to squeeze at most the informational content from data, with the drawback of the autocorrelation: when we slide the window of length h one day ahead, we shift the time interval from $t : t+h$ to $t+1 : t+h+1$ implying an overlap of $h-1$ days between the two windows. This issue gradually mitigates, for time step t, as we move forward and the common elements between the windows are progressively reduced, but to address possible autocorrelation issues, we use the Newey-West procedure with h lags to assess the significance of the time-series average of the $\hat{b}_{k,t}$ coefficients. The procedure, proposed by Newey and West (1987), requires the computation of a weighting scheme to assign a lower weight to elements that are farther apart. Specifically, the correlation matrix used to compute the standard errors, has a weight computed as $w_l = 1 - \frac{l}{1+h}$ for $l = 0, \dots, h$ and $1 = w_0 > w_1 > \dots > 0$ where h is the length of the window. Disturbances that are farther apart

[3]In what follows, as a robustness check, we will also estimate (1.5) using the raw excess returns $R^{i}_{t:t+h} - r^{f}_{t:t+h}$ as a dependent variable, instead of the risk-adjusted excess returns obtained with the Fama-MacBeth procedure described above.

[4]As explained in the previous section, to obtain the variables on the time interval $t : t+h$, we use the daily version of the variable and aggregate it on the corresponding window.

from each other are given lower weight[5], while those with equal subscripts are given a weight of 1. The covariance matrix is given by

$$cov(b) = (Z^T Z)^{-1} Z^T S Z (Z^T Z)^{-1}$$

where

$$Z^T S Z = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 z_t z_t' + \frac{1}{T} \sum_{l=1}^{h} \sum_{t=l+1}^{T} w_l \varepsilon_t \varepsilon_{t-l} (z_t z_{t-l}' + z_{t-l} z_t')$$

Our goal is to assess the incremental explanatory power for (risk-adjusted and raw) excess returns of liquidity-related security characteristics. In this respect, a crucial part of the identification strategy lies in the choice of the proxy for liquidity risk. For this, we include several alternatives, that mainly differ in the level of granularity.

1.3.3 Variables

The security characteristics, $Z_{k,t}$, that we include as explanatory variables in our regression model (1.5) are listed below (we suppress the superscript i to avoid clutter):

- To measure the liquidity level, we use the natural logarithm of the dollar volume, DV_t, computed as the sum of one-minute trading dollar volumes (number of traded shares times price) over the past h days, that is from day $t-h+1$ to day t. When $h = 21$ trading days, that is a trading month, this corresponds to the CSA monthly DVOL measure.[6]

---

[5]An example of weighting matrix scheme with $h = 3$:

$$\begin{bmatrix} 1 & w_1 & w_2 & w_3 \\ w_1 & 1 & w_1 & w_2 \\ w_2 & w_1 & 1 & w_1 \\ w_3 & w_2 & w_1 & 1 \end{bmatrix}$$

[6]A minor difference is that CSA use calendar months, while we use a fixed difference horizon between trading days.

As a robustness check (section 1.6) we also propose two alternative measures of liquidity level: the measure introduced by Amihud (2002a), computed as the daily ratio of absolute stock return to its dollar volume averaged over period from $t-h+1$ to t:

$$AMIHUD_t = \frac{1}{h}\sum_{d=t-h+1}^{t}\frac{|R_d|}{DV_d}$$

The second alternative liquidity measure, $ZEROS_t$, is the average of the percentage of null one-minute returns in a day computed over the period from $t-h+1$ to t. For further details, see section 1.4.

- To measure liquidity uncertainty we use several alternatives:
  - A proxy for liquidity risk that employs *monthly* data:
    * CVVOLM$_t^{(m)}$, that is the natural logarithm of the coefficient of variation of the *monthly* dollar volumes over the past m months. When $m = 36$, this is the original measure used in CSA.
  - Proxies for liquidity risk that employ *daily* data:
    * CVVOL$_t^{(m)}$, the natural logarithm of the coefficient of variation of the *daily* dollar volumes sampled over the past $h \cdot m$ days (when $h = 21$, m is the number of months), that is between day $t-hm+1$ and day t.
    * MADCVVOL$_t^{(m)}$, that is a standardized measure of dispersion of the *daily* dollar volumes sampled over the past $h \cdot m$ days, that is between day $t-hm+1$ and day t. It is defined as the natural logarithm of the ratio of the median absolute deviation of daily volumes to their median. Thus, it is a measure similar to CVVOL$_t^{(m)}$, but with median and median absolute deviation replacing mean and standard deviation;

- * $\text{AMIHUDVOL}_t^{(m)}$, that is the coefficient of variation of the daily Amihud ratio over the past $h \cdot m$, that is between day $t - hm + 1$ and day t. We will use it as a robustness check in Section 1.6.
  * $\text{ZEROSVOL}_t^{(m)}$, that is the coefficient of variation of the daily Zeros measure over the past $h \cdot m$, that is between day $t - hm + 1$ and day t. We will use it as a robustness check in Section 1.6.
- Proxies for liquidity risk that employ *one-minute* data:
  * $\text{HFVOL}_t^{(d)}$, the average of the natural logarithm of the *daily* coefficient of variations of *one-minute* dollar volumes, computed over d past days, that is from day $t - d + 1$ to day t;
  * $\text{HFVOL}_t^{(d)}(\text{AGGR})$, defined as the natural logarithm of the coefficient of variation of *one-minute* dollar volumes, over the past d days. This measure, as opposed to $\text{HFVOL}_t^{(d)}$, that only captures the intra-day variation of the dollar-volume, also captures variation across days.

- We finally include past cumulative returns (also called momentum variables) as explanatory covariates, and precisely $\text{RET}_{2-3} = R_{t-3h:t-h}$, $\text{RET}_{4-6} = R_{t-6h:t-3h}$, $\text{RET}_{7-12} = R_{t-12h:t-6h}$. These are the same variables used by CSA to proxy for momentum effects (Jegadeesh and Titman, 1993). As in CSA, we exclude the return during the immediate prior window (which correspond to the prior month when $h = 21$) to prevent spurious autocorrelation effects due to thin trading.

1.3.4 Estimated Models

The estimation methodology is the following. For each date t for which we have at least 24 months of data available in the past and h days of data available ahead for at least 300 stocks satisfying mild quality cuts[7],

[7] We exclude returns larger than 100%, and volumes that are too low.

we cross-sectionally regress returns from t to $t+h$ on the security characteristics listed above, measured at t. As a result, we have a series of estimates the coefficients $\widehat{b}_{k,t}$ of model (1.5). As mentioned, we then address the significance of the estimates by computing the t-statistics of their time-series average using the Newey-West correction with h lags. To test the impact of the choice of the liquidity risk proxy on the sign and significance of the liquidity risk premium, we estimate (1.5) multiple times, each time including a different proxy for liquidity risk, and keeping all other explanatory variables equal. That is, we estimate

$$\begin{aligned}\overline{R}^{i}_{t:t+h} =& a^{i}_{t} + b^{i,\mathrm{DV}}_{t}\mathrm{DV}^{i}_{t} + b^{i,\mathrm{LVOL}}_{t}\mathrm{LVOL}^{i}_{t} \\ &+ b^{i,\mathrm{RET}_{2-3}}_{t}\mathrm{RET}^{i}_{2-3,t} + b^{i,\mathrm{RET}_{4-6}}_{t}\mathrm{RET}^{i}_{4-6,t} \\ &+ b^{i,\mathrm{RET}_{7-12}}_{t}\mathrm{RET}^{i}_{7-12,t} + \varepsilon_{i,t:t+h}\end{aligned} \tag{1.6}$$

where LVOL^{i}_{t} can be either $\mathrm{CVVOLM}^{(m)}_{t}$ (when $m = 36$ it corresponds to the CSA measure) or one of our increasingly higher frequency measures, that is $\mathrm{CVVOL}^{(m)}_{t}$ and $\mathrm{MADCVVOL}^{(m)}_{t}$ for $m = 1, \ldots, 36$ (obtained from daily data), or $\mathrm{HFVOL}^{(d)}_{t}$ and $\mathrm{HFVOL}^{(d)}_{t}$(AGGR) for $d = 1, \ldots, 36{\cdot}21$ (obtained from 1-minute data). In Table (1) we summarize the main features of the models. We only report the models whose proxies for liquidity level and risk are based on dollar volume. The dependent variable is represented by the risk-adjusted excess returns over the period from t to $t+h$, the liquidity level is always proxied by the dollar volume traded in the previous month and also the liquidity risk is based on different measures of dispersion of the dollar volume. Hence, the models differ for the measure of liquidity risk, namely for the granularity of data and the length of the window h. As stated in the previous section, the sampling frequency of the data can be monthly, daily or one-minute; in addition, the unit window length h can be of 21,10,5 days defining respectively monthly, biweekly, or weekly unit windows (see Table 2). Finally each model presented in Table (1) presents different version of itself, depending on the aggregation period (m or d) over which the the volatility measure is calculated. In the section dedicated to the robustness checks (1.6) we will present the results with measures of liquidity level and risk obtained with the alternative proxy: the Amihud ratio and Zeros.

Model ID	Model Name	Independent Variables: Liquidity Level	Independent Variables: Liquidity Risk	Independent Variables: Control Variables	Window Length (h)	Aggregation Periods (m or d)
(1)	CVVOLM_MONTHLY	DV_t	$CVVOLM_t^{(m)}$	Momentum variables	21	$h \cdot m$, m=12,18,24,30,36
(2)	CVVOLM_MONTHLY _BETAVOL	DV_t	$CVVOLM_t^{(m)}$	Momentum variables, $BETAVOL_t$	21	$h \cdot m$, m=12,18,24,30,36
(3)	CVVOLD_MONTHLY	DV_t	$CVVOL_t^{(m)}$	Momentum variables	21	$h \cdot m$, m=1,...,36
(4)	MADCVVOLD_MONTHLY	DV_t	$MADCVVOL_t^{(m)}$	Momentum variables	21	$h \cdot m$, m=1,...,36
(5)	CVVOLD_BIWEEKLY	DV_t	$CVVOL_t^{(m)}$	Momentum variables	10	$h \cdot m$, m=12:12:36$\cdot h$
(6)	MADCVVOLD_BIWEEKLY	DV_t	$MADCVVOL_t^{(m)}$	Momentum variables	10	$h \cdot m$, m=12:12:36$\cdot h$
(7)	CVVOLD_WEEKLY	DV_t	$CVVOL_t^{(m)}$	Momentum variables	5	$h \cdot m$, m=12:12:36$\cdot h$
(8)	MADCVVOLD_WEEKLY	DV_t	$MADCVVOL_t^{(m)}$	Momentum variables	5	$h \cdot m$, m=12:12:36$\cdot h$
(9)	HFVOL_MONTHLY	DV_t	$HFVOL_t^{(d)}$	Momentum variables	21	$d = 1, \ldots, 20$ d=hm, m=1,...,36
(10)	HFVOLAGGR_MONTHLY	DV_t	$HFVOL_t^{(d)}$(AGGR)	Momentum variables	21	$d = 1, \ldots, 20$ d=hm, m=1,...,36
(11)	HFVOL_BIWEEKLY	DV_t	$HFVOL_t^{(d)}$	Momentum variables	10	$d = 1, \ldots, 9$ d=hm, m=1,...,36
(12)	HFVOLAGGR_BIWEEKLY	DV_t	$HFVOL_t^{(d)}$(AGGR)	Momentum variables	10	$d = 1, \ldots, 9$ d=hm, m=1,...,36
(13)	HFVOL_WEEKLY	DV_t	$HFVOL_t^{(d)}$	Momentum variables	5	$d = 1, \ldots, 4$ d=hm, m=1,...,36
(14)	HFVOLAGGR_WEEKLY	DV_t	$HFVOL_t^{(d)}$(AGGR)	Momentum variables	5	$d = 1, \ldots, 4$ d=hm, m=1,...,36

Table 1: The Table summarizes the models we estimate, whose results are reported in section 1.5. The name of the model specifies the proxy we use for the liquidity risk and the length of the window h days. The dependent variable is always represented by the monthly risk-adjusted excess return from t to $t+h$. The liquidity level is always proxied by the dollar volume computed in the previous month with the respect to the returns. For the liquidity uncertainty we propose several alternatives, that differ in the granularity of data (monthly - models (1) and (2), daily - models from (3) to (8), one-minute - models from (9) to (14)) and the period over which we compute the dispersion measure (which depends on the window length h and the aggregation period m months or d days).

Sampling frequency		Window length (h)		
		21 (monthly)	10 (biweekly)	5 (weekly)
Monthly	$CVVOLM_t^{(m)}$	(1),(2)	-	-
Daily	$CVVOL_t^{(m)}$	(3)	(5)	(7)
	$MADCVVOL_t^{(m)}$	(4)	(6)	(8)
One-Minute	$HFVOL_t^{(d)}$	(9)	(11)	(13)
	$HFVOL(AGGR)_t^{(d)}$	(10)	(12)	(14)

Table 2: The Table presents a classification of models (1)-(13) presented in Table 1, double sorted as a function of the sampling frequency of the data (which can be monthly, daily or one-minute) and the unit window length, which can be of 21 days (monthly case), 10 days (biweekly) or 5 days (weekly).

1.4 Data

The dataset consists of 4809 stocks traded (not over the full period) on NYSE and AMEX in the period from January 1998 to June 2018. We have prices and number of traded shares for every minute of the trading day, from 9:30am to 4pm (a trading day usually lasts 390 minutes).
We squeeze the informativeness of the dataset by aggregating the data at a monthly/daily level or not aggregating them at all (i.e. exploiting the one-minute granularity of data). The section presents some descriptive statistics of the data sampled at a daily and one-minute level and lastly a comparison between monthly, daily and one-minute sampled data.
In the first part of the study, we replicate the analysis of T. Chordia, A. Subrahmanyam, and V. Anshuman (2001) and we remodel our dataset to obtain monthly data, as in the mentor paper.
Monthly return $R^i_{t:t+h}$, with $h = 21$ is computed as the percentage change in the asset price during the period $t : t+h$: $R_{t:t+h} = \frac{P^i_{t+h} - P^i_t}{P^i_t *}$ where P_{t+h} is the closing price at 15:59[8] of day $t+h$ and P_t is the closing price on day t. Monthly volume $Vol^i_{t:t+h}$ is the sum of all the volumes traded during

[8]The American Stock Exchange opens at 9:30am and closes at 16:00pm. Sometimes the last available closing price is not at 15:59 but earlier. On some days the stock exchange closes at 13:00 , hence the last available price is at 12:59; but it might also happen that the time series has some missing values: in that case we use the latest registered price of the day.

the period $t : t + h$ for the stock i.

1.4.1 Daily aggregation of data

When deal with daily data, we dispose of a dataset of daily close price and traded volumes for all the trading days in the mentioned period, for a total of 5157 days. We eliminate the days for which we do not have data due to the closure of the market.
We denote by P_t^i the daily close price for the i-th stock on day t; in case of missing price we consider the close price of the previous day. The daily volume Vol_t^i for the i-th stock is compute as the sum of the one-minute volume in each day t ($Vol_{k,t}^i$):

$$Vol_t^i = \sum_{k=1}^{390} Vol_{k,t}^i$$

The daily dollar volume DV_t^i for the i-th stock is the sum on the one-minute volume multiplied by the corresponding one-minute price on day t:

$$DV_t^i = \sum_{k=1}^{390} Vol_{k,t}^i \cdot P_{k,t}^i$$

The daily dollar volume presents a right skewed distribution (Figure (2)) with an average traded dollar volume of 34.67 million \$ and a median of 4.06 million \$.
For each day t we also compute the realized variance of the 5-minute returns as the sum of squared 5-minute log-returns ($R5m_{k,t}^i = log(P_{k,t}^i - P_{k-5,t}^i)$) on day t:

$$RV5m_t^i = \sum_{k=1}^{78} (R5m_{k,t}^i)^2$$

In a trading day there are 78 5-minute slots.

For each day t we also compute the percentage of null one-minute returns that each stock i presents. For the i-th stock on day t, the measure is computed as the ratio between the number of null one-minute returns for

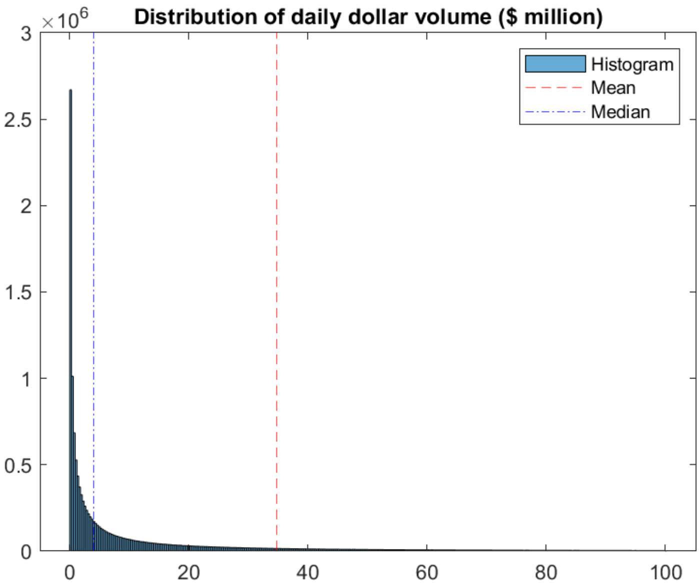


Figure 2: Distribution of the daily dollar volume.

the day and the total number of minutes the i-th stock is traded on that day. Specifically, the numerator is defined as the sum of a one-minute indicator function that assumes value 1 when the k one-minute log-return ($R1m^i_{k,t} = log(P^i_{k,t} - P^i_{k-1,t})$) is null, and 0 otherwise:

$$I_{R1m^i_{k,t}} = \begin{cases} 1 & \text{if } R1m^i_{k,t} \text{ is } 0 \\ 0 & \text{otherwise} \end{cases}$$

Hence, the number of zero one-minute returns in a trading day are computed as:

$$NumberZeros^i_t = \sum_{k=1}^{390} I_{R1m^i_{k,t}}$$

Differently from the numerator the involves the price of the asset, the denominator takes into account the one-minute traded (dollar) volume ($DV^i_{k,t} = Vol^i_{k,t} \cdot P^i_{k,t}$); again, we define an indicator function that assumes value 1 when the one-minute dollar volume is positive, and 0 otherwise:

$$I_{DV^i_{k,t}>0} = \begin{cases} 1 & \text{if } DV^i_{k,t} > 0 \\ 0 & \text{otherwise} \end{cases}$$

The number of trading minutes per day are given by:

$$TradingMinutes^i_t = \sum_{k=1}^{390} I_{DV^i_{k,t}>0}$$

The daily measure $Zeros^i_t$ is finally computed as the ratio of the previously defined variables:

$$Zeros^i_t = \frac{NumberZeros^i_t}{TradingMinutes^i_t}$$

We end up with a matrix of size (number of days $\times$ number of stock), (5157 $\times$ 4809) where each element is the percentage of trading minutes per day that present a zero one-minute return for stock i-th. Figure (3) presents the pooled distribution of the variable $Zeros^i_t$. The majority of observations presents a null rate of zeros: when the stock is traded, the impact on the price is effective and generates a shock in the returns. On

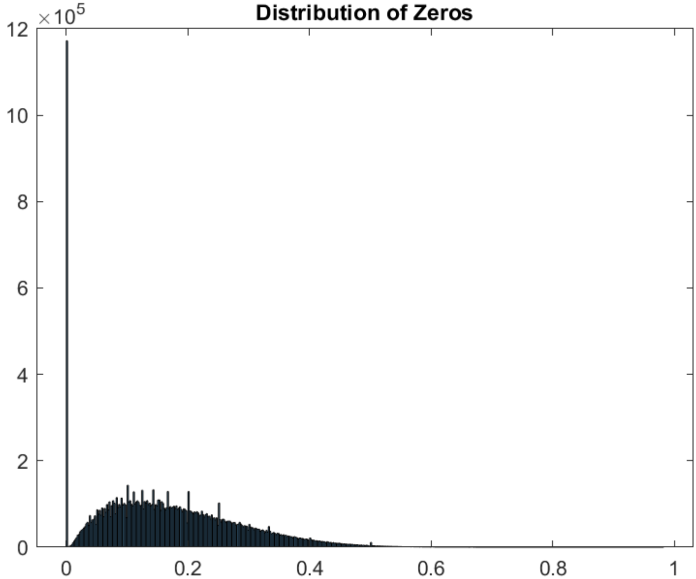


Figure 3: Pooled distribution of the variable $Zeros_t^i$. For each day t and each asset i, we compute the percentage of one-minute null returns when the stock is traded. The histogram shows the values of the variables across stocks and time. The distribution has a mean of 17.20%, a median of 15.79% and a standard deviation of 11.47%.

average, across stocks, we have a 17.20% of trading minutes per day for which we have agents on the market that actively trade the stock i-th but they do not generate any kind of impact on the price level.

By exploiting the daily frequency of data, it is possible to compute the standard deviation of daily dollar volume over several aggregation window. Specifically for each month[9], we compute the measure of dispersion over the past $h \cdot m$ days, where $h = 21$ is the number of business days in a month (here we consider a monthly unit window of length 21 days) and m is the number of months $m = 1, \ldots, 36$. We and up

[9]The month is intended as the fixed difference horizon h between trading days, it is not the calendar month.

with 36 matrices of size 245 $\times$ 4809, which are respectively the number of months and the number of stocks. Each element of the matrix is the standard deviation of the daily dollar volume between day $t - hm + 1$ and t. Clearly, each matrix m-th has the first m rows empty because we need a sufficiently wide initial window of m months to compute the volatility of data. In Table (3) left panel, we report the mean, the median and the standard deviation of the standard deviation of daily dollar volume as a function of the aggregation window length (m months). As we can appreciate, the three statistics (mean, median and standard deviation) are increasing as the aggregation window widens, an indication that the larger the aggregation window, the higher the average value and the dispersion of the volatility measure. Hence, the wider the aggregation window, the more the distribution of the standard deviation shift to the right and increase its dispersion around the mean value. A similar remark applies to the coefficient of variation of daily dollar volume (see Table (3) right panel and Figure (4)). In Figure (5) we display the distributions of the coefficient of variation computed over the previous past month ($m = 1$, $CVVOL^{(1)}$) and over the past 36 months ($m = 36$, $CVVOL^{(36)}$). When the aggregation window widens, the distributions tend to shift to the right and to become more disperse. For sake of clarity, in the picture we only represent the two extremes, but all the other distributions are placed in between.

We finally compute for each aggregation window of length $m = 1, \ldots, 36$ and for each month in the dataset the correlations between the stock characteristics. The variables involved in the correlation analysis are the following. $RET_{t:t+h}$, $DVOL_t$ (the dollar traded volume in the period from $t-h$ to t), $STDVOL_t^{(m)}$ and $CVVOL_t^{(m)}$, respectively, the standard deviation and the coefficient of variation of daily dollar volume over the past $h \cdot m$ days, for $m = 1, \ldots, 36$). Since the measures of dispersion are computed over 36 different aggregation windows, we end up with collection of 36 manifestations of volatility for $STDVOL_t^{(m)}$ and $CVVOL_t^{(m)}$. We compute the correlation between $RET_{t:t+h}$ and $STDVOL_t^{(m)}$, $RET_{t:t+h}$ and $CVVOL_t^{(m)}$, $DVOL_t$ and $STDVOL_t^{(m)}$, $DVOL_t$ and $CVVOL_t^{(m)}$, $STDVOL_t^{(m)}$ and $CVVOL_t^{(m)}$ for each month t and aggregation window $m = 1, \ldots, 36$. The correlations between the mentioned variables as a function of the aggregation window on which we compute the measure of dispersion are plotted in Figure (6). The correlation between the stock returns and the volatility measures are extremely mild, slightly negative when the standard deviation is adopted, moderately positive in the case of coefficient of variation. In both cases, the aggregation window (m) does not have an impact on the relation. On the other side, the correlation between $DVOL_t$ and $STDVOL_t^{(m)}$ is strongly positive, reaches its maximum (around 0.95) when the aggregation window is very tight and mildly declines as the aggregation window increases. Also the correlation between $DVOL_t$ and $CVVOL_t^{(m)}$ is quite consistent and negative, it declines at a decreasing rate as long as the aggregation window enlarges. Finally, the correlations between $STDVOL_t^{(m)}$ and $CVVOL_t^{(m)}$ are negative and do not depend on m. Limiting the analysis to $m = 36$, that corresponds to the CSA measures, our results are perfectly in line with the ones of the original paper. The correlation between $DVOL_t$ and $STDVOL_t^{(36)}$ is 0.901 in our sample and 0.908 in CSA; the correlation between $DVOL_t$ and $CVVOL_t^{(36)}$ is -0.623 in our case and -0.446 in their work; lastly the correlation between $STDVOL_t^{(36)}$ and $CVVOL_t^{(36)}$ is -0.407 for our dataset and -0.243 in their data.

1.4.2 One-minute data

Dealing with one-minute data means to exploit the highest granularity of the dataset and squeeze its informativeness at the maximum level.
For each day t and stock i-th, we have 390 one-minute observations of prices ($P_{k,t}^i$) and number of traded share ($Vol_{k,t}^i$), one for each minute k of the trading day ($k = 1, \ldots, 390$).
Handling one-minute data over a period of almost 20 years and 4809 stocks is not an easy task computationally speaking. Even if the all the stocks are not traded over the full period, each day has 390 one-minute observations that multiplied by the total number of trading days (5157) gives a total of 2011230 one-minute observations per stock. Hence, when

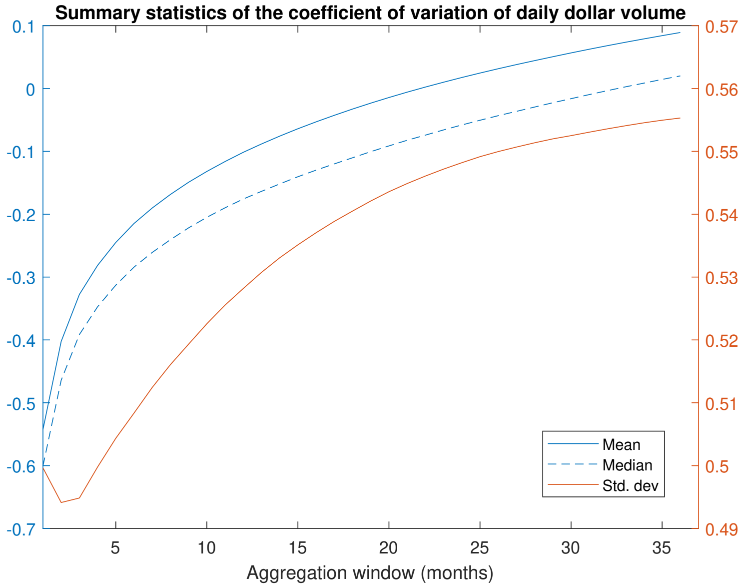


Figure 4: Summary statistics (mean, median and standard deviation) of the coefficient of variation of daily dollar volume computed over different aggregation windows of length $m = 1, \ldots, 36$ months. The statistics are computed as the mean, median, standard deviation respectively of the distribution of the coefficient of variation of daily dollar volume computed over the past m months.

possible, we compute variables that involve one-minute data on the basis of daily aggregated data.

We denote by $DV^i_{k,t} = Vol^i_{k,t} \cdot P^i_{k,t}$ the one-minute dollar volume traded during the minute k on day t for the stock i-th. For each stock i, we compute the average (and the median) one-minute dollar volume over the entire trading period as a time-series average (median) of the one-minute observations. Figure (7) presents the distribution across all the stocks. As in the case with daily dollar volume, the distributions are highly populated in the very left part (especially the distribution of the median), and the densities rapidly declines exhibiting a long right tail. The more pronounced positive skewness that characterizes the distribu-

	Standard Deviation			Coefficient of Variation		
Aggregation window (months)	Mean ($mil)	Median ($mil)	Std. dev. ($mil)	Mean	Median	Std. dev.
1	15.38	2.44	55.77	-0.5426	-0.6005	0.4996
2	17.11	2.9	59.27	-0.4024	-0.4642	0.4941
3	18.04	3.19	61.12	-0.3277	-0.3915	0.4948
4	18.59	3.37	62.31	-0.2814	-0.3473	0.4998
5	19.04	3.52	63.31	-0.2449	-0.3129	0.5044
6	19.43	3.65	64.17	-0.2148	-0.2842	0.5084
7	19.74	3.77	64.85	-0.1901	-0.2608	0.5124
8	20.02	3.87	65.46	-0.1685	-0.2405	0.5161
9	20.29	3.97	66.02	-0.1491	-0.2217	0.5193
10	20.52	4.05	66.54	-0.1319	-0.2049	0.5226
11	20.74	4.13	67.06	-0.1162	-0.1898	0.5255
12	20.95	4.21	67.57	-0.1015	-0.1758	0.5282
13	21.14	4.28	68.05	-0.0882	-0.1636	0.5307
14	21.32	4.35	68.51	-0.0758	-0.1518	0.5331
15	21.5	4.41	68.96	-0.064	-0.1404	0.5351
16	21.66	4.48	69.39	-0.053	-0.1299	0.5371
17	21.82	4.53	69.8	-0.0426	-0.1199	0.5389
18	21.97	4.59	70.2	-0.0327	-0.1102	0.5405
19	22.11	4.64	70.6	-0.0233	-0.1005	0.5421
20	22.25	4.7	70.99	-0.0143	-0.0913	0.5436
21	22.39	4.75	71.37	-0.0058	-0.0823	0.5449
22	22.51	4.8	71.74	0.0023	-0.0742	0.5461
23	22.64	4.84	72.1	0.01	-0.0659	0.5472
24	22.75	4.88	72.42	0.0175	-0.058	0.5482
25	22.86	4.93	72.71	0.0246	-0.0505	0.5492
26	22.96	4.97	72.98	0.0315	-0.0432	0.55
27	23.05	5.02	73.22	0.038	-0.0364	0.5507
28	23.13	5.06	73.44	0.0444	-0.0295	0.5514
29	23.2	5.1	73.65	0.0506	-0.0227	0.552
30	23.28	5.14	73.85	0.0566	-0.0162	0.5525
31	23.36	5.18	74.05	0.0625	-0.0097	0.5531
32	23.42	5.22	74.2	0.0682	-0.0034	0.5536
33	23.49	5.25	74.36	0.0736	0.0027	0.5541
34	23.55	5.29	74.51	0.0789	0.0085	0.5545
35	23.6	5.32	74.64	0.0841	0.0145	0.555
36	23.66	5.36	74.77	0.0891	0.0201	0.5553

Table 3: The summary statistics (mean, median and standard deviation) represent the time-series averages of the cross-sectional statistics for the Standard Deviation (left panel) and the Coefficient of Variation (right panel) of the daily dollar volume computed over the past hm days where $h = 21$ is the number of trading days in a month and $m = 1, \ldots, 36$ is the number of months.

tion of the mean confirms that the high values that form the right tail are mainly due to extreme values in the distribution of dollar volume of some stocks.

We report in Figure (8), by way of example, the distribution of the one-minute dollar traded volume of the stock Apple (AAPL). The choice is not arbitrary, we select a very liquid stock with very few missing observation and a very extended time sample (it is traded over the full period

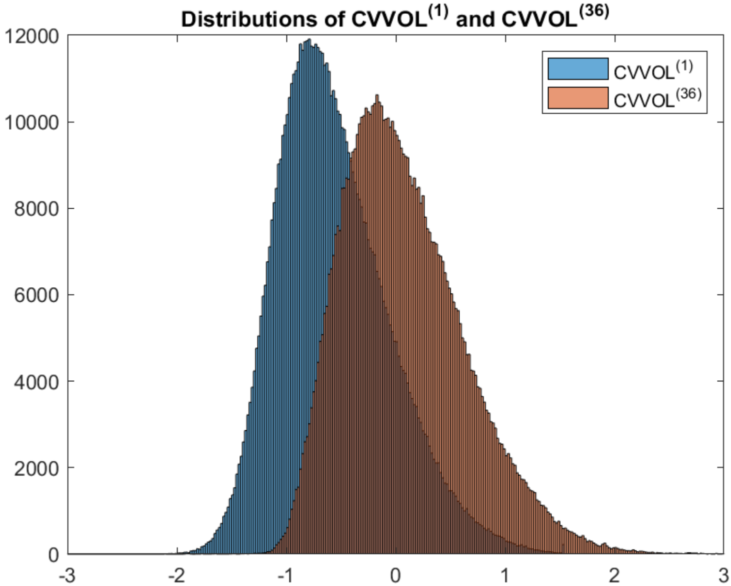


Figure 5: Distributions of the coefficient of variation computed over the previous past month ($m = 1$, $CVVOL^{(1)}$) and over the past 36 months ($m = 36$, $CVVOL^{(36)}$).

covered by the dataset), to show the variability of the data. In fact, considering only the mean and the median of the one-minute dollar volume for each stock might be reductive because many information are lost due to the conciseness of the indicators that do not take into account the data dispersion. In this particular case, the mean and the median of the dollar volume of AAPL are not included in Figure (7) because of the truncation in the horizontal axis; trimming x-axis of Figure (7) implies that the most liquid stocks are not included in the representation of the data.

As for the one-minute version of the dispersion measures, we compute both the standard deviation and the coefficient of variation of the one-minute dollar volume over various aggregation window of length $d = 1, \ldots, 36 \cdot 21$ days.

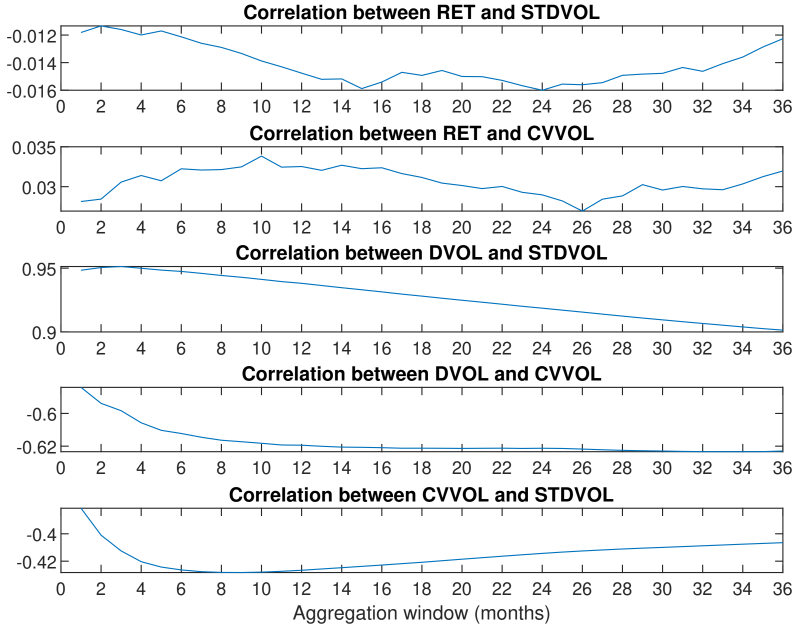


Figure 6: The correlations between $RET_{t:t+h}$ and $STDVOL_t^{(m)}$, $RET_{t:t+h}$ and $CVVOL_t^{(m)}$, $DVOL_t$ and $STDVOL_t^{(m)}$, $DVOL_t$ and $CVVOL_t^{(m)}$, $STDVOL_t^{(m)}$ and $CVVOL_t^{(m)}$ as a function of the aggregation window on which we compute the measure of dispersion.

A measure of dispersion exploiting one-minute data can be computed within a single day (if enough observations are available). We propose two different alternatives to quantify the dispersion measures, one that consider only the intraday variation and a second one that takes into account both the intraday and the interday variability of data:

- The first technique that only incorporates the intraday variability, requires to compute the average of the *daily* standard deviation or the natural logarithm of coefficient of variation of one-minute dol-

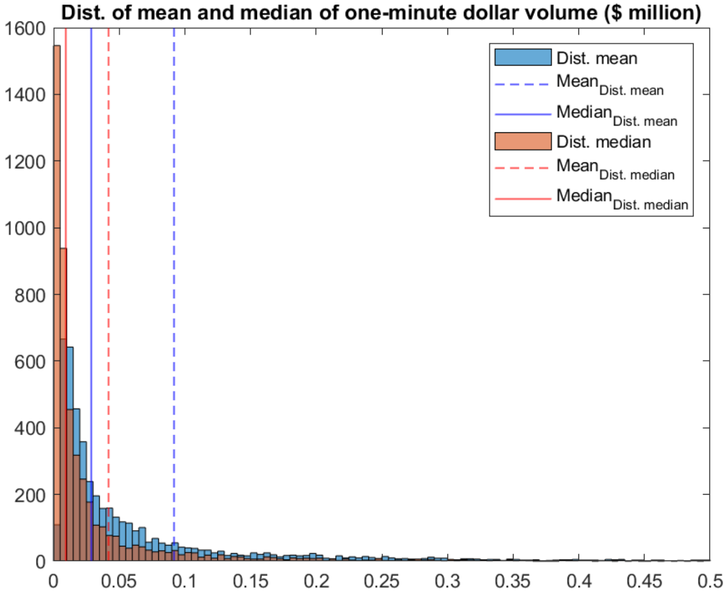


Figure 7: Distribution of the mean and median of the one-minute dollar traded volume for stocks in the sample. The distribution of the mean has an average of 0.09, a median of 0.03 and a standard deviation of 0.26. The distribution of the median has an average of 0.04, a median of 0.01 and a standard deviation of 0.26. For a better resolution of the graph, the horizontal axis is truncated in 0.5.

lar volume[10] over the past d days, from $t-d+1$ to t. The variables computed with this method only evaluate the dispersion within a single day, but they do not involve the variation across days. The following quantities are calculated:

- A matrix nT_t of size $(hm-1) \times 4809$ containing for each day t and stock i-th the number of trading minutes;
- A matrix $AvgDV_t$ of size $(hm-1) \times 4809$ containing the average dollar volume traded each day of the interval for stock

[10]This measure coincides with $HFVOL_t^{(d)}$.

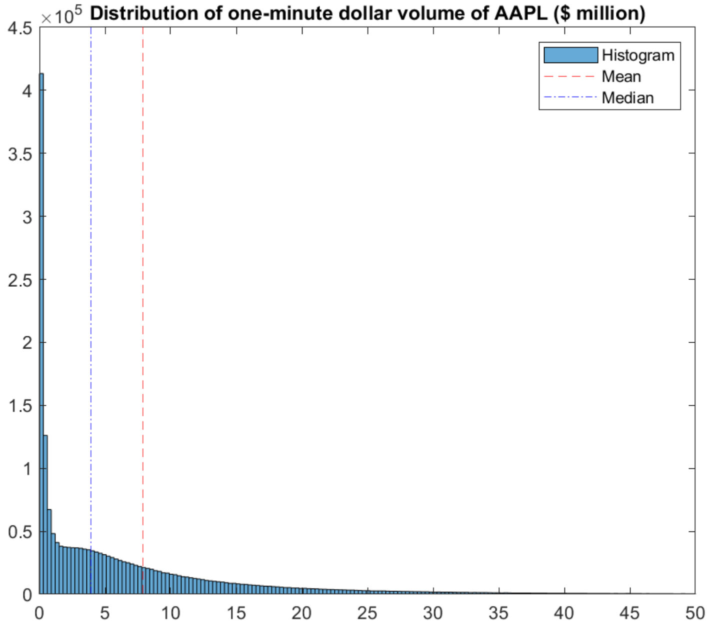


Figure 8: Distribution of one-minute dollar traded volume of the stock Apple (ticker: AAPL). Mean of 7.86, median of 3.93 and standard deviation of 13.93.

i-th; each element of the matrix is given by:

$$AvgDV_t^i = \frac{DV_t^i}{nT_t^i}$$

– A matrix $VarDV_t$ of size $(hm-1) \times 4809$ containing the variance of dollar volume traded in each day of the interval for stock$i$-th; each element of the matrix is given by:

$$VarDV_t^i = \frac{SqrdDV_{t,i}}{nT_t^i} - \left(\frac{DV_t^i}{nT_t^i}\right)^2$$

where $SqrdDV_{t,i} = \sum_{k=1}^{390} (DV_{k,t}^i)^2$ is the sum of the squared one-minute dollar volume for stock i on day t;

The standard deviation of the one-minute dollar volume on day t is defined for each stock i as the mean of the square root of $VarDV_d^i$ in the interval $d = t - hm + 1 : t$:

$$STDVOL_t^{i,(d)} = \frac{\sum_{d=t-hm+1}^{t} \sqrt{VarDV_d^i}}{hm - 1}$$

The natural logarithm of coefficient of variation of the one-minute dollar volume on day t is defined for each stock i as the mean of the square root of $VarDV_d^i$ in the interval $d = t - hm + 1 : t$:

$$HFVOL_t^{i,(d)} = \frac{1}{hm - 1} \sum_{d=t-hm+1}^{t} log\left(\frac{\sqrt{VarDV_d^i}}{AvgDV_d^i}\right)$$

- The second approach that consider both the intraday and interday dispersion of the data, is estimated as the standard deviation or the natural logarithm of the coefficient of variation[11] of the one-minute dollar volume over the period from $t-d+1$ to t. This way allows to evaluate how the data vary within the entire aggregation window in a more broad setting, without limiting the dispersion measure to a single day. The following quantities are calculated:
  - For each stock i on day t we compute the sum of trading minutes in the period from $t - hm + 1$ to t:

    $$nT_{t-hm+1:t}^i = \sum_{d=t-hm+1}^{t} TradingMinutes_d^i$$

  - The average dollar volume traded over the period from $t - hm + 1$ to t for each stock i-th; each element is given by:

    $$AvgDV_t^{i,(d)}(Aggr) = \frac{\sum_{d=t-hm+1}^{t} DV_d^i}{nT_{t-hm+1:t}^i}$$

  - The variance of dollar volume traded over the period from $t - hm + 1$ to t for each stock i-th; each element is given by:

    $$VarDV_t^{i,(d)}(Aggr) = \frac{\sum_{d=t-hm+1}^{t} SqrdDV_d^i}{nT_{t-hm+1:t}^i} - \left(\frac{\sum_{d=t-hm+1}^{t} DV_d^i}{nT_{t-hm+1:t}^i}\right)^2$$

[11]This measure coincides with $HFVOL_t^{(d)}(AGGR)$.

where $SqrdDV_d^i = \sum_{d=t-hm+1}^{t} \sum_{k=1}^{390} (DV_{k,d}^i)^2$ is the sum of the squared one-minute dollar volume for stock i in the period from $t - hm + 1$ to t.

The standard deviation of the one-minute dollar volume on day t is defined for each stock i as the square root of $VarDV_t^i(Aggr)$:

$$HFSTDVOL_t^{i,(d)}(AGGR) = \sqrt{VarDV_t^{i,(d)}(Aggr)}$$

The natural logarithm of coefficient of variation of the one-minute dollar volume on day t is defined for each stock i as:

$$HFVOL_t^{i,(d)}(AGGR) = log\left(\frac{\sqrt{VarDV_t^{i,(d)}(Aggr)}}{AvgDV_t^{i,(d)}(Aggr)}\right)$$

Given the greater exhaustiveness of the second technique, we present in the following the summary statistics obtained by considering both sources of variability in the data. We compute the standard deviation and the natural logarithm of the coefficient of variation over 36 aggregation window. Specifically, for each interval of length d days and stock i, we compute the volatility measures over the past $h \cdot m$ days[12], with $h = 21$ days and $m = 1, \ldots, 36$. The operation is repeated recursively, and at each time step t we shift the window ahead of h days[13]. For both proxies of volatility, we end up with 36 matrices of size 245 $\times$ 4809. Each element of the matrix is the standard deviation or the natural logarithm of the coefficient of variation between day $t - hm + 1$ and t. As in the case with daily sampled data, the matrix m-th has the first m rows empty because we need a sufficient amount of observations to compute the volatility.

In Table (4) left panel, we report the mean, the median and the standard deviation of the standard deviation of one-minute dollar volume as a function of the aggregation windows (m). The summary statistics

[12]Even if it is possible to compute the dispersion measures on a window shorter than a month, for sake of clarity and comparability with the measures obtained with daily data, the shortest aggregation window has length 21 days ($m = 1$).

[13]We do not compute the measures for each day of the sample, but between two subsequent time steps t and $t + h$ the window is shifted of h days.

	Standard Deviation			Coefficient of Variation		
Aggregation window (months)	Mean ($ million)	Median ($ million)	Std. dev. ($ million)	Mean	Median	Std. dev.
1	0.1887	0.0640	0.4797	0.7935	0.7442	0.4187
2	0.1961	0.0697	0.4832	0.8640	0.8106	0.4260
3	0.1997	0.0727	0.4840	0.9030	0.8461	0.4304
4	0.2015	0.0745	0.4836	0.9281	0.8695	0.4339
5	0.2027	0.0757	0.4829	0.9478	0.8881	0.4363
6	0.2035	0.0766	0.4818	0.9637	0.9035	0.4380
7	0.2040	0.0773	0.4804	0.9766	0.9160	0.4394
8	0.2042	0.0779	0.4790	0.9878	0.9268	0.4405
9	0.2044	0.0783	0.4775	0.9977	0.9365	0.4414
10	0.2044	0.0787	0.4760	1.0064	0.9447	0.4421
11	0.2043	0.0790	0.4744	1.0143	0.9523	0.4427
12	0.2042	0.0793	0.4726	1.0215	0.9593	0.4433
13	0.2039	0.0794	0.4710	1.0281	0.9655	0.4438
14	0.2036	0.0795	0.4691	1.0342	0.9716	0.4442
15	0.2032	0.0796	0.4673	1.0398	0.9772	0.4445
16	0.2027	0.0796	0.4645	1.0450	0.9824	0.4447
17	0.2022	0.0796	0.4603	1.0499	0.9872	0.4449
18	0.2016	0.0796	0.4583	1.0545	0.9916	0.4450
19	0.2010	0.0795	0.4564	1.0589	0.9962	0.4451
20	0.2004	0.0795	0.4540	1.0630	1.0005	0.4452
21	0.1998	0.0793	0.4520	1.0670	1.0044	0.4452
22	0.1991	0.0792	0.4501	1.0707	1.0081	0.4453
23	0.1984	0.0792	0.4481	1.0743	1.0119	0.4453
24	0.1976	0.0791	0.4450	1.0777	1.0154	0.4453
25	0.1968	0.0789	0.4423	1.0809	1.0189	0.4452
26	0.1959	0.0787	0.4399	1.0840	1.0221	0.4452
27	0.1950	0.0784	0.4365	1.0870	1.0254	0.4451
28	0.1941	0.0782	0.4342	1.0899	1.0285	0.4451
29	0.1932	0.0779	0.4320	1.0927	1.0315	0.4450
30	0.1923	0.0777	0.4297	1.0955	1.0345	0.4449
31	0.1914	0.0774	0.4275	1.0982	1.0374	0.4448
32	0.1904	0.0771	0.4252	1.1008	1.0401	0.4447
33	0.1895	0.0768	0.4230	1.1034	1.0428	0.4446
34	0.1885	0.0765	0.4207	1.1058	1.0457	0.4445
35	0.1875	0.0761	0.4184	1.1082	1.0484	0.4444
36	0.1865	0.0758	0.4161	1.1106	1.0509	0.4443

Table 4: The summary statistics (mean, median and standard deviation) represent the time-series averages of the cross-sectional statistics for the Standard Deviation (left panel) and the Coefficient of Variation (right panel) of the one-minute dollar volume computed over the past hm days where $h = 21$ is the number of trading days in a month and $m = 1, \ldots, 36$ is the number of months.

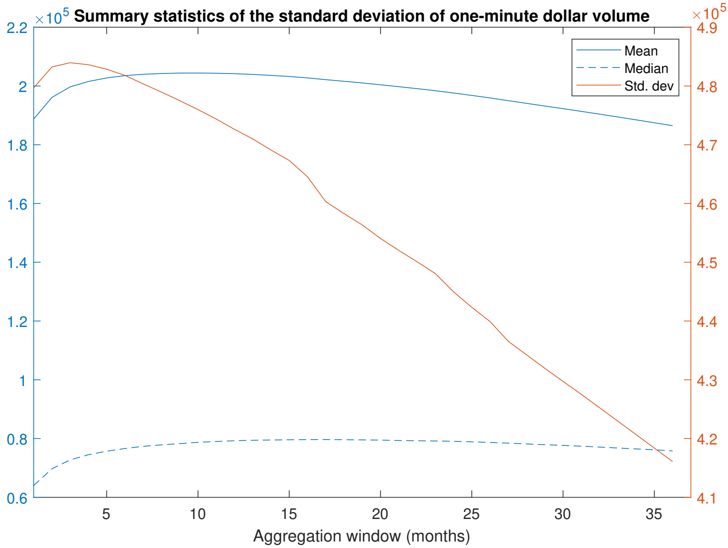


Figure 9: Summary statistics (mean, median and standard deviation) of the standard deviation of one-minute dollar volume computed over different aggregation windows of length $m = 1, \ldots, 36$ months. The statistics are computed as the mean, median, standard deviation respectively of the distribution of the standard deviation of one-minute dollar volume computed over the past months.

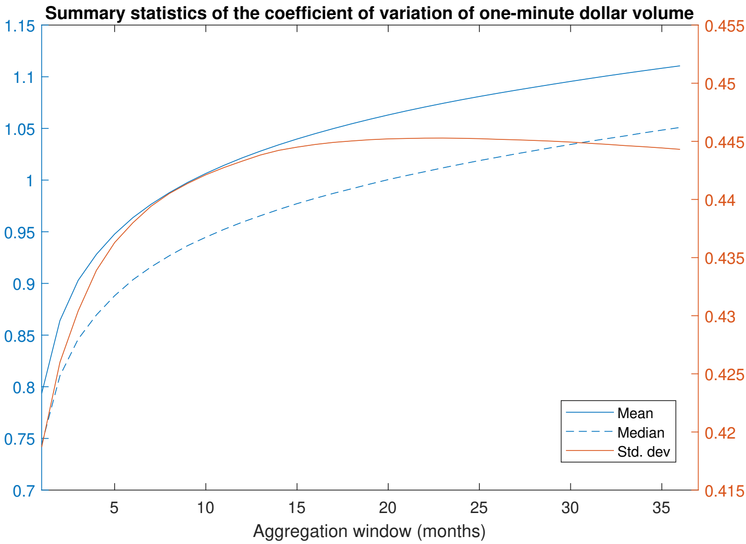


Figure 10: Summary statistics (mean, median and standard deviation) of the natural logarithm of coefficient of variation of one-minute dollar volume computed over different aggregation windows of length $m = 1, \ldots, 36$ months. The statistics are computed as the mean, median, standard deviation respectively of the distribution of the natural logarithm of coefficient of variation of one-minute dollar volume computed over the past months.

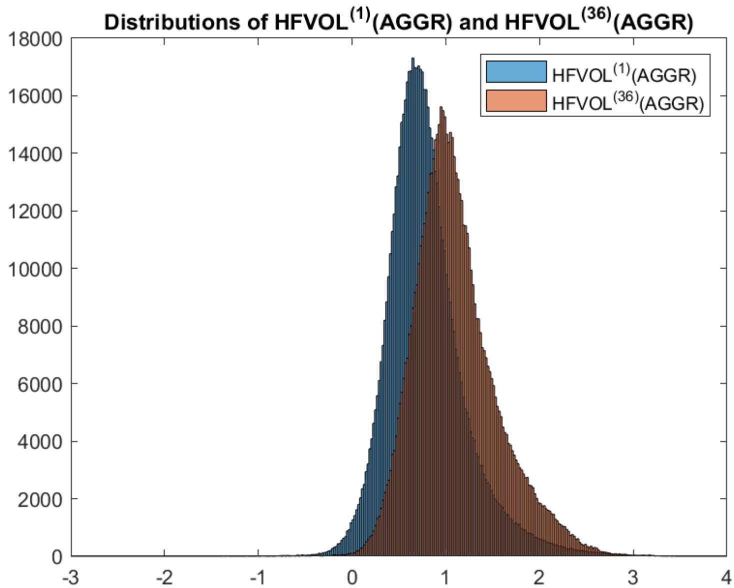


Figure 11: Distribution of the natural logarithm of coefficient of variation computed over the past previous month ($m = 1$, $HFVOL^1(AGGR)$) and over the past 36 months ($m = 36$, $HFVOL^{36}(AGGR)$).

are also plotted in Figure (9). We note that the location measures (mean and median) are initially increasing in m, when $m < 10$ in the case of the mean and $m < 18$ for the median and they smoothly decline when $m > 10$ and $m > 18$, respectively. At the contrary, the volatility of the standard deviation after an initial increase in $m < 3$, it starts a sharp drop as the aggregation window increases. The summary statistics of the natural logarithm the coefficient of variation of one-minute dollar volume are presented in the right panel of Table (4) and plotted in Figure (10). The mean and the median are constantly increasing as the aggregation window widens; also the dispersion continually amplifies until $m = 33$ and it moderately declines in the last three windows. In Figure (11) we display the distributions of the natural logarithm of the coefficient of variation computed over the previous past month (m= 1,$HFVOL(21 \cdot 1)$) and over the past 36 months (m= 36, $HFVOL(21 \cdot 36)$). When the aggregation window widens, the distributions tend to shift to the right and to become more disperse, as in the case with daily data.
We finally compute for each aggregation window of length $m = 1, \ldots, 36$ the correlations between the stock characteristics. The variables involved in the correlation analysis are the following. $RET_{t:t+h}$, $DVOL_t$ (the dollar traded volume in the period from $t-h$ to t), $HFSTDVOL_t^{(d)}(AGGR)$ and $HFVOL_t^{(d)}(AGGR)$, respectively, the standard deviation and the natural logarithm of the coefficient of variation of one-minute dollar volume over the past $d = h \cdot m$ days, for $h = 21$ and $m = 1, \ldots, 36$). Since the measures of dispersion are computed over 36 different aggregation windows, we end up with a collection of 36 manifestations of volatility for $HFSTDVOL_t^{(d)}(AGGR)$ and $HFVOL_t^{(d)}(AGGR)$. We compute the correlation between $RET_{t:t+h}$ and $HFSTDVOL_t^{(d)}(AGGR)$, $RET_{t:t+h}$ and $HFVOL_t^{(d)}(AGGR)$, $DVOL_t$ and $HFSTDVOL_t^{(d)}(AGGR)$, $DVOL_t$ and $HFVOL_t^{(d)}(AGGR)$, $HFSTDVOL_t^{(d)}(AGGR)$ and $HFVOL_t^{(d)}(AGGR)$ for each month t and aggregation window $m = 1, \ldots, 36$. The correlations between the mentioned variables as a function of the aggregation window on which we compute the measure of dispersion are plotted in Figure (12). The correlation between the stock returns and the volatility measures are extremely mild, slightly negative when the standard devi-

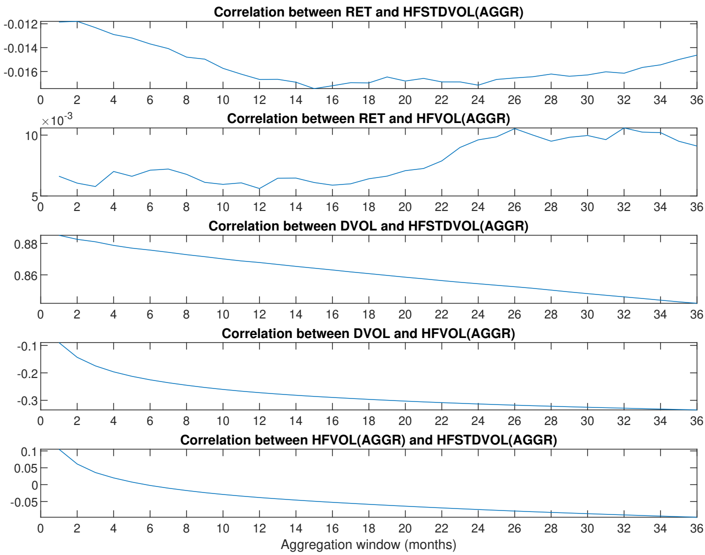


Figure 12: The correlations between $RET_{t:t+h}$ and $HFSTDVOL_t^{(d)}(AGGR)$, $RET_{t:t+h}$ and $HFVOL_t^{(d)}(AGGR)$, $DVOL_t$ and $HFSTDVOL_t^{(d)}(AGGR)$, $DVOL_t$ and $HFVOL_t^{(d)}(AGGR)$, $HFSTDVOL_t^{(d)}(AGGR)$ and $HFVOL_t^{(d)}(AGGR)$ as a function of the aggregation window on which we compute the measure of dispersion. The correlations are computed as the time series average (over the 21-days periods) of the cross-sectional correlations between each pair of variables.

ation is adopted, moderately positive in the case of coefficient of variation. In both cases, the aggregation window (m) does not have an impact on the relation. On the other side, the correlation between $DVOL_t$ and $HFSTDVOL_t^{(d)}(AGGR)$ is strongly positive, reaches its maximum (around 0.88) when the aggregation window is very tight and mildly declines as the aggregation window increases. Also the correlation between $DVOL_t$ and $HFVOL_t^{(d)}(AGGR)$ is quite consistent and negative, it declines at a decreasing rate as long as the aggregation window enlarges. Finally, the correlations between $HFSTDVOL_t^{(d)}(AGGR)$ and $HFVOL_t^{(d)}(AGGR)$ are mildly positive for the first five aggregation windows; the relationship becomes faintly negative for larger windows. Limiting the analysis to $m = 36$, that corresponds to the CSA measures, our results are perfectly in line with the ones of the original paper. The correlation between $DVOL_t$ and $HFSTDVOL_t^{(21\cdot36)}(AGGR)$ is 0.842 in our sample and 0.908 in CSA; the correlation between $DVOL_t$ and $HFVOL_t^{(21\cdot36)}(AGGR)$ is -0.335 in our case and -0.446 in their work; lastly the correlation between
$HFSTDVOL_t^{(21\cdot36)}(AGGR)$ and $HFVOL_t^{(d)}(AGGR)$ is -0.097 for our dataset and -0.243 in their data.

1.4.3 Monthly, Daily, One-Minute data: a comparison

The data aggregated at different time frequencies (monthly, daily, one-minute) allow us to compute variables of dispersion with various granularity and accuracy, compare the performance and the relationship between measures. It is noteworthy comparing the time-series behaviour of the dollar volume aggregated at different frequencies. Figure (13) reports the time-series of the dollar volume of AAPL stock in the last 36 months of the time interval, from June 2015 to June 2018. As already outlined, the monthly dollar volume is computed for each month as the sum of the daily dollar volume traded in the month; the daily dollar volume is computed as the sum of the dollar volume within each day; the one-minute dollar volume represents the dollar volume traded for each trading minute of the sample and it is the quantity with highest granularity. When monthly measure are used, for the entire month (alis for an interval of 21 days or $21 \cdot 390$ minutes) we have a constant quantity, and the set of data we use to compute the uncertainty of liquidity is made of only 36 observations. Dealing with daily observations of dollar volume, implies an higher variability of data generated by the greater sampling frequency: on an aggregation window of 36 months (as the one depicted in Figure (13), the variability of liquidity is measured on a set of $36 \cdot 21$ points. Finally, when we exploit the maximum granularity of the data, the variability is computed on a extremely dynamic time-series that keep its value constant for only one-minute of the full interval. Figure (14) is a zoom on the last week of the sample period for stock AAPL, from June 25^{th} to 29^{th} June 2018. It plots only the last five business days of the dataset and emphasizes the substantial difference between using daily and one-minute data[14]. One-minute data allows to take into account the variation of the time-series within the single day: when the market is about to close and at the very beginning of the trading day, the traded volume is significantly higher compared to the rest of the day, when it shows up more stable and without spikes. With high-frequency data we can appreciate a cyclical pattern that we can not distinguish with data aggregated at a lower frequencies. The seasonality in the data is due to the greater activity during the closure of the trading day and the following opening.

We finally present in Table (5) the correlations between the following variables: RET, DVOL, CVVOLM$^{(36)}$, CVVOL$^{(1)}$, CVVOL$^{(36)}$, HFVOL$^{(21)}$(AGGR), HFVOL$^{(21*36)}$(AGGR), RV5m. It is remarkable to evaluate the correlation between the several measures of liquidity uncertainty computed with different frequencies of data and on different aggregation windows. For sake of expositive clearness, we include, as measures of dispersion, the CSA proxy for liquidity risk (the natural logarithm of coefficient of variation computed with monthly data on the past 36 months) CVVOLM$^{(36)}$; the liquidity uncertainty evaluated with daily data on the past month

[14]The time-series of dollar volume computed with monthly data is flat, because during the week we consider it constant.

and the past 36 months, $\mathrm{CVVOL}^{(1)}$ and $\mathrm{CVVOL}^{(36)}$ respectively; the volatility in liquidity computed with one-minute data over the past month and the past 36 months, $\mathrm{HFVOL}^{(21)}$(AGGR) and $\mathrm{HFVOL}^{(21*36)}$(AGGR) respectively. The correlation of the variable RET with all the others is extremely mild; this evidence is confirmed by CSA for the variables we have in common with their work. The correlation of DVOL with the liquidity risk proxies is always negative; the relationship is particularly intense for both the proxies computed with daily data ($\mathrm{CVVOL}^{(1)}$ and $\mathrm{CVVOL}^{(36)}$) and with monthly values ($\mathrm{CVVOLM}^{(36)}$) while the relation with high-frequency liquidity uncertainty proxies is a bit milder especially for the one computed on the single month ($\mathrm{HFVOL}^{(21)}$(AGGR)). All the variables for the liquidity uncertainty are, taken in pairs, positively correlated meaning that they all catch the same piece of information, regardless the aggregation window and the granularity of the dataset. The highest correlation is between the two measures computed on the past 36 months with monthly and daily data ($\mathrm{CVVOLM}^{(36)}$ and $\mathrm{CVVOL}^{(36)}$): despite the different frequency of the data, the time-series have a similar evolution and move in tandem. In general the measures of dispersion computed over the same aggregation window have a substantially high correlation, regardless their frequency (see correlation between $\mathrm{CVVOLM}^{(36)}$ and $\mathrm{HFVOL}^{(21*36)}$(AGGR), $\mathrm{CVVOL}^{(1)}$ and $\mathrm{HFVOL}^{(21)}$(AGGR), $\mathrm{CVVOL}^{(36)}$ and $\mathrm{HFVOL}^{(21*36)}$(AGGR)). Also the measures computed with the same granularity of data but on a different time span have a remarkably great correlation (see correlation between $\mathrm{CVVOL}^{(36)}$ and $\mathrm{CVVOL}^{(1)}$, $\mathrm{HFVOL}^{(21)}$(AGGR) and $\mathrm{HFVOL}^{(21*36)}$(AGGR)). Note that the lowest correlations are between the dispersion measures constructed with one-minute data and the daily/monthly ones computed on different windows length (see correlation between $\mathrm{CVVOLM}^{(36)}$ and $\mathrm{HFVOL}^{(21)}$(AGGR), $\mathrm{CVVOL}^{(1)}$ and $\mathrm{HFVOL}^{(21*36)}$(AGGR), $\mathrm{CVVOL}^{(36)}$ and $\mathrm{HFVOL}^{(21)}$(AGGR)).

1.5 Results

In this section we report the empirical results of the models presented in Table (1.3.4).

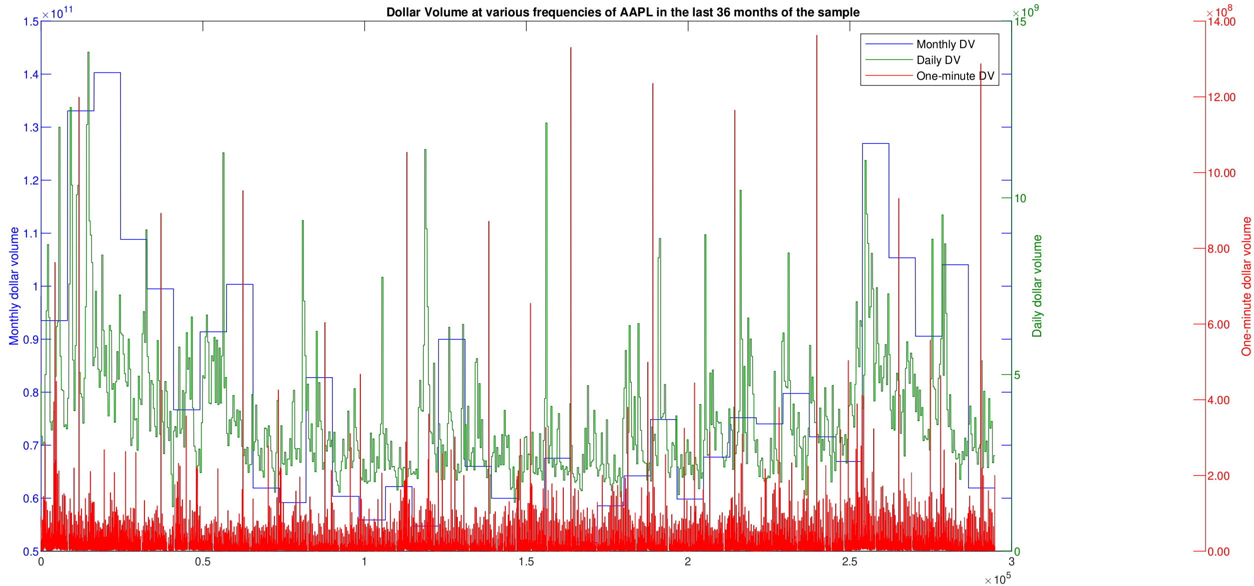


Figure 13: The Figure presents the traded dollar volume aggregated at monthly, daily and one-minute level for AAPL stock in the last 36 months of the time sample, from June 2015 to June 2018. The monthly dollar volume (blu line) is computed for each month as the sum of the dollar volume traded in the month; the daily dollar volume (green line) is computed as the sum of the dollar volume within each day; the one-minute dollar volume (red line) represent the dollar volume traded for each trading minute of the sample.

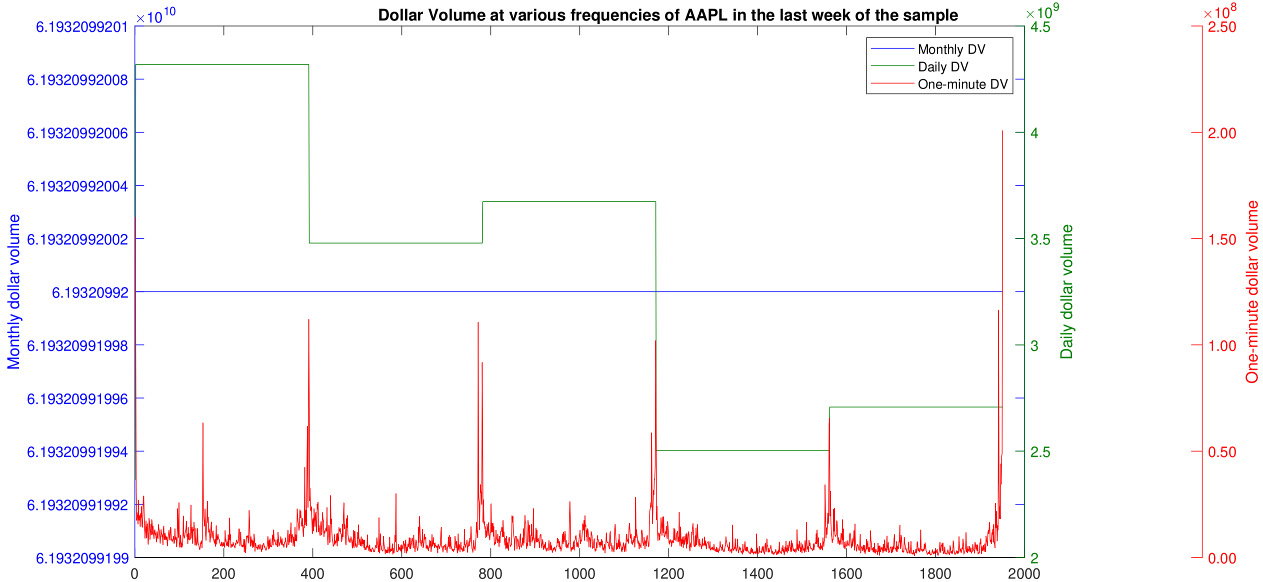


Figure 14: The Figure presents the traded dollar volume aggregated at monthly, daily and one-minute level for AAPL stock in the last week of the time sample, from June 25^{th} to 29^{th} June 2018. The monthly dollar volume (blu line) is computed for each month as the sum of the dollar volume traded in the month; the daily dollar volume (green line) is computed as the sum of the dollar volume within each day; the one-minute dollar volume (red line) represent the dollar volume traded for each trading minute of the sample.

	RET	**DVOL**	**CVVOLM$^{(36)}$**	**CVVOL$^{(1)}$**	**CVVOL$^{(36)}$**	**HFVOL$^{(21)}$ (AGGR)**	**HFVOL$^{(21*36)}$ (AGGR)**	**RV5m**
RET	1.00							
DVOL	-0.05	1.00						
CVVOLM$^{(36)}$	0.03	-0.49	1.00					
CVVOL$^{(1)}$	0.03	-0.55	0.42	1.00				
CVVOL$^{(36)}$	0.03	-0.62	0.86	0.57	1.00			
HFVOL$^{(21)}$ (AGGR)	0.01	-0.12	0.10	0.33	0.17	1.00		
HFVOL$^{(21*36)}$ (AGGR)	0.01	-0.33	0.35	0.26	0.52	0.51	1.00	
RV5m	0.02	-0.20	0.41	0.32	0.33	0.00	0.06	1.00

Table 5: This Table presents time series averages of monthly cross-sectional correlations between stock characteristics. RET denotes the monthly return in the interval $[t : t + h]$, DVOL is the logarithm of dollar volume traded in the previous month with the respect to RET. The measures of dispersion are computed with data at various granularity and on different aggregation window length: CVVOLM$^{(36)}$ is the natural logarithm of the coefficient of variation computed with monthly data over the past 36 months; CVVOL$^{(1)}$ is the natural logarithm of the coefficient of variation computed with daily data over the past month; CVVOL$^{(36)}$ is the natural logarithm of the coefficient of variation computed with daily data over the past 36 months; HFVOL$^{(21)}$(AGGR) is the natural logarithm of the coefficient of variation computed with one-minute data over the past months; HFVOL$^{(21*36)}$(AGGR) is the natural logarithm of the coefficient of variation computed with one-minute data over the past week. RV5m is the realized variance of 5-minute returns. Note: in the Table's labels we omit the time subscripts.

We only introduce the results of the models computed on a window width h of 21 days (models (1), (2), (3), (8), (9) of Table (1.3.4)). Figure (15) shows the t-statistics for the average estimated coefficient of volume $\hat{b}_t^{i,DV}$, as a function of the aggregation window m used. When the CSA measure is involved, (models with $CVVOLM_t^{(m)}$ and $CVVOLM_t^{(m)}(\beta VOL)$ as liquidity risk proxy, hence for models (1) and (2) in Table (1.3.4)), we only use $m = 12, 18, 24, 30, 36$. For each model and each aggregation window on which we compute the volatility of liquidity, we compute the time-series average of $\hat{b}_t^{i,DV}$ for all the time step t and we compute the standard errors of the distribution via Newey-West procedure. We report the t-statistics of the distribution of $\hat{b}^{i,DV}$; in fact, given the definition of the t-statistic as the ratio between estimated value of a parameter ($\hat{b}^{i,DV}$) to its standard error ($\sigma(\hat{b}^{i,DV})$, which is always positive by definition), the sign of the t-statistics is always the same as the one of the estimated parameter. Hence, considering the t-statistics is double informative: the sign of the t-statistics is informative of the sign of the coefficient, while the magnitude of the t-statistic communicates the level of significance of the estimation.

We find a strong and significant negative association between trading dollar volume and stock returns, which is nearly independent of the measure of liquidity uncertainty adopted.

To test the impact of the choice of the liquidity risk proxy on the sign and significance of the liquidity risk premium, we estimate equation (1.6) multiple times, each time including a different proxy for liquidity risk, and keeping all the other explanatory variables equal. The results are in Figure (16), which reports the Newey-West adjusted t-statistics of the average estimated coefficient $\hat{b}_t^{i,LVOL}$ of the measure of liquidity adopted.

The negative relation between expected returns and volatility of volume as defined as in CSA (that is, between expected returns and $CVVOLM_t^{(36)}$ and $CVVOLM_t^{(36)}(\beta VOL)$) is confirmed, although extremely mild. However, the measure used by CSA is extremely flattened and possibly imprecise, since it relies on monthly trading volumes observed in the last three years. The $\text{CVVOL}_t^{(m)}$ and $\text{MADCVVOL}_t^{(m)}$ are similar to the CSA measure, but they are obtained with daily volumes instead of monthly

ones. This allows to gain in resolution, and also to aggregate on less past data. When $m = 36$, we can indeed see that the measures constructed from daily data provide a very similar assessment. However, when the aggregation window goes below 10 months, the statistical significance of the negative coefficient strongly declines. When we use $\mathrm{HFVOL}_t^{(m)}$ or $\mathrm{HFVOL}_t^{(m)}$(AGGR), a precise measure of the volatility of volume based on one-minute data, we observe that the relation becomes positive, and the statistical significance of the results is larger when the aggregation window is smaller.
For the sake of consistency with the previous section, we do not present the results obtained with a window length of width $h = 10$ and $h = 5$ hence models (5), (6), (11), (12) and (7), (8), (13), (14) of Table (1.3.4). The differences with the reported results are not substantial.

1.6 Robustness Check, alternative measures of (il)liquidity

In this section we propose the analysis outlined in (1.3), but using different proxies for liquidity and/or a different specification for the dependent variable (instead of the risk adjusted excess returns we use the excess or the raw returns).
We first present the results obtained by using a different proxy for the liquidity and the (risk *un*adjusted) excess returns. The estimation setup is the same: a cross-sectional regression of the excess returns on K security characteristics $Z_{k,t}$ for each time step t:

$$\tilde{R}^i_{t:t+h} = a_t + \sum_{k=1}^{K} b_{k,t} Z^i_{k,t} + \varepsilon^i_{t:t+h} \tag{1.7}$$

where $\tilde{R}^i_{t:t+h}$ denotes the excess return from t to $t+h$ that is equal to the raw return $R^i_{t:t+h}$ minus the 1-month T-bill rate $r^f_{t:t+h}$ properly standardized, and $\varepsilon^i_{t:t+h}$ is zero-mean random noise which is independent across stocks. As a result, we have a series $\hat{b}_{k,t}$ of model (1.7); a time-series average serves to obtain the estimated coefficient $\hat{b}_k$ and the t-statistics of

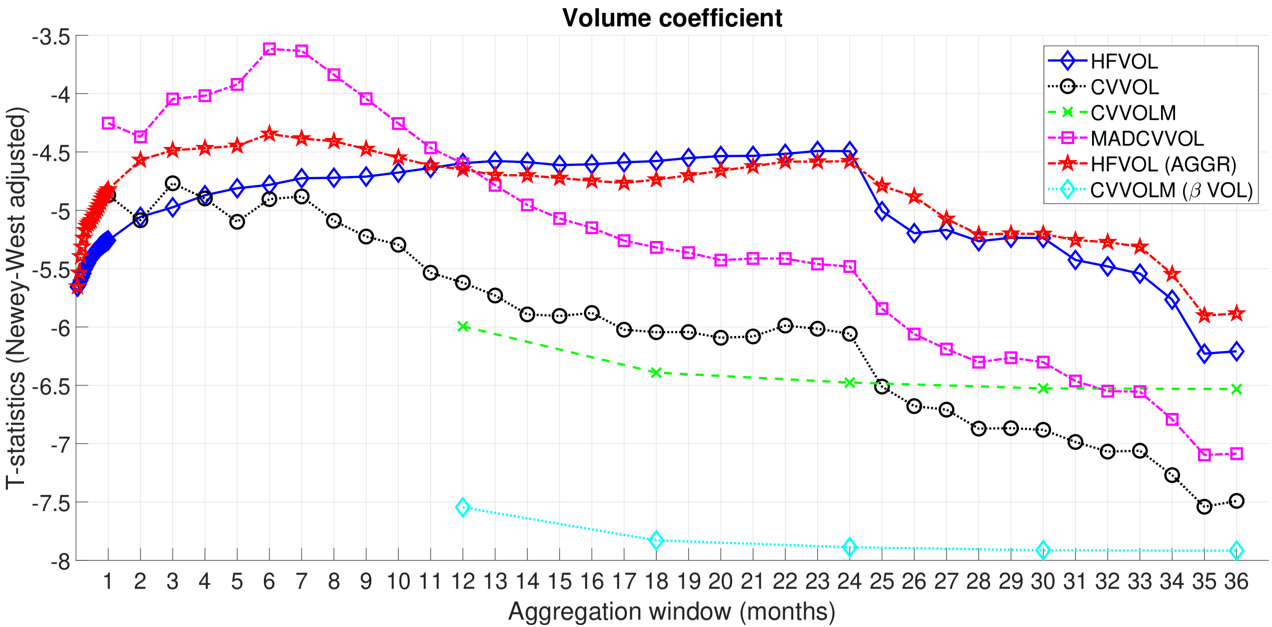


Figure 15: The Figure reports the T-statistic (corrected with Newey-West procedure with 21 lags) of the estimated coefficient $\hat{b}_t^{i,DV}$ (eq. (1.6)) of the volume measure as a function of different aggregation windows, for the following models: $CVVOL\_MONTHLY$ (1) named $CVVOLM$ in the legend, $CVVOLM\_MONTHLY\_BETAVOL$ (2) named $CVVOLM(\beta VOL)$ in the legend, $CVVOLD\_MONTHLY$ (3) $CVVOLD$ named in the legend, $MADCVVOLD\_MONTHLY$ (4) named $MADCVVOLD$ in the legend, $HFVOL\_MONTHLY$ (9) named $HFVOL$ in the legend, $HFVOLAGG\_MONTHLY$ (10) named $HFVOL(AGGR)$ in the legend. The number attached to the model refers to the correspondent ID in Table 1.3.4. The dependent variable is given by the risk adjusted excess returns.

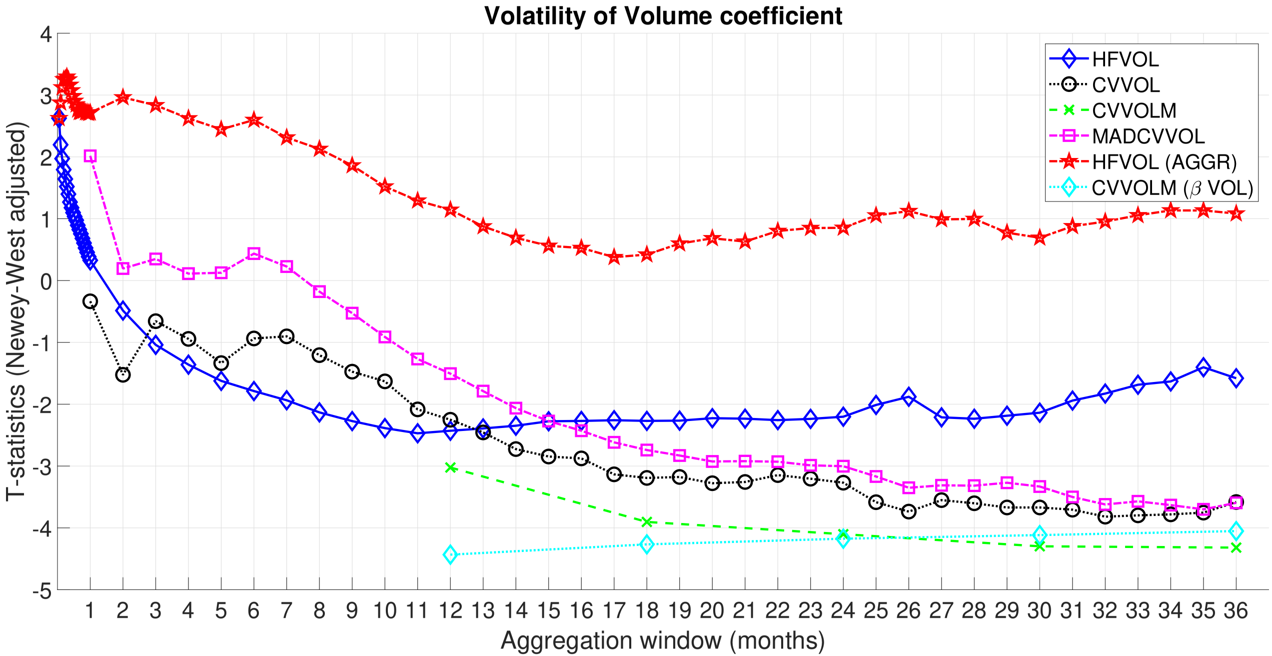


Figure 16: The Figure reports the T-statistic (corrected with Newey-West procedure with 21 lags) of the estimated coefficient $\hat{b}_t^{i,LVOL}$ (eq. (1.6)) of the volatility of volume measure as a function of different aggregation windows, for the following models: $CVVOL\_MONTHLY$ (1) named $CVVOLM$ in the legend, $CVVOLM\_MONTHLY\_BETAVOL$ (2) named $CVVOLM(\beta VOL)$ in the legend, $CVVOLD\_MONTHLY$ (3) $CVVOLD$ named in the legend, $MADCVVOLD\_MONTHLY$ (4) named $MADCVVOLD$ in the legend, $HFVOL\_MONTHLY$ (9) named $HFVOL$ in the legend, $HFVOLAGG\_MONTHLY$ (10) named $HFVOL(AGGR)$ in the legend. The number attached to the model refers to the correspondent ID in Table 1.3.4. The dependent variable is given by the risk adjusted excess returns.

Model ID	Model Name	Independent Variables: Liquidity Level	Independent Variables: Liquidity Risk	Independent Variables: Control Variables	Window Length (h)	Aggregation Periods (m or d)
(15)	AMIHUD_MONTHLY	$AMIHUD_t$	$AMIHUD_t^{(m)}$	Momentum variables	21	$d = 2, \ldots, 20$ d=hm, m=1,...,36
(16)	ZEROS_MONTHLY	$ZEROS_t$	$ZEROS_t^{(m)}$	Momentum variables,	21	$h \cdot m$, m=1,...,36

Table 6: The Table summarizes the models we estimate as a robustness check. The name of the model specifies the proxy used for liquidity. In model (15) we use the Amihud ratio while in model (16) the liquidity is proxied by the measure Zeros. The liquidity uncertainty is computed with daily data on different aggregation periods. The dependent variable is computed as the monthly excess return from t to $t + h$.

the parameter is computed with the Newey-West procedure with h lags. To proxy the (il)liquidity we use two alternative specifications: the Amihud ratio and the Zeros measure. In particular, the Amihud ratio catches the impact in absolute terms on the daily returns of a unit of traded dollar volume in the same time period: the higher the absolute variation of returns when a unit of dollar volume is traded, the more illiquid the stock is. When a unit of dollar volume is traded, it generates an effect on the stock returns: if the price impact is significant, it means that the price is not able to absorb the movement of a unit of dollar volume and it reacts substantially. Hence, if we define the liquidity as the easiness of trading without generating consistent effect on the price, the Amihud ratio clearly represent a measure of illiquidity. With regards to Zeros measure, it computes the percentage of trading minutes per day that present zero one-minute returns: that's why we consider it as a measure of liquidity. In model (15) of Table (6), the illiquidity level is computed over the period from $t-h+1$ to t with the measure $AMIHUD_t$, while the illiquidity uncertainty is proxied by $AMIHUDVOL_t^{(m)}$ over the past $h \cdot m$ days, with $h = 21$ (section (1.3.3) for additional details). In model (16) of Table (6), the illiquidity level is computed over the period from $t - h + 1$ to t with the measure $ZEROS_t$, while the liquidity uncertainty is proxied by $ZEROSVOL_t^{(m)}$ computed over the past $h \cdot m$ days, with $h = 21$. In concrete terms, the models (15) and (16) read respectively as

$$\tilde{R}^i_{t:t+h} = a_t + b_t^{i,AMIHUD} AMIHUD_t^i + b_t^{i,AMIHUDVOL} AMIHUDVOL_t^i + \sum_{k=1}^{K} b_{k,t} Z^i_{k,t} + \varepsilon^i_{t:t+h} \tag{1.8}$$

$$\tilde{R}^i_{t:t+h} = a_t + b_t^{i,ZEROS} ZEROS^i + b_t^{i,ZEROSVOL} ZEROSVOL^i + \sum_{k=1}^{K} b_{k,t} Z^i_{k,t} + \varepsilon^i_{t:t+h} \tag{1.9}$$

where we include as control variables, $Z^i_{k,t}$ the past cumulative returns. Figure (17) presents the Newey-West adjusted t-statistics of the average estimated coefficients of the measure of (il)liquidity adopted, $\hat{b}_t^{AMIHUD}$

and $\hat{b}_t^{ZEROS}$. In line with the economic intuition, the illiquidity level and the expected returns have a positive relationship: the estimated coefficient $\hat{b}^{AMIHUD}$ is always positive and strongly significant regardless of the aggregation window on which the illiquidity risk is computed. On the contrary, the measure of liquidity level, proxied by $ZEROS_t$, always presents a negative relation with the returns but pretty mild. The outcome sheds light on how this facet of liquidity does not have a strong relationship with the returns. Differently from the classical measures of (il)liquidity like the dollar volume, which captures the magnitude of the money flow traded during a certain period, or the Amihud ratio, which measures the reaction (in absolute value) of the stock price when a unit of dollar volume is traded, the variable Zeros highlights how many times, in relative terms, the agents are active on the market and they trade stocks without generating impact on the time-series of the price. The measure is constructed with one-minute data: it computes within a day how many trading minutes generate no price impact (hence, a zero return); then, the monthly measure is obtained as an average of the daily variables in the correspondent period. A possible explanation of this mild relationship between the liquidity level and the returns is in the fact that what happens at the very high-frequency level does not have a remarkable impact at a monthly level. The two scales are far enough away that the micro-level events are not impactful on the grand vision.

As additional robustness check we also propose the estimation of models (1), (3), (4), (9) and (10) of Table (1.3.4) using as dependent variable the excess returns instead of the risk adjusted excess returns. The cross-sectional regressions are in the format of equation (1.6) with the only variation that the dependent variable is given by the excess returns. Figure (19) reports the Newey-West adjusted t-statistics of the average estimated coefficients of the liquidity level $\hat{b}_t^{i,DV}$ proxied by the dollar traded volume in the previous month with the respect to the time when the returns are computed. Results are perfectly in line with the ones reported in Figure (15): the relationship between liquidity level and the excess returns is strongly negative, even if the significance is slightly less

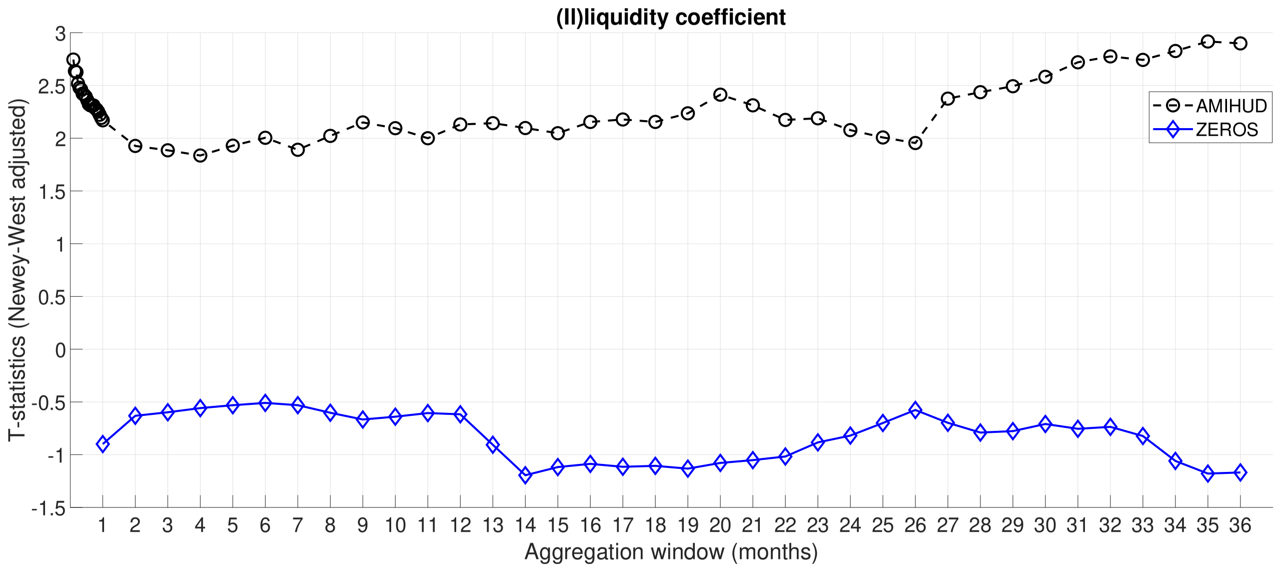


Figure 17: The Figure reports the T-statistic (corrected with Newey-West procedure with 21 lags) of the estimated coefficient $\hat{b}^i_{t,AMIHUD}$ and $\hat{b}^i_{t,ZEROS}$ as a function of different aggregation windows on which (il)liquidity risk measure is computed.

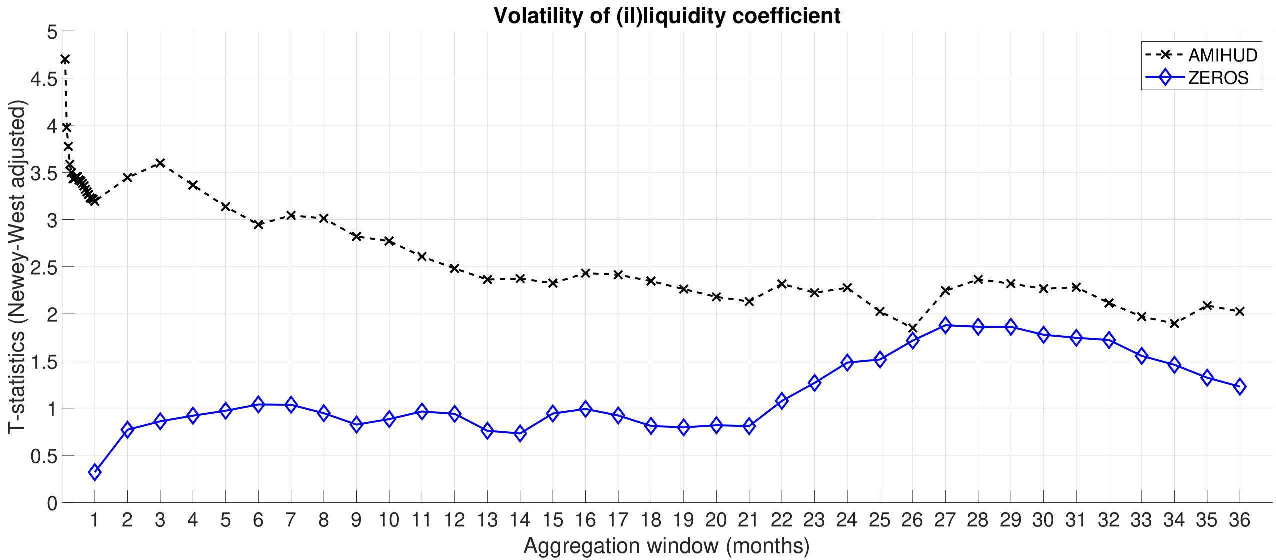


Figure 18: The Figure reports the T-statistic (corrected with Newey-West procedure with 21 lags) of the estimated coefficient $\hat{b}^i_{t,AMIHUDVOL}$ and $\hat{b}^i_{t,ZEROSVOL}$ as a function of different aggregation windows on which (il)liquidity risk measure is computed.

strong than in the case with risk adjusted excess returns. Also the relationship between the coefficient of variation of dollar volume and the excess returns, reported in Figure (20), is similar to the one between the liquidity risk and the risk adjusted excess returns: when low frequency measures (montlhy and daily) are employed to estimate the liquidity uncertainty and the aggregation window is very wide and consequently distant to the period over which returns are computed, the significance of the coefficients is very mild, the sign of estimated parameters is negative for aggregation windows larger than 12 months. On the contrary, when one-minute data are involved in the evaluation of liquidity risk, the estimated coefficients are strongly positive for very short windows while the robustness of the results smoothly declines as we enlarge the aggregation window.

1.7 Self-financing portfolio performance

This section presents the performance of self-financing with monthly rebalancing portfolios built with different selection criteria to test whether it might be economically convenient to construct a portfolio sorted on liquidity uncertainty. We propose an analysis of portfolios' performance in two different frameworks: a less realistic case, when the performance is not eroded by the transaction costs, and a more realistic scenario with the performance of a long strategy net of the transaction costs.

1.7.1 Portfolio gross of transaction cost

We construct the self-financing portfolios based on a monthly re-balancing, sorting the involved stocks according to various criteria.
At each time step, i.e. at each month t, we identify the active stocks, namely the stocks which have the absolute return in the current and following month lower than 100% ($|R_t| < 100\%$ and $|R_{t+1}| < 100\%$), and the values for the monthly dollar volume (DV_t^i), the natural logarithm of

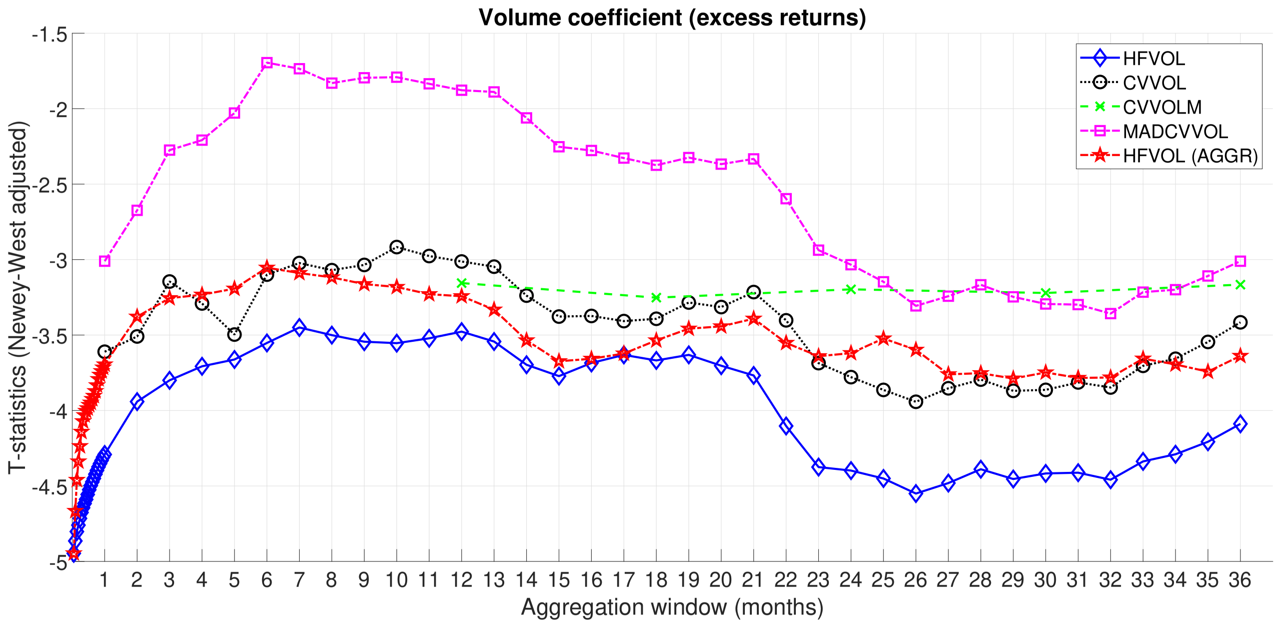


Figure 19: The Figure reports the T-statistic (corrected with Newey-West procedure with 21 lags) of the estimated coefficient $\hat{b}_t^{i,DV}$ (eq. (1.6)) of the volume measure as a function of different aggregation windows, for the following models: $CVVOL\_MONTHLY$ (1) named $CVVOLM$ in the legend, $CVVOLD\_MONTHLY$ (3) $CVVOLD$ named in the legend, $MADCVVOLD\_MONTHLY$ (4) named $MADCVVOLD$ in the legend, $HFVOL\_MONTHLY$ (9) named $HFVOL$ in the legend, $HFVOLAGG\_MONTHLY$ (10) named $HFVOL(AGGR)$ in the legend. The number attached to the model refers to the correspondent ID in Table 1.3.4. The dependent variable is given by the excess returns.

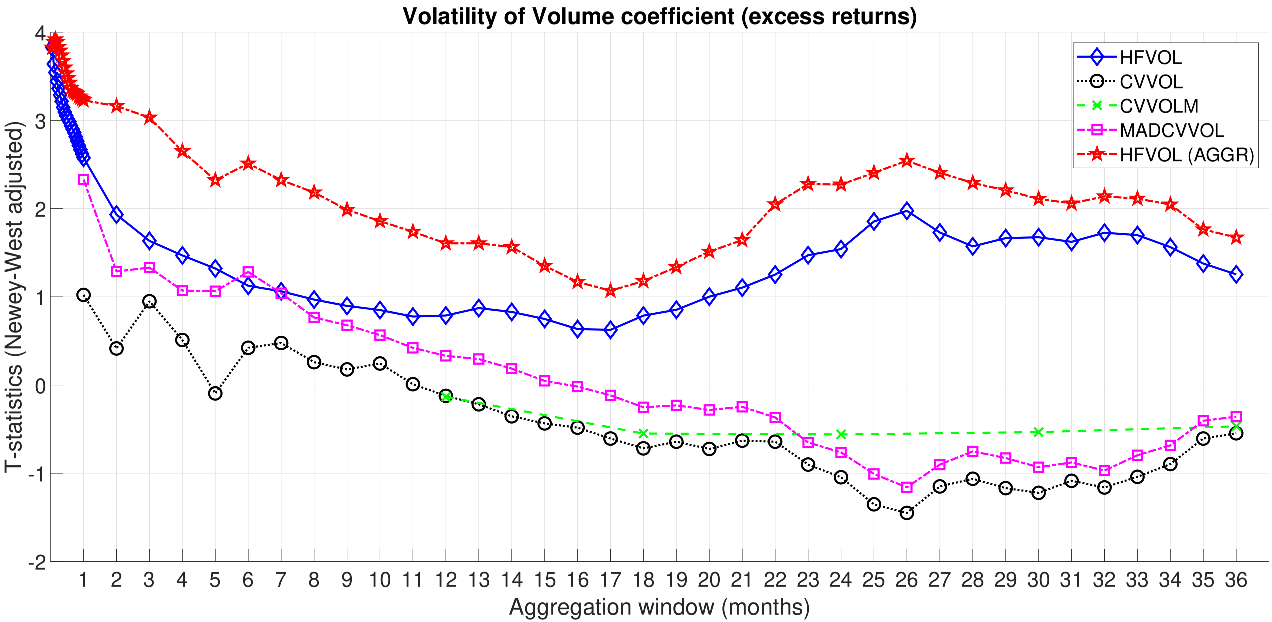


Figure 20: The Figure reports the T-statistic (corrected with Newey-West procedure with 21 lags) of the estimated coefficient $\hat{b}_t^{i,LVOL}$ (eq. (1.6)) of the volatility of volume measure as a function of different aggregation windows, for the following models: $CVVOL\_MONTHLY$ (1) named $CVVOLM$ in the legend, $CVVOLD\_MONTHLY$ (3) $CVVOLD$ named in the legend, $MADCVVOLD\_MONTHLY$ (4) named $MADCVVOLD$ in the legend, $HFVOL\_MONTHLY$ (9) named $HFVOL$ in the legend, $HFVOLAGG\_MONTHLY$ (10) named $HFVOL(AGGR)$ in the legend. The number attached to the model refers to the correspondent ID in Table 1.3.4. The dependent variable is given by the excess returns.

the coefficient of variation computed with monthly data over the past 36 months ($CVVOLM_t^{(36)}$), the natural logarithm of the coefficient of variation computed with one-minute data over the past month ($HFVOL_t^{(21)}$).
We construct portfolios sorting stock on their values of liquidity and/or liquidity risk to test whether a strategy based on liquidity uncertainty can outperform the market portfolio.
We use as a benchmark the equally weighted portfolio, an investment strategy that includes all the available stocks in the formation of the portfolio, attributing each stock the same weight. If in month t we dispose of N_t stocks, each stock has a weigh of $w\_eq_t = 1/N_t$, and the portfolio return in month t is the cross sectional average of stock returns

$$R_t^{Mkt} = \frac{1}{N_t} \sum_{i=1}^{N_t} R_t^i \tag{1.10}$$

Its cumulative performance evolution during the sample period 1998-2018 is reported in Figure (21), marked with the line 'Market'.
In what follows, we present five different long-short strategies obtained by sorting stocks on liquidity and/or liquidity risk measure:

- Portfolio sorted on monthly dollar volume: each month stocks are sorted (in descending order) according to their dollar-volume value in the previous month; the first and forth quartile of the sorted stocks implement the long-short strategy: the portfolio goes long on lowest volume stocks and short on the highest volume stocks (liquidity level has a negative impact on stock returns, hence the stocks with higher liquidity have on average a lower expected return. Going short on these stocks and long on the opposite ones make the approach profitable).

- Portfolio sorted on the 'low-frequency' measure used by CSA: each month stocks are sorted according to their $CVVOLM^{(36)}$ value in the previous month; the portfolio goes long on the stocks in the quartile with the lowest liquidity uncertainty and short on stocks in the quartile with the highest liquidity uncertainty ($CVVOLM^{(36)}$

has a negative relationship with expected stock returns, stock with lower $CVVOLM^{(36)}$ should command higher returns).

- Portfolio sorted on the 'high-frequency' measure of liquidity risk ($HFVOL^{(21)}$): each month stocks are sorted according to their $HFVOL^{(21)}$ value in the previous month; the portfolio goes long on the stocks in the quartile with the highest $HFVOL^{(21)}$ and short on stocks in the quartile with the lowest $HFVOL^{(21)}$ ($HFVOL^{(21)}$ has a positive relationship with expected stock returns, stock with higher HFVOL should command higher returns).

- Portfolio double sorted on Volume and $CVVOLM^{(36)}$: we first sort based on the last month traded dollar volume, and we consider stocks among and above the median. Then, we sort again based on liquidity uncertainty. When this is proxied by $CVVOLM_t^{(36)}$, we select, among stocks with volume above the median, the 50% of stocks with the highest volatility of volume, and we go short on these. We then select, among stocks with volume below the median, the 50% of stocks with the lowest volatility of volume, and we go long on these.

- Portfolio double sorted on Volume and $HFVOL^{(21)}$: we first sort based on the last month traded dollar volume, and we consider stocks among and above the median. Then, we sort again based on liquidity uncertainty. The double-sorted porfolio based on $HFVOL_t^{(21)}$ is composed in a similar way, but now we select, among stocks with volume above the median, the 50% of stocks with *the lowest* volatility of volume and we go short, and among stocks with volume below the median, the 50% of stocks with *the highest* volatility of volume and we go long, since, again, the sign for this characteristic is positive.

The performances of all the portolios' strategies descibed above are reported in Figure (21). We compute the performance as the cumulative product of monthly portfolios' returns; each month t, the performance is $R_t^p = (1 + R_1^p)(1 + R_2^p) \dots (1 + R_t^p)$ where p stands for one of the strategies outlined above. It is worth noticing that all the returns are gross of transaction cost, and therefore this scheme is not representative of the reality. Estimating the transaction costs that erode the portfolio's performance makes the exercise more realistic.

1.7.2 Transaction cost estimation

Since we do not dispose of the monthly transaction costs, we follow the procedure introduced by Abdi and Ranaldo (2017), a recent paper where the authors propose a method to estimate the bid-ask spread based on close, high and low daily prices.
For each stock, we take into account the close price for each day on which the security is traded (if the close price on the last trading minute is not available, we consider the most recent accessible price); the high (low) price is the maximum (minimum) traded price of the day. We define c_t^i the observable daily close log-price on day t for the $i-th$ stock, h_t^i and l_t^i as the high and low log-price. The authors propose as proxy for the efficient price, the daily mid-range, computed as the mean of the daily high and low log-prices (we omit the summit i to avoid clutter):

$$\eta_t = \frac{l_t + h_t}{2} \tag{1.11}$$

They prove that the squared distance between the close log-price c_t and the midpoint proxy η_t is made of two components: the efficient price variation and the squared effective spread; we are interested in the estimation of the second item, which can be computed as:

$$S^2 = 4\mathbb{E}[(c_t - \eta_t)(c_t - \eta_{t+1})] \tag{1.12}$$

where the expectation is computed as the sample average on the past 21 daily values (with a minimum of 12 active trading days). To estimate the effective proportional spread for every month-stock we take the square root of equation (1.12). However, due to estimation errors, the value of S^2 might be negative; hence, to correct it, the proportional spread is

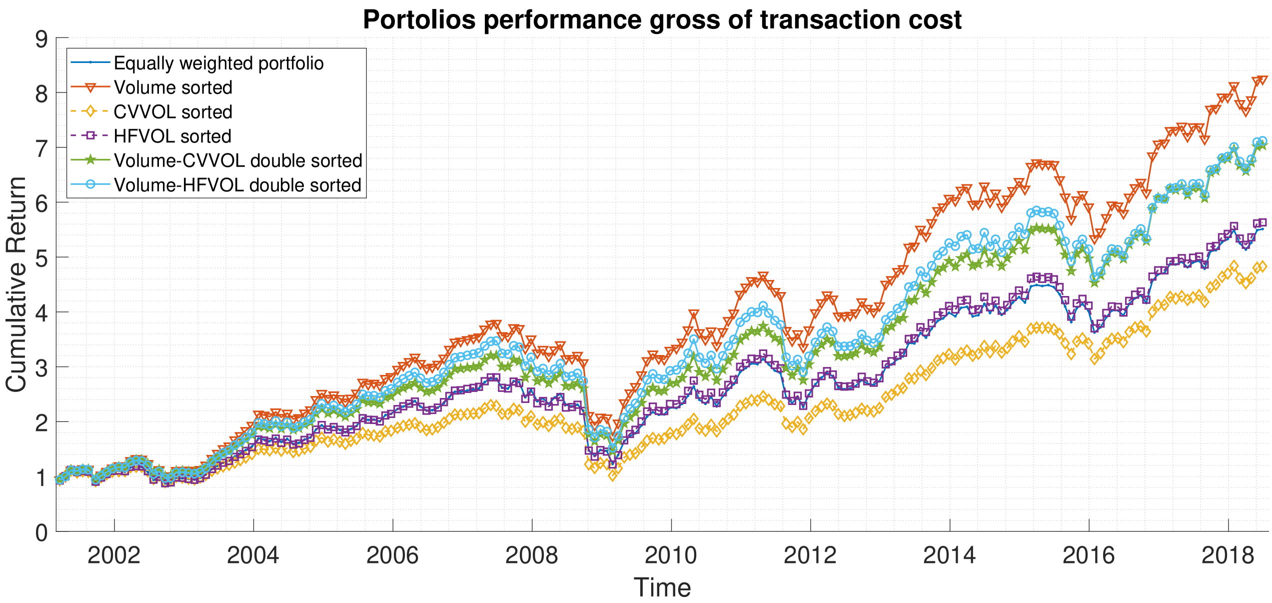


Figure 21: The Figure presents the performance of portfolios built according to different strategies plus the equally weighted portfolio (used as a benchmark). The performance is computed as the percentage monthly cumulative return of the portfolios in the sample period from 1998 to 2018 and it is gross of transaction costs.

computed as:

$$Spread = \sqrt{max(0, S^2)} = \sqrt{max(0, 4\mathbb{E}[(c_t - \eta_t)(c_t - \eta_{t+1})])} \qquad (1.13)$$

The spread can be either 0 or $2\sqrt{\mathbb{E}[(c_t - \eta_t)(c_t - \eta_{t+1})]}$.
Figure (22) shows the daily time series evolution of the spread (in $\%$) computed for the equally weighted or market portfolio: each day, spread is computed as the average across stock of all the stock's spread traded during the past 21 days (or at least active for 12 days). At each time step the 21-days-window is slided one day ahead. In line with the findings of Abdi and Ranaldo (2017), the model display a relatively stable variation over time; the only period with a significant turmoil is the one during the financial crisis.
It is a well known fact the strong negative relationship between the size, proxied by the market capitalization, and the bid-ask spread. Brandt, Santa-Clara, and Valkanov (2009), in their novel approach to optimizing portfolios with large numbers of assets by directly modelling the portfolio weight in each asset as a function of the asset's characteristics, they also calibrate the transaction costs to decline uniformly over time and to decrease with the relative size of the firms. Novy-Marx and Velikov (2016) in Figure 1 of their work, shows cross-sectional and time-series variation in trading costs, by looking at the estimated mean effective spreads of the largest 2,000 firms, by decade over the period 1963-2009. Smaller cap stocks are more expensive to trade. They also show a general trend towards lower costs over time, and a dramatic reduction in the cost of trading stocks outside the mega-cap universe over the last decade. Size is strongly negatively correlated with transaction costs in the cross section, but the effect is nonlinear. The coefficient on the squared market cap variable is positive and significant, implying a convex relation between trading costs and size. Korajczyk and Sadka (2004) are the only one, to the extent of our knowledge, who study the cross-sectional relationship between transaction cost and both the size and the volume. They evaluate the average cross-sectional relation between the transaction costs (estimated with three different models) and firm-specific pre-determined variables, among which the market cap at

the end of last month divided by the average market cap, and the total volume traded during the last three months divided by the average firm volume on NYSE. They uncover the following results: a positive and statistically significant impact of the size measure on the transaction costs, and a negative and significant impact of volume on the transaction costs. The relationship between transaction costs and volume is perfectly in line with our findings in Figure (23), where we present for each stock the relation between the time-series average of estimated transaction costs (with the model outlined above) and DVOL (the natural logarithm of the dollar volume traded in the previous month). The negative relation between the two variables witness that, on average, the higher the liquidity level of a stock, the lower is the average transaction cost. Gerhold et al. (2014) uncover an asymptotic relation among the liquidity premium, trading volume, and transaction costs. They show that share turnover (ShTu), the liquidity premium (LiPr), and the bid-ask spread (ε) satisfy the following approximate equation: $LiPr = 3/4\varepsilon * ShTu$.
It is well known from the literature, that the transaction costs tend to decrease over time and that the more liquid stocks are inclined to have lower costs. Table 7 presents the average transaction costs as a function of time and volume. Each month we sort the stocks according to their dollar-volume level in a descending order and split them into four quartiles: the stocks in the first quartile are the 25% stocks with the highest liquidity level, while the stocks in the bottom quartile are the most illiquid ones. We compute the average transaction cost for each quartile and each month. We also compute for each month the average transaction cost, unconditonal f volume level. We then divide our time sample in 3 intervals: from February 1998 to December 2004, from January 2005 to December 2011 and from January 2012 to June 2018. We aggregate the monthly measures in the three time intervals by computing the time-series mean of the monthly indicators. Results are in Table 7. In accordance with the literature on the theme, the transaction costs are decreasing in time and in liquidity level.
The transaction costs estimated with the procedure described above are appropriate to be applied to a long only portfolio; in fact, in the case of

	1998-2004	2005-2011	2012-2018	1998-2018
Top 25%	0.56	0.34	0.23	0.38
2^{nd} **quartile**	0.60	0.39	0.28	0.43
3^{rd} **quartile**	0.80	0.54	0.37	0.57
Bottom 25%	1.35	0.80	0.69	0.95
Unconditional	0.83	0.52	0.40	0.58

Table 7: The Table presents the average transaction costs as a function of time and liquidity level. For each of the three time intervals (1998-2004, 2005-2011, 2012-2018), and each quartile of the monthly volume-sorted distribution of stocks we compute the average transaction costs. The costs are decreasing in time and liquidity level. In fact, the first quartile of the distribution ('Top 25%') contains the most liquid stocks. The last row presents the average transaction costs as a function of the time interval, unconditional of volume ranking.

long-short strategy, it is necessary to compute transaction costs distinguishing whether we go long or short on the specific stock. Going short on a stock is much more expensive than going long because the agent has to pay a deposit and some extra fee to the stock lender. Hence, to be more realistic in a long-short strategy, transaction cost must be estimated asymmetrically. Symmetric transaction cost, estimated with the Abdi and Ranaldo (2017) procedure, are suitable to simulate strategies that only go long in the portfolio construction.

1.7.3 Asymmetric estimation of transaction cost

Transaction costs can be decomposed in various constituents: the bid-ask spread (estimated with the model of Abdi and Ranaldo (2017)), the exchange fee (not included in this estimation), and in case of shorting strategy the additional constituent is the shorting fee. A short sale is defined as "*the sale of a security that the seller does not own or that the seller owns but does not deliver. In order to deliver the security to the purchaser, the short seller will borrow the security, typically from a broker-dealer or an institutional investor. The short seller later closes out the position by returning the security to the lender, typically by purchasing equivalent securities on the open market. In general, short selling is utilized to profit from an expected downward*

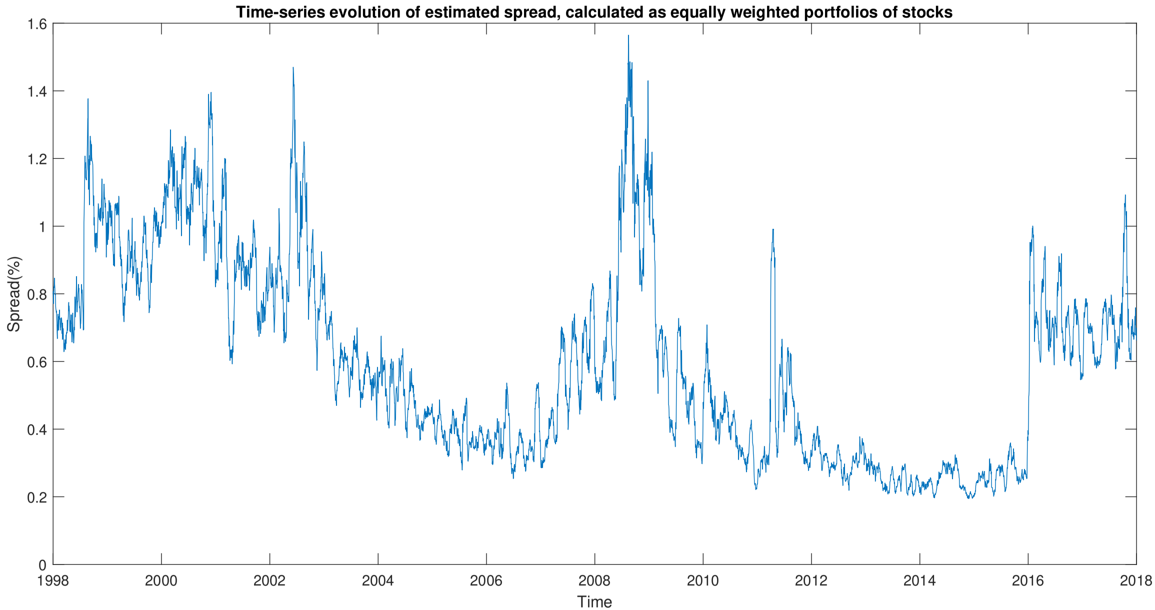


Figure 22: The Figure plots the daily historical evolution of the estimation of the spread with the model proposed by Abdi and Ranaldo (2017). Each value is computed as the spread of an equally weighted portfolio over the previous 21 days. At each time step, we slide the rolling-window one day ahead by including the most recent value and removing the most far away one. We compute the estimate over the time-span from 1998 to 2018 for all the stocks traded on the NYSE and AMEX.

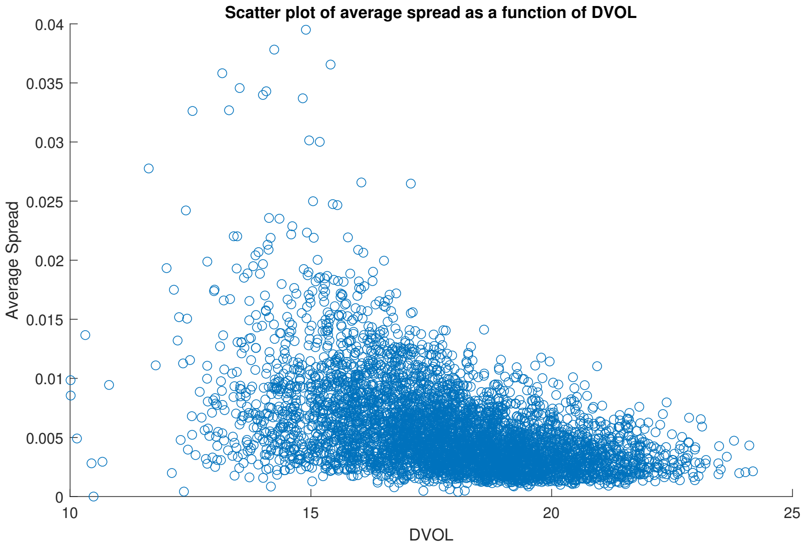


Figure 23: The Figure represents for each stock the relation between the average estimated spread (vertical axis) and the average of variable DVOL (which is the natural logarithm of the dollar volume traded in the previous month).We compute the estimate over the time-span from 1998 to 2018 for all the stocks traded on the NYSE and AMEX.

price movement, or to hedge the risk of a long position in the same security or in a related security[15]". The lender requires a recompense for the loan: for making the shorting possible and for the risk of the securities are not returned to her.
Hence, the cost of borrowing the stock is encapsulated in the shorting (or loan) fee; it is essential taking into account the shorting fee in the simulation of a long-short strategy because the additional cost, the investor encounters when she goes short, may affect the economic convenience of the strategy. The literature proposes some studies related to the shorting strategies and asymmetry of transaction costs.
In particular, D'Avolio (2002) studies the loan position and transaction information for U.S. equity security over period April 2000-September 2001. To summarize his results, 91.3% of analyzed stocks have loan fee $< 1\%$; value-weighted mean loan fee is 0.17%. 8.7% of stocks have a special regime of loan market with an average value-weighted fee of 4.30%.
In the stylized model by Nutz and Scheinkman (2020), investors pay costs which are proportional to the square of their positions (i.e. number of shares, y) but the constant (α) of proportionality that defines the cost of going short may be larger than the corresponding constant for going long. The transaction cost are modeled as

$$\begin{cases} \frac{1}{2\alpha_+}y^2 & \text{if } y > 0 \\ \frac{1}{2\alpha_-}y^2 + \beta_-|y| & \text{if } y < 0 \end{cases} \tag{1.14}$$

where $0 < \alpha_- \leq \alpha_+$: the cost of shorting is higher than the cost of going long.

1.8 Portfolio net of transaction cost

In absence of transaction cost, each month the rebalance of the portfolio is implemented by selling all the stocks and (re)purchasing the new ones. But rebalancing the portfolio is a costly operation, hence only the stocks

[15] Definition proposed by SEC on their web site https://www.sec.gov/rules/concept/34-42037.htm

presenting a variation in their weight in the portfolio are involved in the operation. Specifically, referring to the procedure in DeMiguel et al. (2020), and denoting by R_t^i, the percentage return of stock i in month t, and by $w_{i,t}$ the desired weight of a given portfolio on stock i in month t, with $i = 1, \ldots, N = 4809$ and $t = 1, \ldots, T = 245$, we compute daily turnover (TO_t^w) as

$$TO_t^w = \sum_{i=1}^{N} \left| w_{i,t} - \widetilde{w}_{i,t} \right|, \tag{1.15}$$

where

$$\widetilde{w}_{i,t} = \frac{w_{i,t-1}(1 + R_{i,t})}{\sum_{i=1}^{N} w_{i,t-1}(1 + R_{i,t})}. \tag{1.16}$$

Indicating with $k_{i,t}$ the average (relative, round-trip) transaction cost estimated for stock i in month t, the portfolio return $R_{t+1}^{w,*}$ at month $t+1$ net of transaction cost is obtained by:

$$R_{t+1}^{w,*} = \sum_{i=1}^{N} w_{i,t} R_{i,t} - \sum_{i=1}^{N} \frac{k_{i,t}}{2} \left| w_{i,t} - \widetilde{w}_{i,t} \right|. \tag{1.17}$$

The performance of the six portfolios is evaluated using Sharpe Ratios net of transaction costs. For each of the portfolios under consideration, we compute average monthly turnover TO^w (see Figure (24)) and out-of-sample annualized mean, standard deviation and Sharpe Ratio as follows:

$$TO^w = \frac{1}{T-37} \sum_{t=37}^{T-1} TO_t^w,$$

$$\mu^w = \frac{12}{T-37} \sum_{t=37}^{T-1} (R_{t+1}^{w,*} - r_t^f),$$

$$\sigma^w = \left(\frac{12}{T-37} \sum_{t=37}^{T-1} (R_{t+1}^{w,*} - r_t^f - \mu^w)^2 \right)^{1/2},$$

$$SR^w = \frac{\mu^w}{\sigma^w},$$

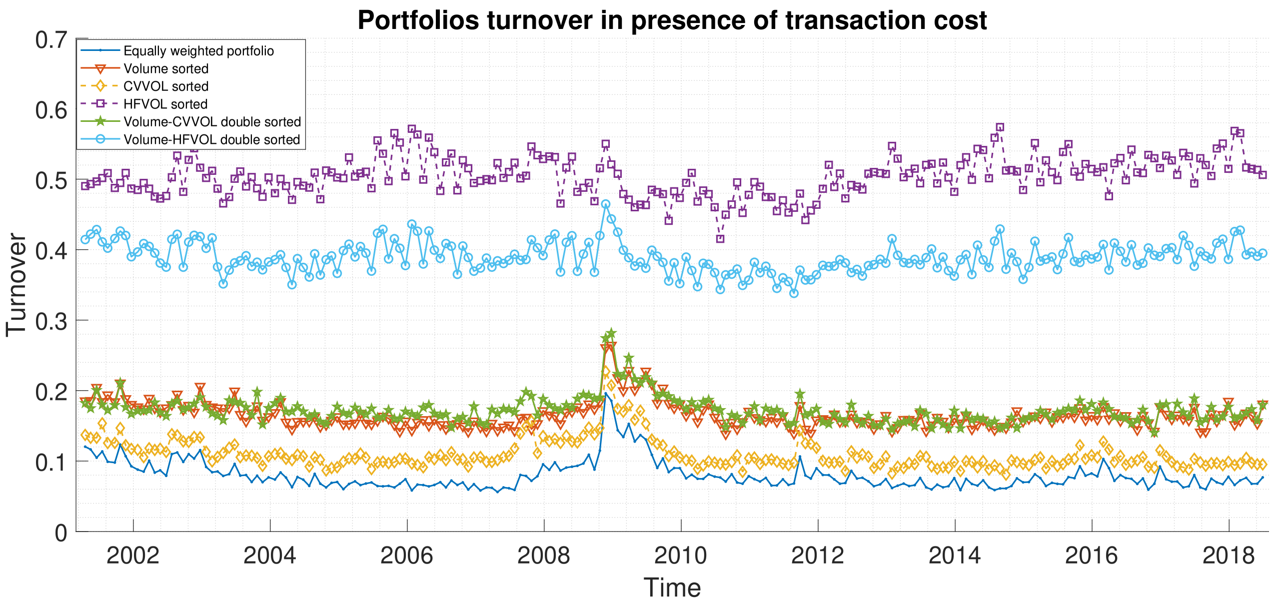


Figure 24: The Figure presents the monthly turnover over the sample period for the implemented portfolio's strategies accounting for the transaction costs.

where r_t^f is the risk-free rate of month t and $T = 245$. To test whether the out-of-sample performance of the six portfolio is statistically significantly different from the equally weighted portfolio using the iid bootstrap method in Ledoit and Wolf (2008), with 10,000 bootstrap samples to construct a one-sided confidence interval for the difference between Sharpe ratios. We use three/two/one asterisks ($*$) to indicate that the difference is significant at the 0.01/0.05/0.10 level. Table 8 reports the out-of-sample performance of the considered portfolios, without transaction costs (Panel A) and with transaction costs (Panel B). From Panel A we see that the performance of the volume-sorted portfolio is significantly larger than the equally weighted portfolio, that is that we can earn an out-of-sample premium by exploiting the liquidity premium. Interestingly, while double-sorted portfolio have a higher Sharpe ratio than the equally weighted portfolio, they compare unfavorably to the single-sorted portfolio based on traded volume, and their advantage with respect to the equally weighted portfolio is less statistically significant. The performance of the volume portfolio is robust to the cost of transacting both economically and statistically, even if the turnover of this strategy is more than twice than the turnover of the benchmark. Instead, the statistical significance of double-sorted portfolios disappears when considering transaction costs, and the comparison with the volume portfolio is even more unfavorable. Portfolio sorts based on liquidity uncertainty proxies only performs very poorly. Figure 25 shows the magnitude of the economic gains of the portfolios with respect to the benchmark by showing the out-of-sample cumulative returns of the various trading strategies, all net of transaction costs. The Figure shows that the best portfolio is the one based on single-sorting with trading volume.

We can summarize the findings in portfolios' exercise as follows. There is little doubt that liquidity carries a premium, which has been shown to be economically and statistically significant with our out-of-sample exercise, as well as robust to the additional cost of trading stemming from the higher turnover needed to implement the strategy. However, while the existing literature has revealed a significant premium in proxies of liquidity uncertainty, this premium seems not to be profitable out-

Table 8: This table reports the out-of-sample performance of six portfolios in the absence of transaction costs (Panel A) and in the presence of transaction costs (Panel B). The first row of each Panel is the equally weighted portfolio. Single-sorted portfolios are based on three characteristics: volume, $CVVOLM_t^{(36)}$ and $HFVOL_t^{(21)}$. Double-sorted portfolios are based on volume and either $CVVOLM_t^{(36)}$ or $HFVOL_t^{(21)}$. For each portfolio, the first column reports the average monthly turnover, and the next three columns report the out-of-sample annualized mean, standard deviation, and Sharpe ratio of returns, net of transaction costs. We test the significance of the difference of the Sharpe ratio of each portfolio with that of the equally weighted portfolio using the bootstrap technique of Ledoit and Wolf (2008). $^{*}p < 0.10;^{**}p < 0.05;^{***}p < 0.01$.

Panel A: gross of transaction costs

Portfolio	Mean	SD	SR
Equally weighted	0.106	0.196	0.541
Single-sorted portfolios:			
Volume	0.129	0.194	0.665^{***}
CVVOL	0.096	0.185	0.521
HFVOL	0.108	0.200	0.540
Double-sorted portfolios:			
Volume-CVVOL	0.119	0.188	0.630^{*}
Volume-HFVOL	0.121	0.199	0.611^{**}

Panel B: net of transaction costs

Portfolio	Turnover	Mean	SD	SR
Equally weighted	0.079	0.103	0.196	0.525
Single-sorted portfolios:				
Volume	0.165	0.122	0.194	0.631^{**}
CVVOL	0.109	0.093	0.185	0.503
HFVOL	0.502	0.093	0.200	0.462^{**}
Double-sorted portfolios:				
Volume-CVVOL	0.171	0.113	0.188	0.599
Volume-HFVOL	0.387	0.109	0.199	0.549

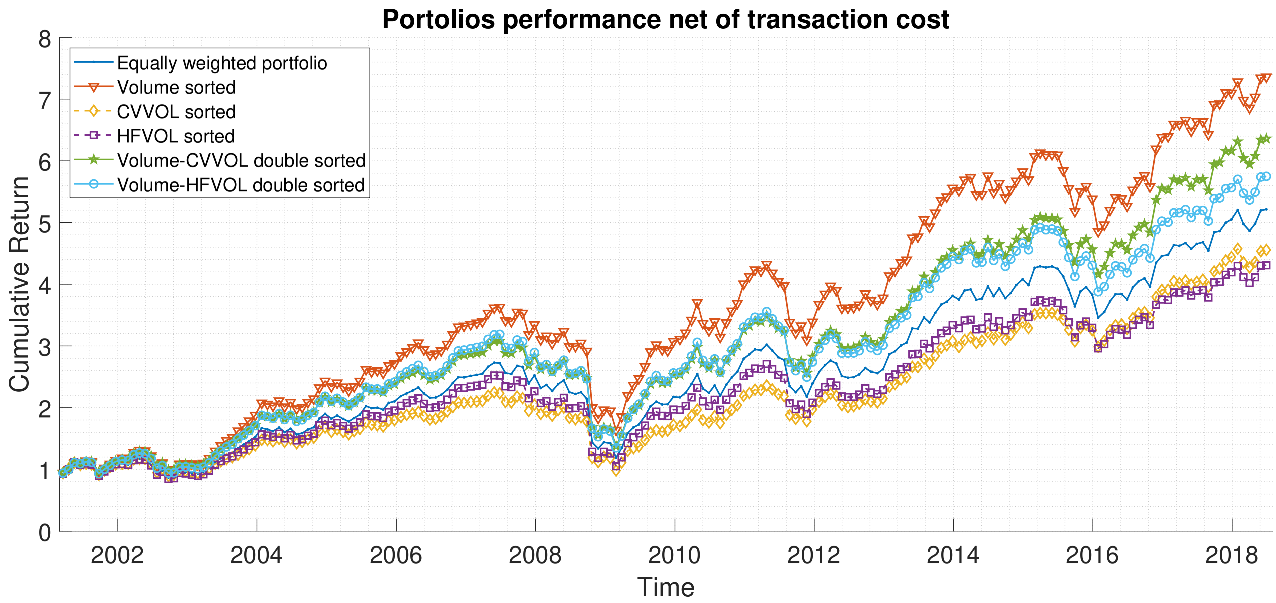


Figure 25: The Figure presents the performance of portfolios built according to different strategies plus the equally weighted portfolio (used as a benchmark). The performance is computed as the percentage monthly cumulative return of the portfolios in the sample period from 1998 to 2018 and it is net of transaction costs.

of-sample, or at least not able to gain on top of the liquidity premium. We then conclude that the data strongly support the presence of a strong liquidity premium, but do not allow to reveal a liquidity uncertainty premium.

1.9 Final discussion

This chapter tackles one of the oldest puzzles in financial economics, that is the empirical negative relationship between the firm-specific volatility of volume and the expected stock returns. The empirical works that present this (apparently puzzling) negative relation have an important commonality: they all use data sampled at a monthly level, hence a low frequency dataset that force to use a window length sufficiently large (and consequently back in the past) to compute the volatility of dollar volume in a statistically-reliable manner. We also demonstrate that by exploiting data sampled at monthly frequency, the relationship between the liquidity risk and the expected returns is negative (and particularly strong when risk adjusted excess returns are used, milder in the case of excess returns). When data are sampled on a daily basis, the results based on large aggregation windows do not differ that much from the one obtained with monthly data. On the contrary, when we employ one-minute data we make the best use of all the informativeness of the data: when the role of dependent variable is played by the excess returns the impact of volatility of volume on expected returns is always positive, regardless the aggregation window and the computation method to compute the liquidity risk; when the risk adjusted excess returns are used, the relationship is more blurred: always positive when we consider both the intra-day variation and the one across days, positive for aggregation window shorter than a month with the volatility measure that capture only the intra-day variation. To wrap up, we show that the relation between volatility of volume and expected stock returns is not robust to the way the uncertainty is computed. In particular, with the most reliable and precise measure, the one based on aggregating the coefficient of variation of high-frequency one-minute data, the relation is positive and

robust regardless the aggregation period.
The analysis presented in the chapter might be further developed; the ideas below are the results of fruitful discussions and suggestions by some experts. We may try to clean the one-minute series with a filter to eliminate noise and see if the positive high-frequency liquidity risk premium persists. In addition, we can try to split the sample between very liquid stocks and not very liquid stocks to assess whether the results continue to hold or are driven by one of the two sub-categories. Finally, we can assess whether the results we find are consistent over time: our data sample spans from 1998 to 2018, and in these years, there have been significant changes in technology that have led to changes in investors' modus operandi. The proportion of high-frequency investors is now higher than in the past. This could lead to a change in the type of liquidity risk proxies that investors observe because their needs change: whereas once most investors acted at low frequency in the financial market, now a more significant proportion operates at high frequency and therefore has an interest in observing variability at high frequency and over short horizons. Therefore, the idea would be to split the sample into sub-samples of three to five years and see whether the coefficients for the various liquidity risk proxies maintain their significance and sign or change over time.
Our findings might support the idea of a change in the investors' characteristics and an evolution in the financial market technologies.
Especially in the last few years, trading platforms have become more accessible to a broad public and favor high-frequency trading with bots and automatic trading (primarily for the big players). Heterogeneous agents with different needs, financial knowledge, and investment time horizons populate the market. At the extremes of the ample spectrum of investors' prototypes, there are the low-frequency and long-term investors and, on the other side, the high-frequency with short-term necessities. The first type of agent is unconcerned about the intraday evolution of the market because she has a long-term horizon, and she only cares about the long-run trend, waiting for the appropriate moment to act on the market. On the contrary, the diametrically opposite investor type is deeply

concerned about the intraday movements on the market. From this, their sensitivity to liquidity risk depends on different factors: the first agent cares about its evolution on a longer horizon, while the second prototype is more reckless. Technically, we can model the two sensitivities using different proxies based on the horizon and trading frequency. A long-term, low-frequency investor's proxy for liquidity uncertainty can be computed with monthly data on a long time horizon because it reflects her needs. On the contrary, a short-term, high-frequency investor's proxy for liquidity risk is based on one-minute data on a concise time horizon. The different needs of agents justify the switch in the sign. For an agent in a rush with close deadlines, the uncertainty is something bad, and that's why she requires a positive premium for the liquidity uncertainty. Differently, an agent with plenty of time to wait for the best moment to (dis)invest might see an opportunity to spot the best moment to act in liquidity volatility.
Although the premium is statistically significant, it is not profitable economically because it is too mild compared to other more considerable forces on the market.

Chapter 2

The echo chamber effect resounds on financial markets: a social media alert system for meme stocks

The short squeeze of Gamestop (GME) has revealed to the world how retail investors, pooling through social media, can severely impact financial markets. In this chapter, we devise an early warning signal to detect suspicious users' social network activity, which might affect the financial market stability. We apply our approach to the subreddit *r/WallStreetBets*, selecting both meme and non-meme stocks as case studies. The alert system is structured in two stages; the first one is based on extraordinary activity on the social network, while the second aims at identifying whether the movement seeks to coordinate the users to a bulk action. We run an event study analysis to see the reaction of the financial markets when the alert system catches social network turmoil.
This chapter is based on the work 'The echo chamber effect resounds on financial markets: a social media alert system for meme stocks' in col-

laboration with Luigi Longo and Massimo Riccaboni (see Gianstefani, Longo, and Riccaboni (2022)).

2.1 Framework and scope

Retail investors have always been considered as noise traders in the financial and market microstructure literature: their choices are believed not to be driven by the knowledge of the fundamentals of a stock or any sophisticated analysis of the market, but guided mainly by their emotional and irrational beliefs. Noise traders market impact has always been considered negligible compared to the influence of large players (such as investment banks and hedge funds). However, this picture of nonthreatening amateur investors seems outdated: the progressive diffusion of social media combined with low cost trading platforms are making investment strategies more and more widespread. The impact they have on the financial market is anything but negligible. The most striking episode is the short-squeeze that some retail investors triggered on GameStop stock (ticker symbol on NYSE: GME) by coordinating themselves on Reddit, a micro-blogging social network (Anand and Pathak, 2021).

Reddit is a website[1] composed of user-generated content and related discussions. The site's content is divided into forums, communities known as "subreddits", which deal with a specific topic. As a network of communities, Reddit's core content consists of posts (submissions) from its users. Users can comment on others' posts to continue the conversation, and they can collect positive or negative votes (score). The number of upvotes or downvotes determines the posts' visibility on the site; the more popular the content is, the higher the number of people it is displayed.

One of the most popular and active subreddit is *r/WallStreetBets*, a community focused on financial markets and stock trading. In this community, users boast very aggressive trading strategies and what they did

[1] In the top-10 of the most visited websites in the US according to various websites such as `https://www.alexa.com/topsites/countries/US` and `https://www.semrush.com/blog/most-visited-websites/`

at the end of January 2021 with GameStop is no exception. GameStop (GME) is the world's largest American retailer of video games and accessories. The market for physical video games started to plummet in 2017 due to the downloadable version of many games and services offered by the main consoles. GameStop started facing a sharp decline in sales that determined a share-price dropping in the financial market. The COVID-19 pandemic was a severe beating for the company, and its fundamentals faced another shock. The stock price decline led many institutional investors to sell the stock short. Conversely, some retail investors, considering the stock undervalued, went against the trend of the big players. In January 2021, a coordinated effort orchestrated by the community of the subreddit *r/WallStreetBets* surged the price of GME (US Securities and Exchange Commission (2021)).

Apart from what happened on the financial market, the squeeze of the price and the consequent losses faced by short-sellers, the high volatility of traded volumes and the liquidity issues, the most striking part of this episode is to comprehend how an apparently harmless group of noise traders were able to provoke such a substantial effect on a market usually dominated by the big players. Stylizing the phenomenon can be helpful to detect eventual anomalies based on indicators of coordination. This is relevant in a perspective of policy-making to prevent this kind of shock from happening or at least to tackle the harmful effects.

Besides the specific case of GameStop, which is dramatic in terms of magnitude and subsequent effects, the interest in how social media networks impact the financial price formation of meme stocks is gaining growing attention (see L. Pedersen, 2021; Costola, Iacopini, and Santagiustina, 2021). Furthermore, an ever-increasing decentralization of the financial system and the ease to access it via user-friendly online trading platforms are potential destabilizers for the financial ecosystem[2]. The fintech (r)evolution and the capillary diffusion of the online social network, which allow us to receive the up-to-date information as they happen, lay the foundations for a reinterpretation of market manipulation steer-

[2]See for example: https://www.risk.net/investing/7810026/aqr-quant-on-the-network-effects-behind-gamestop-frenzy

ing through social networks. Hence, here is the urgent need to create an alert system to pinpoint episodes of misconduct behaviors in online social network that can result in potential market manipulation. With misconduct behaviors, in this chapter we refer to an inappropriate activity at the social network level that might potentially translate into market manipulation, a deliberate attempt to interfere with the free and fair operation of the market.
Although there is no regulation legislating on the relationship between social media coordination and the financial market, the U.S. securities and exchange commission (SEC) defines as market manipulation[3] all the actions where someone artificially affects the supply or demand for a security.
Social network variables on mass coordination are valuable tools for building nowcasting systems and scheduling real-time interventions to ensure stability. We contribute to the extant literature by designing an alert system to detect potential misconducting behaviors or suspicious activity on the social network to eventually prevent the harmful coordination from creating instability in financial markets. The methodology relies on social listening and social network analysis to identify the red light days to monitor.

2.2 Literature Review

Digital and online social network revolutions are deeply affecting the functioning of financial markets. One manifestation of this revolution is the rising importance of retail traders. In the classical market microstructure models (like Glosten and Milgrom (1985) and Kyle (1985)), noise traders are considered as a residual category because of their randomness in the trades and are usually ignored in the price formation process because of their irrational impact on the market (which temporary makes the price to diverge from the fundamental value) is predominated and counterbalanced by rational agents on the market.

[3] https://www.investor.gov/introduction-investing/investing-basics/glossary/market-manipulation

One of the first models to recognize the relevance of noise traders in the financial ecosystem is the one described by De Long et al. (1990) where the authors model them as irrational traders with erroneous stochastic beliefs whose *unpredictability creates a risk in the price of an asset. As a result, prices can diverge significantly from fundamental values even without fundamental risk*. This paper was highly enlightened and forward-looking as it perceived how the agents, so far targeted as irrelevant, are effectively impacting the asset price formation in certain circumstances.
Thanks to user-friendly, low-cost online trading platforms, the widespread diffusion of social media and the easiness of accessing financial markets have significantly transformed the market dynamics. As pointed out by Zheludev, Smith, and Aste (2014), the proliferation of the internet has improved our ability to access information in real-time, and in particular, the diffusion of social media allows us to get in contact with the moods, thoughts, and opinions of a large part of the world's traders in an aggregated and real-time manner. Many are the researches aim to analyze the impact of social media on the financial markets both under a perspective of the volume of the social activity, the search engine traffic, and the prevailing sentiment of the agents. A consistent branch of the literature focuses, using various techniques, on the impact of social media activity on the financial markets Mao, Wang, and Liu (2012), Bordino et al. (2012), Ruiz et al. (2012), and Preis, Susannah Moat, and Stanley (2013) . Most of the results presented so far are based on empirical extrapolation.
In Renault (2018), the author focuses on a specific type of market manipulation: the pump-and-dump scheme. He finds that an abnormal activity on social media about a specific stock is associated with a sharp increase in volume and price on the day of the event, while the price presents a reversal over the following trading week. L. Pedersen (2021) proposes a new model that revolutionizes the vision of the so-called noise traders. He describes how investment strategies propagate in a social network and how they affect the market. Four typologies of investors are considered: besides the classic prototypes of rational short- (who tries to predict the sentiment changes among naive investors based on social network dynamics) and long-term investor (focused on the fundamental value of

assets), it portrays two new types of agents: the 'fanatics' (investors with a stubborn view that can influence many people thanks to their popularity on social media) and naive investors (agents learning and relying on investment strategies proposed on social networks). The model explains the belief formation process on the social network and how it affects the fluctuations of prices and trading volumes on the financial market. Modern social media, such as Reddit, allows envisioning (and downloading) the data generating process that leads to the coordination of the agents on the network and analyzes the underlying forces behind the event. To the extent of our knowledge, there are no works devoted to studying the network evolution and consequent market impact. Suppose social network activity does generate a force that drives the financial market dynamics. In such a case, the GameStop saga can be an exciting case study to understand the network indicators to monitor to prevent extreme phenomena like the one that happened at the end of January 2021. Hence, as recommended by L. Pedersen, 2021 in the conclusion of his paper, the availability of data from social media might open *'new research possibilities to test model's prediction using data on networks and market behaviors'*.
Dim (2021) analyzes the implications on the financial markets of *non-preofessional social media investment analysts* that publish investment strategies shaping the views and actions of many retail investors. This study highlights how the interplay of social media and retail trading poses new challenges for financial markets and regulators, which requires a deeper understanding of belief formation on social media.
We work towards defining some indicators and parameters to monitor on the social network to detect extreme situations that might affects the financial market stability. Inspired by the setting proposed in Costola, Iacopini, and Santagiustina (2021), our alert system has two consecutive red flags: if the first one, based on extraordinary activity on the social network, activates, we start monitoring the structure of the user social network.
In the network analysis, distinguishing the various roles the users can play is crucial. Many works are devoted to this categorization (see Ríos et al. (2019), Choi et al. (2015), and Thukral et al. (2018)). A special role is

played by the influencers, aka the leaders, or to use a term proposed L. Pedersen, 2021, the "fanatics". We track the users' activity within the network, catch the agents distinguishing for their ability to be central and vocal in the network by proposing relevant contributions.
Before delving into the core of our chapter, we cross-reference all the works dealing with the GameStop case, the triggering factor of this piece of literature devoted to creating an alert system to prevent the dysfunctional social network activity destabilizing the financial market. Investment recommendation (Bradley et al. (2021)), social network activity volume and tone (obtained with sentiment analysis in Long, Lucey, and Yarovaya (2021) and Umar et al. (2021)) influence the GME returns and traded volume on the financial market. Also, the Google trend researches with keywords related to the event are positively correlated with the financial GameStop performance (seeKlein (2021) and Vasileiou, Bartzou, and Tzanakis (2021)). Hasso et al. (2021) profile the agents participating in the frenzy and describe their average performance; they have proven to be relatively inexperienced and unsophisticated (Eaton et al. (2021)). Eaton et al. (2021) also infers that a large portion of agents acting on the subreddit *r/WallStreetBets* uses Robinhood as a trading app. Robinhood is a zero commission, no account minimum, and an easy-to-use interface trading app widely used among young investors. During the most turbulent period of GME frenzy, the trading app went down or malfunctioned several times, avoiding investors from acting on the financial market and loosing the best moments to trade. Many were the complains about this malfunction reported on a website *DownDetector.com*[4], an online platform that provides users with information about the status of various websites and services based upon user outage reports.
The disruptive effect would have been milder if only the financial market presented more substantial barriers to entry. Finally, in Fusari, Jarrow, and Lamichhane (2021), they report as a case study the extreme event of GME, demonstrating that they can predict the bubble using a model based on options.

[4]https://downdetector.com/

2.3 Data

Our analysis is run on four NYSE-listed stocks in the period October 2020 - June 2021 for the following stocks (NYSE ticker is reported in parenthesis): GameStop (GME), American Multi-Cinema Entertainment (AMC), Apple Inc. (AAPL) and Microsoft Corporation (MSFT). For each stock, we download daily data on price and trading volume and all the post on the social network Reddit containing the ticker. GME and AMC are two examples of meme stocks, meaning stock that gains popularity (measured in terms of mentions on social networks) among retail investors through social media. As reported in US Securities and Exchange Commission (2021), a meme stock is characterized by a confluence of all these factors: large price moves, large volume changes, large short interest, frequent mentions on social media and significant coverage in the mainstream media. At the contrary, AAPL and MSFT are two non-meme stock (as they are not characterized by factors described above) to use as controls.

2.3.1 Market data

We download the time series with daily resolution of price and traded number of share (volume) from October 1^{st} 2020 to June 30^{th} 2021. For each stock we compute the daily percentage returns:

$$R_{t,i} = \frac{P_{t,i} - P_{t-1,i}}{P_{t-1,i}} \tag{2.1}$$

where $P_{t,i}$ is the closing price of stock i on day t. The daily trading volume is denoted as $Vol_{t,i}$.

2.3.2 Reddit data

Rooting our analysis in the theoretical framework proposed by L. Pedersen, 2021, we design a model to study the interaction of agents on the social platform, evaluate their coordination effort and quantify the im-

pact of their action on the financial market. Modern social media contain an outstanding informative potential related to the users' sentiment evolution and opinion formation. If properly squeezed, the confidential information can be a forerunner for upcoming events.
Relying on the informative power of the social network and with the help of PRAW (an acronym for 'Python Reddit API Wrapper', a Python package that allows accessing Reddit's API), we downloaded all the posts containing as keyword the ticker of the stock to investigate. We limit the data download to the subreddit *r/WallStreetBets* in the period from October 2020 to June 2021. For each post we obtained the related comment forests and attributes. We downloaded the data and run the analysis for the following keywords, standing for the stock tickers: GME, AMC, AAPL, and MSFT. Precisely, we extracted every post (in Reddit jargon 'Submission') satisfying the conditions above and the related comment tree.[5]

We are interested in the emerging collective phenomena when the retail investors cooperate to determine a significant effect on the financial market. We exclude from our analysis the messages generated automatically from the bots (that in our dataset are denoted with the tag 'moderator' in the variable `distinguished`). A conversation thread can be modeled as a directed tree $T_{t,i}^k = (M_{t,i}^k, C_{t,i}^k)$. $M_{t,i}^k$ the set of nodes represented by the messages in the tree k (where the initial submission represents the root of the tree and the comments are the following-up branching) and $C_{t,i}^k$ is the set of edges, each of them connecting two messages linked by commenting activity. The direction of the edge points to the parent node to which the comment is addressed.
Figure 26 reports the number of daily posts (i.e. the number of trees) containing as keyword the ticker in the title of each subgraph and related comments (i.e. the number of nodes excluding the initial submission). Each subgraph has a double scale y-axis; on the left in red is the scale measuring the absolute value of daily submissions, on the right in blue is the corresponding measure for the number of related comments. We note that for the meme stocks (GME and AMC), the activity on the

[5]See the Appendix A.1 for more details of downloaded data.

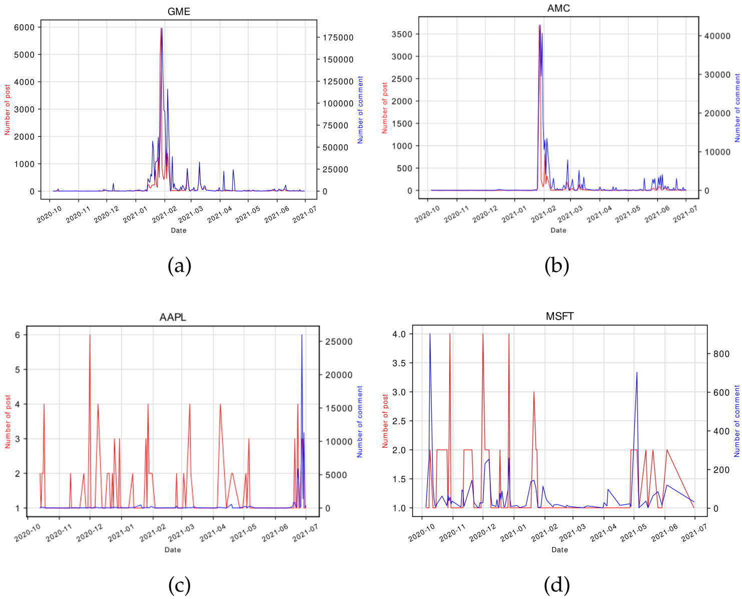


Figure 26: Time series of the daily number of post (submission) containing the word reported in the title of each subgraph (GME, AMC, AAPL, MSFT) and related comments on the subreddit *r/Wallstreetbets* in the period October 2020 - June 2021. The post y-axis scale measure is on the left in red, while the corresponding scale for the comments is on the right of each graph and in blue.

social network is massive compared to other well-known but not meme stocks (like AAPL and MSFT). Furthermore, when the social network is active on a specific topic, the correlation between creating new content and commenting activity is high. Similarly, Figure 27 shows the time series of the daily number of users who write content containing the ticker in the title of each subgraph and the daily number of users who take part in the conversation threads. In red on the left is the y-axis for the submitters; in blue on the right side is the corresponding one for commenters. The analysis is ideally in line with the one done for Figure 26.
The social network data informativeness is not limited to its extent over

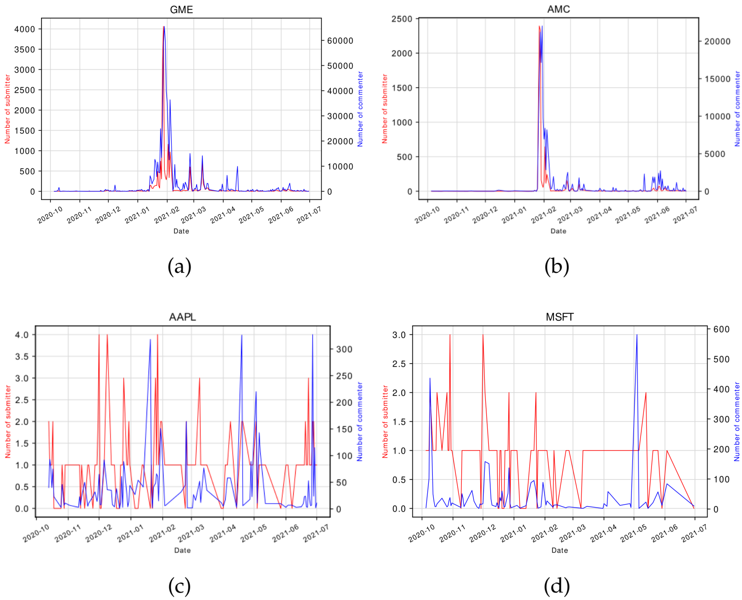


Figure 27: Time series of the daily number of submitters citing the word reported in the title of each subgraph (GME, AMC, AAPL, MSFT) and related comments by the users on the subreddit *r/Wallstreetbets* in the period October 2020 - June 2021. The submitter y-axis scale measure is on the left in red, while the corresponding scale for the commenters is on the right of each graph and in blue.

time. It can be further squeezed by analyzing the interaction among users to identify their roles within the network.

Social network analysis

We extrapolate the daily agent interplay from the tree-level raw data by reconstructing their discussing interchanges. We define a network where the nodes represent the active users on the day t, and the edges are users' interactions through the comments: the network structure is outlined by $G_{t,i} = (N_{t,i}, E_{t,i})$. The set of active users on the day t for ticker i, $N_{t,i}$, represents the graph's nodes. The commenting activity establishes the

links among them: $E_{t,i}$ is the set of directed edges from the user writing the comment to the author of the initial submission.
Following Ríos et al., 2019, we adopt a simplification in the user-network reconstruction (see Figure 28): the links always point to the author of the initial submission even if the comment is a reply to a second or higher-order comment. The ratio underlying this network reduction is due to the need to identify the users acting like hubs, gaining popularity, consensus, and driving a potentially impactful movement with their contents. In addition, the stylization is still a realistic representation of the social dynamic: when scrolling the blog, the user first reads the main submission; then, if attracted by the topic, she opens the cascade of comments. Our reduction considers all the comments as first-level comments. This structure emphasizes the submitter that becomes the central actor of a thread and, eventually, the driver of a particular message or idea. Figure 29 depicts an example of interaction among users. The network represents the interaction on Reddit on January 14st, 2021, based on submissions containing the ticker GME. There are 6,465 users (represented by the nodes) interacting on the platform throughout commenting activity (the 8,741 edges connecting them). The social network has a hub-and-spoke structure, with the colored users representing the hubs (i.e., central nodes in the network). In Appendix A.2 we report similar examples of networks for AMC, AAPL and MSFT.

2.4 Methodology

This section presents the backbone of our analysis. We describe the design and the functioning of the alert system to detect situations of misconducting behavior at social network level. Subsequently, we illustrate how we structure an event study analysis to check whether the alert system is capable of anticipating potential attempts of market manipulation.

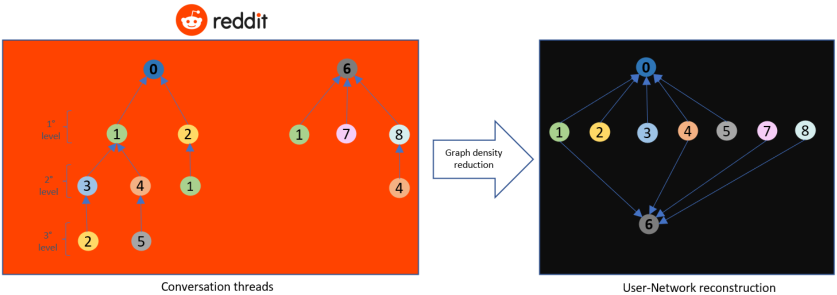


Figure 28: Example of user network reconstruction from two conversation threads. Starting from two conversation threads $T_{t,i}^k = (M_{t,i}^k, C_{t,i}^k), k = 1, 2$ stylized as a comment trees (left-hand panel), we reconstruct the interaction among users $G_{t,i} = (N_{t,i}, E_{t,i})$ by reducing the density and the level of detail. We consider all the comments to be addressed to the submission creators (in the example the users labeled with 0 and 6), as all the comments were at the first level of the tree. We do not consider how many times a user interacts with another one: we consider whether a user comments to the submitter to streamline the social network.

2.4.1 Alert system

We devise an alert system based on social-network-retrieved information. Cooperation among users can translate into a dangerous impact on financial market stability. Detecting potential misconduct behaviors and anticipating a coordinated action concocted on social media might be beneficial for the financial market well-functioning. The alert system is query-dependent, meaning that we have to instruct the system on which keyword we want to monitor. For the sake of practical usability, two consecutive stages compose the structure.

First stage

The first stage of the alert system consists of a screening of the days where the ticker-related activity on the social network is extensive compared to the previous days. We use the volume-related metrics to implement the first stage: we skim the days, identifying when ticker-related activity is extraordinary. Technically speaking, for every day and each ticker, we determine:

- The number of submissions citing the ticker in their content;
- The number of active users discussing that ticker;
- The overall[6] activity on the subreddit, both in terms of submissions and users.

And construct the following variables:

(1) Relative number of daily submission: the ratio between the submissions citing ticker and the total submissions on the subreddit, to identify the portion of the activity on that topic during the day;

(2) Absolute number of daily submission: the total number of daily submissions about that ticker, to identify the magnitude of the movement in absolute terms;

(3) Relative percentage change in the number of daily submissions: the percentage variation of number of ticker-related submissions with respect to the previous day;

(4) Relative number of daily users: the ratio between the number of users citing ticker and total users on the subreddit, to identify the portion of the community discussing that topic during the day;

(5) Absolute number of daily users: computed as the total number of daily users citing that ticker, to identify the magnitude of the user-trend in absolute terms;

(6) Relative percentage change in the number of daily users: the percentage variation of number of ticker-related users with respect to the previous day;

[6] Meaning the whole activity during the day on the subreddit *r/WallStreetBets*, not specifically related to a ticker.

The first stage of the alert system switches on when at least four variables exceed their threshold. For the relative and the absolute number of daily submissions (users), variables (1), (2), (4), (5), the threshold for the day t is the mean value of the variable over the previous ten days plus one mean absolute deviation computed over the same period. We define the mean value of the variable as

$$\bar{V}_{t,i}^{(v)} = \frac{1}{H}\sum_{h=1}^{H} V_{t-h,s}^{(v)} \tag{2.2}$$

The mean absolute deviation as

$$MAD(V_{t,i}^{(v)}) = \frac{1}{H}\sum_{h=1}^{H}\left(V_{t-h,i}^{(v)} - \bar{V}_{t,i}^{(v)}\right) \tag{2.3}$$

Hence, the threshold is defined as (method proposed in Leys et al. (2013)):

$$T_{t,i}^{(v)} = \bar{V}_{t,i}^{(v)} + MAD(V_{t,i}^{(v)}) \tag{2.4}$$

where v refers to the variables $j = 1, 2, 4, 5$, t indicates the day, i the ticker $i = GME, AMC, AAPL, MSFT$ and H is the length of the window over which we compute the threshold, in our case $H = 10$. For the relative percentage change in the number of daily submissions (users), variables (3) and (6), the threshold to overcome is 100%, hence when on the day t they double the value with respect to the previous day $t-1$.
When the social network conditions determine the activation of the alert, we approach the situation in a prudential and conservative way: we keep monitoring the stock until the average number of active indicators over the previous three days is below three. In this way, we keep controlling the situation even if it is not exceptional compared to the earlier days, but it is still turbulent.
A single indicator (or variable) switches on when it overcomes its critical threshold defined in (2.4):

$$I_{V_{t,i}^{(v)}} = \begin{cases} 1, & \text{if } V_{t,i}^{(v)} > T_{t,i}^{(v)} \\ 0, & \text{otherwise} \end{cases} \tag{2.5}$$

The minimum alert-activation condition is:

$$\sum_{v=1}^{6} I_{V_{t,i}^{(v)}} \geq 4 \tag{2.6}$$

The alert remains on until the following condition verifies:

$$\frac{1}{3}\sum_{h=1}^{3}\sum_{v=1}^{6} I_{V_{t-h,i}^{(v)}} \leq 3 \tag{2.7}$$

Step one of the alert system detects the days when the activity is exceptional, calling for further controls on the network structure.

Second stage

The second stage of the alert system activates only for those days recognized as an alert state by the first one.
For each day selected by the first step, we reconstruct the network structure to model the interaction among the agents, $G_{t,i} = (N_{t,i}, E_{t,i})$ in the manners set in the previous paragraph: the links always point to the author of the initial submission even if the comment is a reply to a second or higher-order comment.
We also implement other filters in the network modeling: creating a nowcasting alert system requires the tool to be quick and smartly devised, beyond the fact that it has to detect the bigwigs. Hence, the network reconstruction is made of only the users whose submission obtained a score above the median and with a cascade of at least ten comments.
We now move to the detection of the users acting like leaders, able to gain trustworthiness, popularity, prestige, leadership, and authority, essential features for a virtual user to lead a movement that can translate its effects on the real economy.
We rank the users according to their in-degree centrality (fraction of nodes its incoming edges are connected to) for each day in the subset of first-stage-detected days:

$$C_D(n_{t,i}^{(h)}) = \frac{\sum_j a_{t,i}^{kj}}{N_{t,i}} \tag{2.8}$$

The indicator identifies the ten authors with the highest relative incoming links: users able to attract a vast portion of the community. To test whether the detected authors are trending and critically acclaimed, for each day t, we define a set of the most critical users according to the algorithm of PageRank (Page et al., 1999). We consider the network in the window $[t-20, t-1]$ and identify the users who were standing out for the published content. According to the ranking, the first twenty users belong to the set of influencers at time t. We finally check whether some of the top in-degree centrality authors on the day t belong to the influencers set over the past twenty days[7]. If the intersection between the two sets is not empty, the second stage of the alert system switches on.
The methodology narrows down the set of agents we monitor to avoid misconducting behaviors on the online social network and prevent repercussions on the financial stability. The method aims at pin down the users who might manage suspicious movements by promoting extreme investment strategies masked by financial pieces of advice. In a perspective of macroprudential stability, a regulatory authority using the tool can quickly pinpoint the users to check to analyze their contents throughout textual analysis tools and eventually ban the profiles.

2.4.2 Analysis of abnormal returns

We finally check whether the algorithm based on social network analysis can adequately detect the financially unstable days. In order to evaluate the accuracy of our algorithm, we develop an event study analysis following MacKinlay (1997). We measure whether abnormal returns occur for the stock discussed in the Reddit community right after the alert turns on.
The abnormal returns are constructed by defining an estimation window that goes from T_0 to T_1, and an event date τ. The event window goes from T_1 to $T_2 =$. $L_1 = T_1 - T_0$ and $L_2 = T_2 - T_1$ are respectively the

[7]We do some robustness checks by changing the number of users with the highest in-degree centrality, and also modifying the parameters of the influencer set (window length and number of critical users identified by the PageRank algorithm). The results are unaffected.

length of the estimation and the event window. The abnormal returns are defined with the specification of the market model, to purge it by the market fluctuations:

$$AR_{\tau,i} = R_{\tau,i} - \hat{\alpha}_i - \hat{\beta}_i R_{\tau,m} \tag{2.9}$$

where $R_{\tau,i}$ is the return for security i at time τ, while $R_{\tau,m}$ is the market return. The parameters of the model are estimated via OLS over the estimation window $[T_0 : T_1]$. Under the null hypothesis, the abnormal returns, $AR_{\tau,i}$, are normally distributed with zero conditional mean and conditional variance $\sigma^2 (AR_{\tau,i})$ is set to be the variance of the OLS residuals $\sigma^2_{\varepsilon_t}$.

Under the null hypothesis, that event has no impact on the returns, the distribution of the abnormal returns in the event window is:

$$AR_{\tau,i} \sim N\left(0, \sigma^2 (AR_{\tau,i})\right) \tag{2.10}$$

The significance of the abnormal returns is tested via the non-parametric rank test by Corrado (1989). The abnormal returns are standardized as a ranked descending variable defined by:

$$K_{\tau,i} = \frac{\text{rank}\,(AR_{\tau,i})}{1 + M_i} \tag{2.11}$$

where M_i are the the number of observations in the sample for security i. The ranked variable is uniformly distributed in a [0,1] interval.
The variance computed across all the stock i observations is:

$$S^2_K = \frac{1}{M_i} \sum_{t=0}^{T} (K_{i,t} - 0.5)^2 \tag{2.12}$$

where 0 and T stand for the first and last date of the sample. The significance is computed as a t-rank statistic for a standard uniform distribution with expected value of 0.5:

$$t_{\text{rank}} = \left(\frac{K_{i,\tau} - 0.5}{S_K} \right) \tag{2.13}$$

These results can be used to make inference on the absolute returns for every security in the event window.

We also test the impact of the detected events on the trading volume. Specifically, we consider the abnormal trading volume computed as the volume traded on a given day divided by the average volume traded on the estimation window $[T_0 : T_1]$. For each day τ, the abnormal trading volume is defined as:

$$AVol_{\tau,i} = \frac{Vol_{\tau,i}}{\frac{1}{L_1}\sum_{t=T_0}^{T_1} Vol_{t,i}} \tag{2.14}$$

2.5 Results

This section consists of an empirical application of the methodology we design in the previous section. We present the days spotted by the alert tool and evaluate the associated abnormal returns with an event study analysis. Subsequently, we provide a regression analysis to understand the main drivers of abnormal returns and how they differ between the meme and non-meme stocks. We find that social networks-related variables significantly explain the meme stock performance, while market-related variables primarily drive non-meme stocks.

2.5.1 Event detection

In Figures 30, 31, 32, and 33 we report the daily time series of price and traded volume for the stocks under analysis, together with the days the alert system detects extraordinary activity and cooperation on the social network.

The alert tool identifies 21 suspicious movements on the social network for GME, 4 for AMC, 2 for AAPL, and 1 for MSFT.

GME is the stock facing the most incredible attention and chattering on social media. Let us consider the short squeeze that happened at the end of January 2020 as the most striking episode of social network cooperation impacting the financial economy. We appreciate that our alert tool can detect abnormal and suspecting movements on the network many days before the social coordination affects the financial market. In addition, the tool can narrow down the few users that drive the movement, simplifying the eventual inspection by a surveillance authority. In Table 9, we report all the dates and the users noticed by the alert system. The tool can trace the user acting like leaders for the movement. In Appendix A.2, we show a user network to highlight how agents interacts on the social network.

Despite the lower media frenzy compared to GME, AMC experiences considerable attention on social media. Retail investors attempt a pump-and-dump scheme during the same days of the GME frenzy, but they create actual instability on the second mid of May 2021 when they considerably purge the price. The price has risen about seven times in ten days since the warning first lit on May 19th, 2021.

Unlike GME and AMC, the financial performance of prices and volumes of non-meme-stocks (AAPL and MSFT) is entirely unaffected by social media activity. Leading users ('fanatics') on social networks are aware that stocks with such high capitalization are challenging to implement pump-and-dump schemes or drive the price away from fundamental value. For this reason, the alert is rarely triggered and the identified alerting events are insignificant compared to those of meme-stocks.

2.5.2 Analysis of the abnormal returns

The submissions from the influencers during the days of unusual activity, detected by the two steps of the alert system, are considered as events potentially triggering a response in the financial market. We analyze abnormal returns constructed with a market model to evaluate whether the submission from the influencers triggers a reaction in the financial markets. For each detected event, i.e., when the alert system turns on ($\tau = 0$), we compute the abnormal return on the estimation window [-21:-11], $L_1 = 10$ and we consider as event window the period [-10:10], $L_2 = 21$ trading days. The abnormal return is defined with a market model, specified in (2.9), with the CRSP index as a proxy for the market returns. We impose a minimum of 10 days between two events to avoid contagion on the event window and consider the events independent. If

GME

Alert date (τ)	Influencer(s)
09/11/2020	DeNovaCain
10/11/2020	Veryforestgreen
20/11/2020	Neothedreamer Imboredsoyh
21/11/2021	Ackilles
22/11/2021	Ackilles
25//11/2021	Ackilles
27/11/2020	SIR_JACK_A_LOT
26/12/2020	Uberkikz11
03/01/2021	Uberkikz11
14/01/2021	DeepFuckingValue
18/01/2021	Its-Loki
19/01/2021	Gardeeon DeepFuckingValue
21/01/2021	Unlucky-Prize
23/01/2021	Unlucky-Prize
25/01/2021	DeepFuckingValue
28/01/2021	DeepFuckingValue
29/01/2021	DeepFuckingValue
09/03/2021	dumbledoreRothIRA
15/04/2021	OPINION_IS_UNPOPULAR
16/04/2021	DeepFuckingValue
25/06/2021	Chillznday
# events = 21	

AMC

Alert date (τ)	Influencer(s)
09/11/2020	Killtrend
31/01/2021	dhiral1994 BrandinoGames
19/05/2021	Realplayer16
01/06/2021	Nobjos
# events = 4	

AAPL

Alert date (τ)	Influencer(s)
24/12/2020	Nafizzaki
22/06/2021	Tilthefatladysings
# events = 2	

MSFT

Alert date (τ)	Influencer(s)
20/05/2021	Mysterious—- EmphasisOk3036
# events = 1	

Table 9: The Table reports for each stock under analysis the events spotted by the alert system. For each detected event, it indicates the date and the and the user(s) under investigation.

an event happens on a non-trading day, we consider the next trading day as the event date.

Under these conditions, we end up with eight events for GME, four events for AMC, two events for AAPL, and one for MSFT over the sample period. It is clear that influencers driving unusual activity on social are more present for memes than non-meme stocks.

Considering that Reddit users are mostly inexperienced retail investors, we assume they primarily trade with a long position on the stock. Indeed, their primary concern is to move the price up by massively buying the stock. For this reason, we report and evaluate the significance of the event that generates the highest positive abnormal returns in the ten days after. Results are reported only for the case of meme stock. The event for the non-meme stock does not generate a clear upward trend after the submission, although some abnormal returns are significant. The events reported for GME and AMC are respectively submissions by *u/DeepFuckingValue* on January 14th, 2021 and by *u/realplayer16* on May 19th, 2021.

Table 10 and 11 represents the abnormal returns, the cumulative abnormal returns and the abnormal volumes for the stocks GME and AMC over the event window [-10:10]. The event window presents significant abnormal returns for both stocks, but the significance is generally higher for GME and more concentrated after the event. Interestingly, in both stocks, we find a clear upward trend of the abnormal returns after the event, perfectly in line with the assumptions that users are coordinated on the Reddit community to buy the stock massively. Figure 34 corroborates the claim.

In Figure 35 we report the graphical representation of the event analysis for the abnormal volume (last column of Tables 10 and 11). In line with the analysis of (cumulative) abnormal returns, the abnormal volume has a similar reaction after the event happens. It presents a significant upward trend after detecting unusual and cooperative activity on the social network.

Overall, the event analysis suggests significant abnormal returns following the alert dates for the meme-stock only, GME and AMC. The alert for non-meme stocks (AAPL and MSFT) turns on rarely and without generating abnormal returns. The alert results mark significant differences between the meme and non-meme stocks. When the alert system turns on for the meme stocks, it is related to an attempt by retail traders to coordinate an action to drive the stock market price up. The alert system on the non-meme stock detects agents on the social network discussing

Day (τ)	$AR_{GME,t}$	$CAR_{GME,t}$	$AVol_{GME,t}$
-10	0.012	0.012	0.464
-9	-0.009	0.004	0.651
-8	-0.075	-0.072	1.001
-7	0.013	-0.057	0.494
-6	0.057	-0.0009	0.551
-5	-0.011	-0.011	0.554
-4	-0.043	-0.054	0.492
-3	0.104	0.049	1.060
-2	-0.038	0.011	0.568
-1	0.534^{**}	0.545	11.521
0	0.233^{*}	0.778	7.379
+1	-0.152	0.626	3.801
+2	0.078	0.704	6.117
+3	-0.030	0.674	2.657
+4	0.082	0.756	4.571
+5	0.493^{*}	1.249	17.338
+6	0.164^{*}	1.412	16.173
+7	0.935^{**}	2.347	**20.292**
+8	**1.349****	3.697	11.961
+9	-0.443^{**}	3.253	7.463
+10	0.630^{**}	**3.883**	2.503

Table 10: Event study for GME on the event date January 14th, 2021 ($\tau = 0$). This table shows the abnormal returns, the cumulative abnormal returns and the abnormal volumes on a [-10:+10] days event window around the event day. ***, **, * represents significance at 1%, 5% and 10% for the abnormal returns only. The highest value for each column in bold.

Day (τ)	$AR_{AMC,t}$	$CAR_{AMC,t}$	$AVol_{AMC,t}$
-10	-0.012	-0.012	0.710
-9	-0.016	-0.028	1.069
-8	0.054^{*}	0.026	0.982
-7	0.020	0.045	1.067
-6	0.012	0.057	1.156
-5	-0.0005	0.057	1.267
-4	0.217^{**}	0.274	6.972
-3	0.007	0.281	5.024
-2	0.072^{*}	0.353	4.001
-1	0.0007	0.354	4.438
0	-0.098	0.256	2.277
+1	-0.007	0.248	1.569
+2	-0.031	0.217	1.329
+3	0.135^{**}	0.353	2.866
+4	0.202^{**}	0.555	5.241
+5	0.203^{**}	0.758	9.941
+6	0.365^{**}	1.123	**18.382**
+7	-0.033	1.090	10.702
+8	0.202^{**}	1.293	6.592
+9	**0.919****	**2.212**	8.603
+10	-0.215	1.996	5.868

Table 11: Event study for AMC on the event date May 19th, 2021 ($\tau = 0$). This table shows the abnormal returns, the cumulative abnormal returns and the abnormal volumes on a [-10:+10] days event window around the event day. ***, **, * represents significance at 1%, 5% and 10% for the abnormal returns only. The highest value for each column in bold.

the stock's performance without necessarily coordinating moves onto the market.

2.6 Final discussion

Social media are a powerful and impacting tool to disseminate information and stir a vast mass of people. This chapter provides empirical evidence of the echo chamber effect on the financial market from the social network, a form of market manipulation (albeit indirect). Reddit and, in general, social media are great places to share advice and manipulate poorly-informed, unsophisticated, and prone to be convinced investors.
We provide evidence that the manipulators, or to use a word a là Pedersen, the fanatics can coordinate many naive investors to provoke the desired stock price movement. The fanatics can effectively undermine the financial market stability by persuading inexperienced and easily reachable people.
We design an alert system to detect abnormal movement related to a specific stock on social media based on the extraordinary activity in terms of volume and the detection of a potential manipulator that coordinates the mass movement.
While it is far from our scope to assess whether the promotion and persuasion practice falls within the boundaries of the law, our consideration concerns market micro-structure models. In front of these episodes, the retail investors can no longer be relegated to a residual category of 'noise traders', but models should consider that many small and apparently harmless investors if aggregated and coordinated, can generate a disruptive effect on the financial market.
In the end, the entire analysis spots significant differences between the meme and non-meme stocks. The detection system finds alerts for both categories. Still, they never turn into abnormal returns for the non-meme stocks, suggesting fewer chances of suspicious trading activity or market manipulation in those cases. Moreover, the regression analysis indicates that social network-retrieved information is significant for meme stock abnormal returns only, resulting in structural differences between the

price formation of the two categories. Indeed, noisy traders may determine the price of meme stocks through social activity and potential coordination.

The analysis can be further extended and strengthen by adding more stock; we are currently working with the co-authors in this direction. Furthermore, we are refining the alert tool taking into account the severity of the signal. Finally, we would like to test whether the content analysis of the leaders and/or sentiment analysis of the users might be useful to contain the spillover effect of social media on financial stability. The difficulty here is to instruct the system to recognize slangs, irony and jokes which are very common on Reddit.

The thresholds we use for the alert system are arbitrarily chosen, but we also tried with other values, and the results remain the same. For example, we do robustness checks by changing the number of users with the highest in-degree centrality and modifying the influencer set's parameters (window length and the number of critical users identified by the PageRank algorithm). The results are unaffected. Endogenizing the fine-tuning parameters process would be ideal if we dispose of more data; we are working to extend the period, which might improve the calibration of the model without the risk of overfitting.

The decision to use a daily network of users (even if the conversation threads might last for several days) is to align the social media data and financial market data frequency. Financial data are collected daily, while social media data have a higher frequency; despite the same conversation might trigger comments for more than a day, the relevant ones occur over the next few hours. People on social are very active and react quickly, especially when dealing with hot topics.

In addition, even with an extension of data (both in period and the number of stocks), the average abnormal returns are higher already a couple of days before the event detection: the are many factors contributing to the market turbulence and disentangling the fraction due to retail investors is very ambitious.

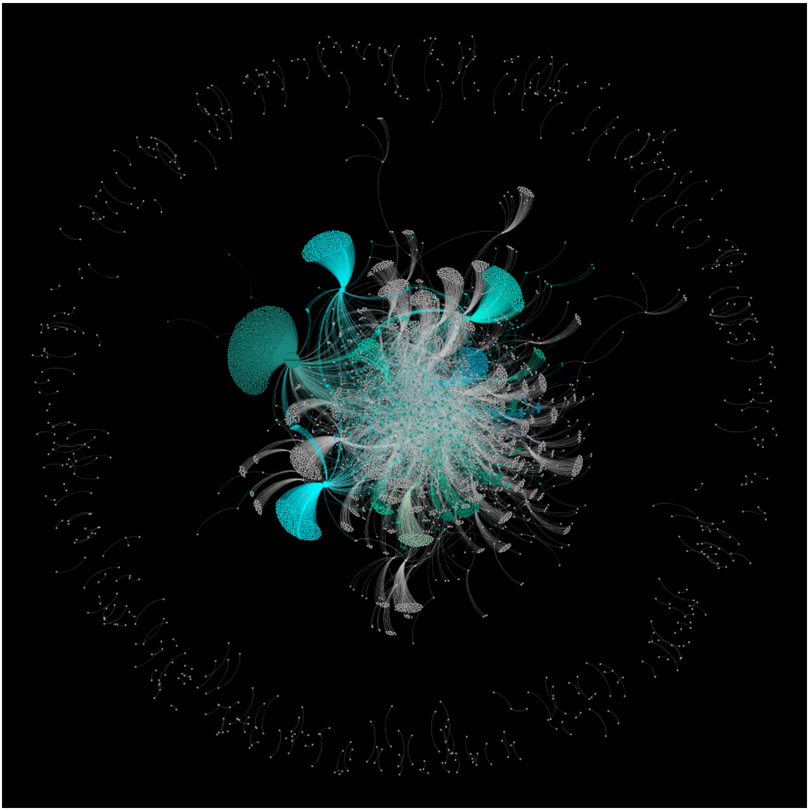


Figure 29: The Figure shows the network of users interacting on Reddit on January 14st, 2021. The network reports the interactions of users posting a submission containing the wording 'GME'. The network has 6,465 nodes and 8,741 edges. The top 10 nodes with the highest in-degree centrality are colored in blue.

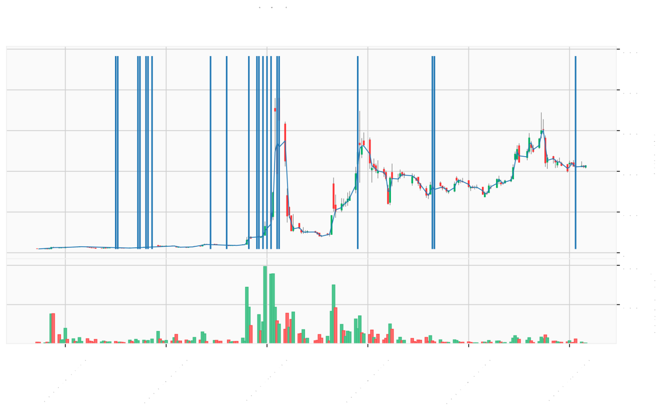


Figure 30: The figure shows financial data together with the alert days for the stock GME over the period October 2020 - June 2021. The above panel of the figure presents the time series of the daily price with a candlestick chart and the vertical blue lines are the days when the alert system turns on; in the bottom panel, the daily traded volume over the corresponding period.

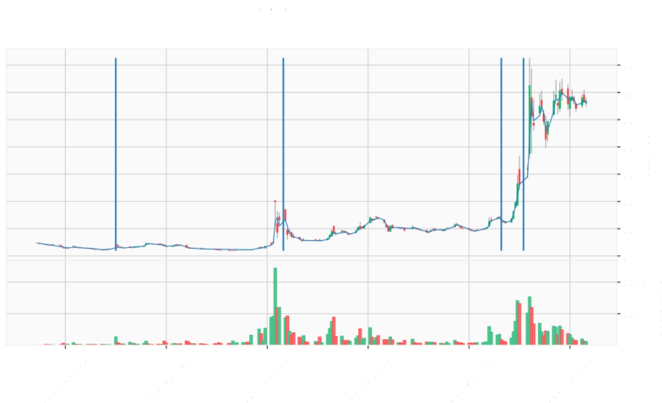
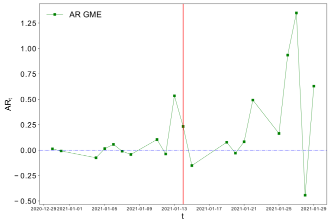


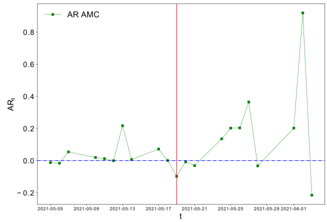
Figure 31: The figure shows financial data together with the alert days for the stock AMC over the period October 2020 - June 2021. The above panel of the figure presents the time series of the daily price with a candlestick chart and the vertical blue lines are the days when the alert system turns on; in the bottom panel, the daily traded volume over the corresponding period.

Figure 32: The figure shows financial data together with the alert days for the stock AAPL over the period October 2020 - June 2021. The above panel of the figure presents the time series of the daily price with a candlestick chart and the vertical blue lines are the days when the alert system turns on; in the bottom panel, the daily traded volume over the corresponding period.

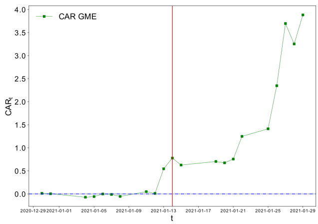
Figure 33: The figure shows financial data together with the alert days for the stock MSFT over the period October 2020 - June 2021. The above panel of the figure presents the time series of the daily price with a candlestick chart and the vertical blue lines are the days when the alert system turns on; in the bottom panel, the daily traded volume over the corresponding period.



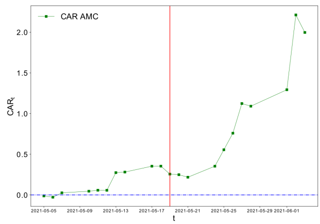
(a)



(b)

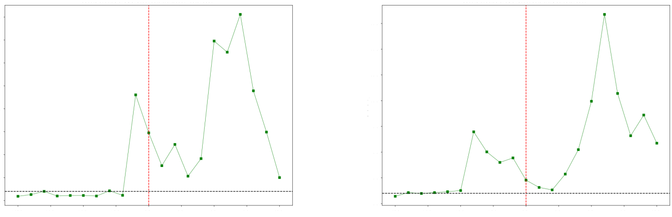


(c)



(d)

Figure 34: The figure shows the abnormal returns and the cumulative abnormal returns on a [-10:10] days event window centered in $\tau = 0$ that corresponds to the event date, marked with the red vertical line. The event date for GME is January 14th, 2021. The event date for AMC is May 19th, 2021. The horizontal dashed line indicates when the (cumulative) abnormal return equals 0. AR_t and CAR_t plot for GME and AMC single event triggering the highest (cumulative) abnormal returns in the following 10 days.



(a) GME (b) AMC

Figure 35: The figure shows the abnormal volume on a [-10:10] days event window centered in $\tau = 0$ that corresponds to the event date, marked with the red vertical dashed line. The event date for GME is January 14th, 2021. The event date for AMC is May 19th, 2021. The horizontal dashed line indicates when the abnormal volume equals 1.

Chapter 3

Interest Rate Sensitivity of Irish Bond Funds

We measure the losses bond funds would suffer when an exogenous interest rate (and credit) shocks occur in the market. We use the duration and convexity of funds' bond exposures to estimate their sensitivity to such shocks. This enables us to assess the shock-absorbing capacity concerns within the investment funds sector. We first stress test the sector by simulating a parallel shock to the yield curve; then, we calibrate the shock to a stress test scenario developed by ESMA, taking into account the domicile and credit quality of the underlying bonds. The results, split by fund categories, show that funds containing a prevalence of government securities are the most affected category. Finally, we draft a network model to stylize the systemic risk due to common asset holdings in the investment funds.
This chapter is based on the financial stability note (FSN) 'Interest Rate Sensitivity of Irish Bond Funds' in collaboration with a team of experts (Paweł Fiedor, senior economist; Kitty Moloney, head of function; Naoise Metadjer, senior economist) in the International Finance division in the Financial Stability department of Central Bank of Ireland. The views expressed in this chapter are those of the authors and do not represent the official views of the Central Bank of Ireland or the European System of

Central Banks. Any errors remain the responsibility of the authors. The work will be published soon on the Central Bank of Ireland official website (Gianstefani, Metadjer, and Moloney, 2023).

3.1 Framework and scope

Since the financial crisis in 2008, the global financial system has undergone significant structural changes, with considerable growth of non-banking financial institutions (NBFI). The worldwide NBFI financial assets grew from 42% of the total financial asset before the financial crisis to more than 50% in 2020 (Financial Stability Board (2021)). The shift from a traditional bank-centric model to an NBFI contributes to an alternative finance or investment source and diversifies risk across investors and countries (Lane and Moloney (2018)). The downside of this transformation is that it poses new challenges concerning risk sources, which need to be monitored and eventually tamed: hence the need to understand the amplification and transmission mechanism of shocks and the interaction and propagation of risks from a system-wide perspective.[1]
Over the past decade, Ireland's non-banking assets almost tripled in size. The sector increased from 1.8 trillion at the end of 2009 to approximately 4.5 trillion in the second quarter of 2019, making it the fifth largest in the world (Cima, Killeen, and Madouros (2019)). The Irish non-banking sector has developed a lot, presenting a constant and rampant growth that has slowed down in the first two quarters of 2022.[2] The composition of the sector in Ireland is dominated by investment funds and money market funds, which account for around two-thirds of total assets (Cima, Killeen, and Madouros (2019)). Within the investment fund category, there is an ample assortment of fund types, differing mainly in the invest-

[1]These topics were central in the last conference of FSB, Understanding and addressing systemic risks in non-bank financial intermediation, in June 2022 (Financial Stability Board (2022b)).

[2]As reported by the non-bank financial statistics of Central Bank of Ireland (Central Bank of Ireland (2022b)), the net asset values (NAVs) of Irish-resident funds continued to decline this year, falling to 3,7 billion in end-June 2022. The total NAV decreased by 230 billion during the second quarter (5.9%), (See Figures 3,4 in Cima, Killeen, and Madouros (2019))

ment strategy and security type: equity, bond, mixed investment funds, hedge funds, and real estate funds are the main ones. Specifically, the bond funds, investing in debt securities, are the second largest fund category in Ireland (See Figure 6 in Cima, Killeen, and Madouros (2019)) accounting for 835 billion in Q2 2019. The bond market played a central role in the recent turmoil of the economy. In particular, the measures to tame inflation combined with the UK government crisis in September 2022 generated significant losses in the UK government bonds, incentivizing investors to redeem their investments. Asset managers encountered difficulties liquidating the positions (due to liquidity mismatches) requiring the Bank of England's intervention to ensure the funds cover future payouts. Over the past few years, a series of unfortunate events demonstrated how NBFI might be a tool for propagating financial distress. The combination of the pandemic, the war, the subsequent energy crisis, and the UK turmoil made the inflation skyrocket, forcing the central banks to increase the interest rate to tame the harmful repercussions. The mechanisms underlying the financial instability propagation are pretty straightforward: an interest rate shock mainly and directly affects the bond funds. An increase in the interest rate reduces bond prices and favors the investors' redemptions, depending on the fund's sensitivity to the interest rate change. Liquidation of the fund's shares negatively impacts the fund's value generating losses. The financial instability problem manifests when the funds are hit with redemption requests and forced to sell bonds in a coordinated way, potentially leading to a reinforcing spiral. This mechanism is further amplified due to the interconnectedness of NBFI.

It is essential to assess whether the bond funds have enough shock - absorbing capacity when interest rate risk manifests. When an exogenous positive interest rate shock happens on the market, it negatively affects the bond prices favoring fund redemptions by investors, which might generate liquidity issues. On top of that, fire sales might provoke a feedback loop that induces a second round of fund redemption with an additional negative price impact. In this regard, the interaction between liquidity mismatches and leverage is at the core of financial instability.

As explained in (Carstens (2021)), liquidity mismatches are due to the on-demand convertibility of illiquid investments into cash. Many funds offer daily redemptions, meaning that investors can subscribe or redeem from a fund, at its current net asset value (NAV), on any day they choose (King and Semark (2022)). But if the fund invests in assets that take more than one day to sell at a fair price, the liquidity mismatch occurs, meaning that the fund's liabilities are more liquid than its assets.
The main contribution of this work is to assess interest rate risk on Irish-domiciled bond funds by stress testing them. We implement a forward-looking exercise to evaluate the impact of severe but plausible adverse scenarios on the resilience of bond funds. It is essential to clarify that stress tests do not predict the likelihood of such events or scenarios. Still, they are tools to assess the shock-absorbing capacity of a portion of the financial sector. Our stress test is part of the Central Bank of Ireland's risk identification and assessment framework and comes under the analytical tools and models to inform judgements around the risk environment (Central Bank of Ireland (2022a)).

3.2 Literature Review

In its last global report analyzing the critical vulnerabilities in the global financial system (International Monetary Fund (2022a)), the IMF reports that such a high level of inflation has not been seen for a long time, deteriorating the global outlook. Central banks worldwide have introduced monetary policies to normalize the situation and prevent inflationary pressure from becoming entrenched. The Euro Area is no exception to this. The European Central Bank (ECB) has been raising interest rates to tame inflation and bring it back to the target level. The monetary policy implemented by ECB to cool down inflation has some repercussions on the NBFI. An interest rate increase translates into interest rate risk for NBFI exposures in a volatile monetary policy environment.
Specifically for Ireland, the IMF positively evaluates the institutional framework and macroprudential policy for the non-banking sector (Interna-

tional Monetary Fund (2022b)). Nevertheless, considering its importance and magnitude, IMF recommends continuing to address non-bank risks and developing a more robust and sophisticated non-bank macroprudential framework. An open question is how much the new structure of the financial system is resilient to exogenous shocks. As pointed out by Goldstein, Jiang, and Ng (2017), the market participants are sensitive to the market stress level: their outflows are sensitive to bad performance more than their inflows are to good performance. Moreover, agents tend to have greater sensitivity of outflows to bad performance when they have more illiquid assets and high overall market illiquidity. Hence the importance of robust and realistic stress testing tools to test the shock-absorbing capacity of the non-bank sector.

Unlike the bank sector, the literature for non-bank stress testing tools is still nascent since market-based finance expansion started after the great financial crisis. Stress tests are forward-looking exercises that aim to evaluate the impact of severe but plausible adverse scenarios on the resilience of financial firms (Bank for International Settlements (2021)). The use of stress tests developed widely and rapidly for the bank sector to assess the banks' loss-bearing capacity after the crisis in 2008. Hence the vast literature, both theoretical and empirical, on the topic.[3]

At this stage, the NBFI stress testing tools aim to assess the non-bank sector's resilience; in the long run, the goal is to combine the two domains (bank and non-bank) in a single model to create a system-wide stress simulation.[4] We must be aware that the operativity of these tests and the good functioning strongly depend on (see Constâncio (2017)):

- The data availability;

- The ability to discern different business models in the wide cate-

[3]We suggest as references Quagliarello (2009) for an overview of the methodologies proposed in the field, Jerome Henry and Kok (2013) for a description of the analytical framework employed by the ECB for macro stress testing of banks' solvency, and Dees, Jérôme Henry, and Martin (2017) for a real application.

[4]A first step towards a more comprehensive framework on how banks and NBFIs react to an exogenous shock is proposed in Sydow et al. (2021). In Caccioli, Ferrara, and Ramadiah (2022), the authors study how fire sales affect the common (bank and non-bank) asset holdings, amplifying the losses in the UK financial system.

gory of NBFI and tailor a suitable test for each;

- Accounting for the agents' interactions (see in this regard Dees, Jérôme Henry, and Martin (2017)).

In terms of policy orientation (Baudino et al. (2018)), a stress test can be "macroprudential" (if designed to assess the system-wide resilience to financial and economic shocks) or "microprudential" (if intended to evaluate the strength of an individual institution). We will focus on the first type. Based on the Irish data availability, Fiedor and Katsoulis (2019) develop a macroprudential stress testing framework for investment funds. It injects exogenous redemption shocks into the funds, which forces them to sell assets at depressed prices, creating other negative fund returns and subsequent endogenous redemptions. Baranova, Coen, et al. (2017) test the liquidity resilience of European-domiciled corporate bond funds induced by an initial fund redemption shock. In subsequent work, Baranova, Douglas, and Silvestri (2019) simulate a stress framework to evaluate the amplification mechanism of asset price shocks in the UK corporate bond market. They note that the amplification is more significant for shocks to credit risk and risk-free interest rate than for a shock to liquidity risk.
One of the most straightforward, assumption- and calibration-free ways to simulate stress dynamics for the bond market is to parallel shift the yield curve of a fixed amount (usually 100bps). Cetorelli, Duarte, and Eisenbach (2016) clearly explains how handy this method is in analyzing the system's vulnerability and estimating the fund losses. The stress simulation is implemented in European Central Bank (2017), where they present the impact for Europen funds split by fund sector, disentangling the effect due to price and volume. European Systemic Risk Board (2022) proposes a similar framework for a sample of 200 bond funds domiciled in Ireland, Luxemburg, France, and Germany. The results, split by country, show that in front of an interest rate shock of 100bps, the average fund loss is 4% of the NAV.
Given the increasing centrality of the NBFI sector, it is of utmost importance to consider how the monetary policy is transmitted via the non-

bank sector. Carstens (2021) notes how NBFI are determinant in monetary policy transmission to the economy mainly due to liquidity mismatch and hidden leverage. The paper calls for adjustments in the regulatory framework for NBFI. Also, an expansionary monetary policy (like the one adopted in 2009-2019 and analyzed by Giuzio et al. (2021)) shifted investment funds towards riskier assets and a reduction in cash holdings. The authors noted that the shift increased the liquidity risk leaving the fund sector less resilient to large outflows. Limiting the funds' capacity to take excessive liquidity risk is an idea to prevent the funds from taking an excessive risk that might reveal destructive during hard times. Finally, the financial stability board (FSB), in a recent report (Financial Stability Board (2022a)) aimed at enhancing the NBFI resilience, recognizes the crucial centrality of the non-banking sector in transmitting economic instability due to liquidity imbalances and proposes a set of policies to address the related systemic risk.

3.3 Data

We analyse the reaction of the Irish-domiciled bond funds in front of simulated exogenous stress dynamics in the yield curve. In this Section, we introduce the Irish investment fund sector composition and some summary statistics to overview this slice of the market-based sector.
We run the stress test for the Irish-domiciled bond funds in the third quarter of 2022 (Q3 2022). The total amount outstanding is of almost 813 euro billion. We consider the following fund cohorts (see Figure 36 for the categorization and order flow we used):

- Government bond funds: funds containing at least 70% of government debt securities in their portfolios with respect to their total closing position;
- Emerging market bond funds: funds containing at least 70% of non-advanced economies exposures.[5]

[5] The list of non-advanced economies consists of countries not belonging to the ad-

- Corporate bond funds: funds containing at least 70% of corporate debt securities in their portfolios with respect to their total closing position. Corporate bond funds are further split in:
  - Investment grade corporate bond funds: corporate bond funds with at least 70% of investment grade bonds;[6]
  - High yield corporate bond funds: corporate bond funds with at least 70% of high yield bonds;[7]
  - Mixed corporate bond funds: a residual category of funds that do not reach the threshold to be included in one of the two previous categories.
- Mixed bond funds: a residual category for the bond funds whose portfolios do not have a sufficiently large (70%) portion of the government, or emerging market or corporate debt securities;

Table 12 reports the number of funds composing each category, the total net asset value (expressed in euro billion), and the relative weight of each category as a portion of the total NAV. The largest category contains the investment grade corporate bond funds, which, together with the other two corporate bond-based categories (High Yield Corporate Bond Funds and Mixed Corporate Bond Funds), weight for more than 60% of the total NAV. Even if less consistent as category, the Government Bond Funds are of extreme importance because, especially in the latest quarters, they were badly hit by the recent turmoil, generating a lot of instability.
We also have data on the securities (identified by the ISIN codes) composing the portfolio for each fund. Security-specific data allow for in-depth analysis of fund composition, its features, and sensitivity to risks.

vanced economies list compiled by IMF. The following countries are classified as advanced economies: Australia, Austria, Belgium, Canada, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hong Kong, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Macao, Malta, Netherlands, New Zealand, Norway, Portugal, Puerto Rico, San Marino, Singapore, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Taiwan Province of China, United Kingdom, United States.

[6] We consider as investment grade the bonds with a rating AAA, AA, A, BBB and equivalent nomenclatures proposed by the credit rating agencies.

[7] We consider as high yield the bonds with a rating below BBB (BBB, BB, B, CCC, CC, C, D) and equivalent nomenclatures proposed by the credit rating agencies.

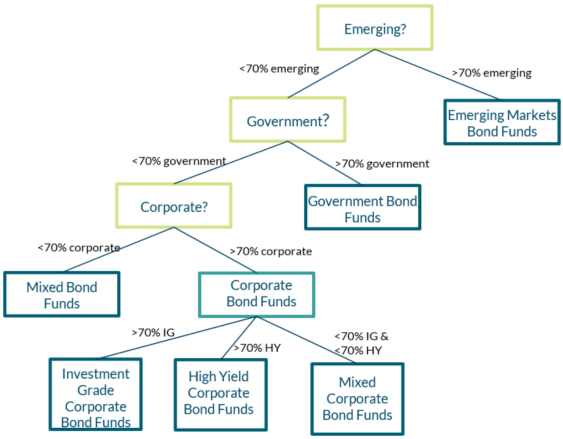


Figure 36: The Figure reports the flow we used to categorize the funds within each category. The classification process starts by selecting the funds investing in emerging market bonds (at least 70% of their portfolios). Subsequently, we identify the funds containing a prevalence (70%) of Government bonds, and we finally select the corporate bond funds by further splitting them according to their investment grade. Source: our elaboration.

Category	Q3 2022		
	Number of Funds	NAV (bln)	NAV (%)
Investment Grade Corporate Bond Funds	308	229.23	28.19
Mixed Corporate Bond Funds	155	206.21	25.36
Government Bond Funds	200	142.88	17.57
Mixed Bond Funds	150	104.39	12.84
High Yield Corporate Bond Funds	156	70.16	8.63
Emerging Markets Bond Funds	123	60.32	7.42
Total	1092	813.19	100

Table 12: The Table reports the fund cohorts we analyse, specifying the number of funds populating each category, and the total exposure, both in absolute (NAV (bln)) and relative terms (NAV (%)) with respect to the total amount of fund exposure.

We collect data on the bonds' maturity, coupon type (zero coupon bond, fixed, floating, inflation-linked, step-up), coupon rate, payment frequency, duration (a measure of how much bond prices are likely to change if interest rates move), convexity (a measure of the curvature in the relationship between bond prices and interest rates), domiciled country and investment grade (credit rating). On IHS Markit and CSDB databases we retrieve the security-specific data for 46.422 ISIN codes. Data coverage on duration is almost complete (89.4% of securities have data), while the convexity coverage is more limited (43.6% of securities have data). For the securities with missing duration and/or convexity, we calibrate two models (see below) to estimate the values. The duration model is able to produce estimates for 9.4% of total ISINs, guaranteeing an almost perfect data coverage for duration (98.7% of the securities have duration value - real or estimated). The convexity model gives a considerable contribution in broadening the data, predicting values for 30.1% of the total securities, expanding the data coverage for convexity to 73.6%.
Figure 37 reports the composition of fund categories in terms of bond coupon type composing each category. Across all the cohorts, the fixed coupon bonds (indicated by FIX) prevail. The government bond funds and the mixed bond funds contain a discrete portion of inflation-linked (IDX) bonds, reflecting the government's strategy to issue securities to protect investors from inflation. Finally, the three corporate-bonds predominated cohorts present a consistent proportion of floating (FLO) and step-up (STE) bonds. Figure 38 presents the composition of funds categories in terms of investment grade levels of the securities composing each type. The percentages are computed as the weight of each investment grade category with respect to the total NAV of each fund category. As expected, the investment - grade corporate bond funds contain more than 96% of investment-grade securities. In comparison, the high-yield corporate bond funds have in their portfolios more than 95% of speculative bonds. Funds with a high proportion of sovereign bonds (government bond funds and mixed bond funds) are dominated by prime/high-grade securities. Mixed categories and emerging market bond funds present heterogeneous investment grades, with a remarkable presence

(more than 20%) of very speculative (<=CCC) securities.
Fund categories present considerable variability in maturity (see Figure 39 and Table 13). Mixed Bond funds have, on average, the highest fund maturity, heavily exposing them to interest rate risk; the long maturity is representative of a good portion of government securities in their portfolio. As the Government Bond funds witnesses, securities issued by governments have longer expiry dates. Corporate bond funds have, on average, a lower maturity, but they present a higher variability within each cohort: the three distributions are notably right-skewed compared to other categories. The most homogeneous classes in terms of maturity are the government and the emerging market bond funds.
Modified duration distributions reflect the maturity's ones (see Figure 40 and Table 14). As we will deepen later, the fund duration is an accurate proxy for the sensitivity to interest rate risk and strongly depends on bonds' maturities. Due to the high maturities, government bond funds and mixed bond funds are notably exposed to interest rate risk. Emerging market bond funds are very homogeneous in terms of modified duration, as the high-yield corporate bond funds are. On the contrary, investment grade corporate bond funds and the residual categories (mixed corporate bond funds) present very spread-out levels of sensitivity to market fluctuations.

The bond price - yield relationship is not linear as proxied by the duration, hence to catch the curvature in the relation we need the convexity adjustment (see the focus box below for further details). In Figure 41 we report the distribution of fund's convexity split by cohorts.

3.4 Methodology

This section describes the frameworks we implement to test the absorbing capacity of IFs when a shock in the interest rate level perturbs the financial market.

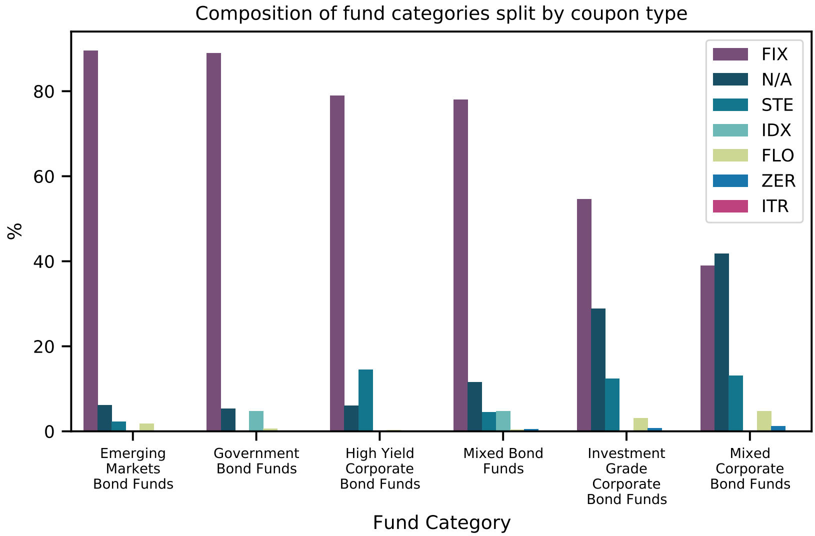


Figure 37: The Figure reports the coupon type breakdown for the fund categories included in the analysis. The abbreviations in the legend refer to the coupon types, and specifically: FIX stands for Fixed coupon bonds, FLO for Floating, IDX for inflation-linked, ITR for interest rate-linked , STE for Stepped and ZER for Zero Coupon, N/A when not specified. Source: our elaboration on HIS Markit and CSDB data.

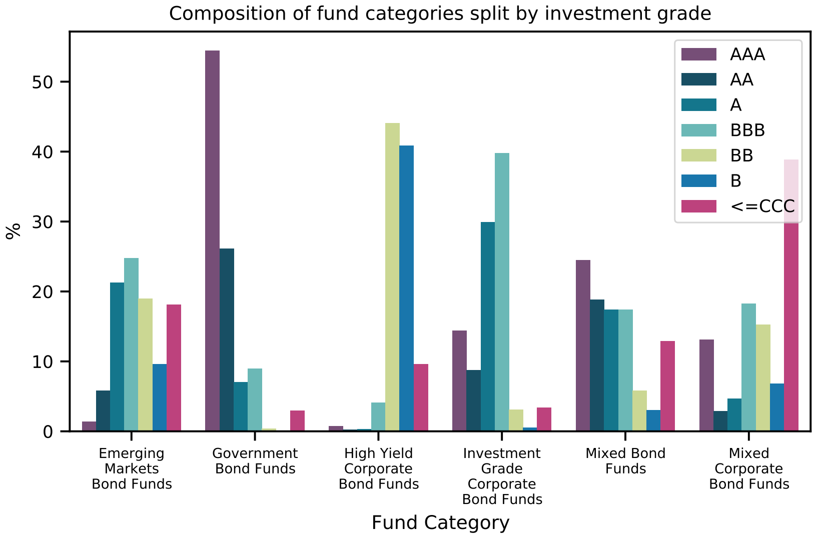


Figure 38: The Figure reports the investment grade breakdown for the fund categories included in the analysis. The percentage belonging to each investment grade level (for each category) is computed as the fraction with respect to the total NAV of the fund category it belongs to. Securities categorized as AAA, AA, A and BBB are investment grade, the remaining ones (BB, B, <=CCC) are speculative.

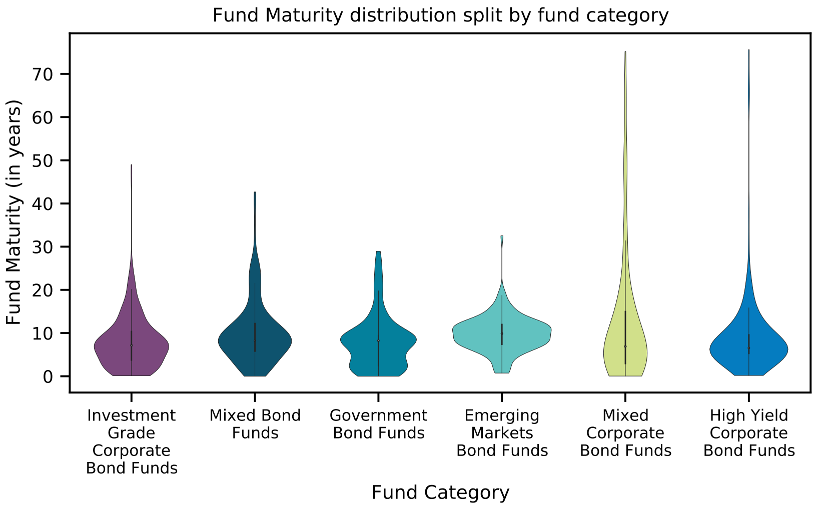


Figure 39: The Figure reports the distribution of the funds' maturities split by cohorts. Each fund maturity is computed as the market-value-weighted-average of the days to maturity of individual bonds composing the fund's portfolio. Source: our elaboration on CBI data.

Fund Category	Mean	Std. Dev.	Min	25th perc	50th perc	75th perc	Max
Mixed Bond Funds	12.91	12.26	0.04	5.85	10.04	16.01	75.19
Government Bond Funds	11.12	7.25	0.04	6.57	10.66	12.87	42.67
Emerging Markets Bond Funds	10.30	4.06	1.77	7.59	10.10	12.37	32.42
IG Corporate Bond Funds	9.03	6.17	0.07	5.11	8.72	11.25	49.81
Mixed Corporate Bond Funds	8.55	6.10	0.08	3.95	8.57	10.27	28.95
HY Corporate Bond Funds	8.00	8.28	0.00	3.50	6.42	9.72	75.63

Table 13: The Table reports the distribution of the funds' maturities split by cohorts. In addition to the mean value and the standard deviation, we report the minimum and maximum values and the 25^{th},50^{th},75^{th} percentile of the distribution. Source: our elaboration on CBI data.

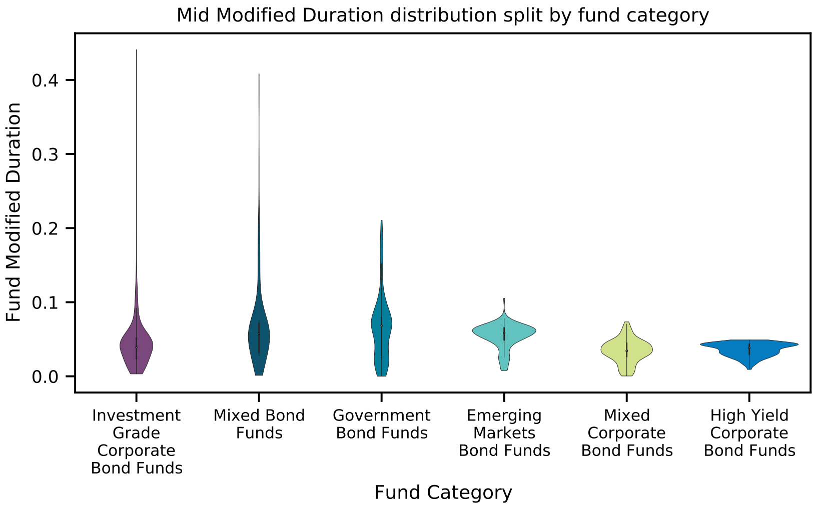


Figure 40: The Figure reports the distribution of the funds' modified duration split by cohorts. Each fund modified duration is computed as the market-value-weighted-average of the modified duration of individual bonds composing the fund's portfolio. Source: our elaboration on CBI data.

Fund Category	Mean	Std. Dev.	Min	25th perc	50th perc	75th perc	Max
Mixed Bond Funds	0.22	0.18	0.00	0.13	0.19	0.23	1.23
Government Bond Funds	0.21	0.13	0.00	0.11	0.22	0.26	0.63
Emerging Markets Bond Funds	0.18	0.04	0.07	0.16	0.18	0.20	0.32
IG Corporate Bond Funds	0.16	0.09	0.01	0.12	0.16	0.20	0.43
Mixed Corporate Bond Funds	0.16	0.12	0.00	0.10	0.14	0.18	1.32
HY Corporate Bond Funds	0.12	0.07	0.00	0.09	0.12	0.14	0.41

Table 14: The Table reports the distribution of the funds' modified duration split by cohorts. In addition to the mean value and the standard deviation, we report the minimum and maximum values and the 25^{th},50^{th},75^{th} percentile of the distribution. Source: our elaboration on CBI data.

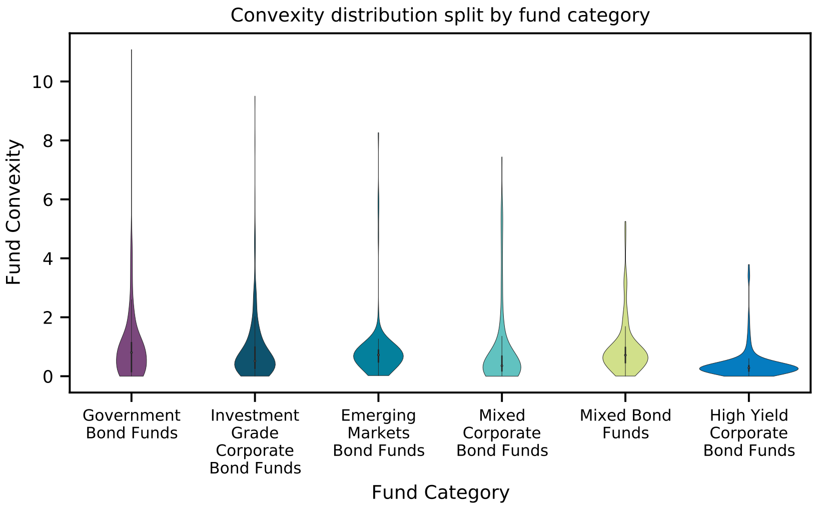


Figure 41: The Figure reports the distribution of the funds' convexity split by cohorts. Each fund convexity is computed as the market-value-weighted-average of the convexity of individual bonds composing the fund's portfolio. Source: our elaboration on CBI data.

3.4.1 Parallel shock of 300bps across the yield curve

The first stress test we execute aim to evaluate the reaction of funds to an exogenous parallel shock to the yield curve as a function of their sensitivity to the interest rate risk.

We simulate an initial shock of 300bps to the yield curve that generates a fall in bond prices. When the market is efficient, the bond price is the discounted expected value of its future cash flows. Investment funds react differently to the shock. The fund's sensitivity to interest rate risk depends mainly on the features of its securities:

- The coupon rate (ceteris paribus, a bond with a lower coupon rate experiences a greater price decrease as the interest rate increases);
- The coupon frequency (ceteris paribus, a bond with a lower coupon frequency experiences a greater price decrease as the interest rate increases);
- The face value (ceteris paribus, a bond with a lower face value experiences a greater price decrease as the interest rate increases);
- The maturity (ceteris paribus, a bond with longer maturity experiences a greater price decrease as the interest rate increases);
- Eventual callable features (not modeled for simplicity);
- Credit quality.

Bond prices usually have an inverse convex relationship with the interest rates: if the market interest rate rises, the bond price declines to remain attractive in the changing environment.

An ideal proxy to measure a bond (i) price's sensitivity to interest rate risk is the modified duration (MD_i).[8] The proxy captures the percentage change in a bond price due to a 100bps change in the interest rates. Securities with more sensitivity to interest rate risk (i.e., with higher duration) have more significant price fluctuation than those with less sensitivity.

A fund's modified duration (MD_f) is the weighted average duration of

[8] For the basic principles of bond pricing and risk exposures, we refer to Fabozzi (1999).

the bonds composing its portfolio. Security-specific data help in this purpose: the modified duration is computed by weighted-averaging the securities' duration and using their relative contribution to the fund's NAV as weights.

$$MD_f = \sum_{i=1}^{I^f} MD_i \cdot w_i^f \tag{3.1}$$

where w_i^f is the weight of security i within the fund f and is computed as the relative contribution of the security closing position (CP_i^f) to the fund NAV (NAV^f):[9]

$$w_i^f = \frac{CP_i^f}{NAV^f} \tag{3.2}$$

The impact of a variation in the interest rate (Δy_{IR}, in decimal format) translates into a fund loss (L_f) that depends on the sensitivity of the fund to interest rate risk (MD_f):

$$L_f = -(-\Delta y_{IR} \cdot MD_f \cdot CP_f) \tag{3.3}$$

where CP_f is the total closing position of fund f. The fund's modified duration acts as an amplifier/multiplier for the impact of the shock on the fund price reduction.[10] The only limit of duration is that it assumes a linear relationship between the interest rate and the bond price when, in reality, it is convex.

We complement the model with convexity to account for non-linear relationships and better proxy the degree of curvature. Convexity measures (C_f) how much the duration change when the interest rate moves, hence the second-order derivative of the price-yield function. Hence, the loss formula adjusted for the convexity part becomes:

$$L_f = -(-\Delta y_{IR} \cdot MD_f + (\Delta y_{IR})^2/2 \cdot C_f) \cdot CP_f \tag{3.4}$$

[9] $\sum_{i=1}^{I^f} w_i^f = 1$.

[10] The quantity inside the round brackets computes the price variation of the bond fund when the shock manifests; given that the bond price moves inversely to the interest rate, adding the minus is a convention to identify a price movement with opposite sign respect to interest rate shock. But since we compute the fund loss (L_f), which implicitly evokes a value reduction for bonds, we prefix the brackets with a minus. A positive value of L_f implies a reduction in the fund value.

3.4.2 ESMA interest rate and credit shock stress test

In the spirit of a more realistic exercise, we calibrate the shock to trace the level of risk a security supports as a function of its domicile and creditworthiness. Every year, the European Security and Market Authority (ESMA) calibrates some adverse scenarios (European Securities and Markets Authority, 2022) split by country or region depending on the prevailing sources of risk affecting regional stability (see the Appendix). The calibration process considers the geopolitical, economic, monetary, and fiscal trends and the effect of undergoing macroeconomic policies. The scenarios try to match the systemic risk result from a simulation set that aims to identify the main sources of financial instability. The scenarios apply to the MMF stress test. Still, they are a good starting point for calibrating the shocks for the IF sector because they embed the country-specific uncertainty caused by the Russian-Ukrainian conflict, the pandemic, and all the cascade effects due to the current turmoil. In addition, ESMA proposes various credit shocks depending on the security creditworthiness. In the model we assume the credit shock to propagate throughout the same mechanism of interest rate one, with duration and convexity acting as shock multiplier. Based on the ESMA scenarios (European Securities and Markets Authority, 2022) for country-specific interest rate shock and investment grade specific credit shock, we implement a stress test security-domicile- and security-investment-grade-specific, with a uniform shift in the yield curve depending on the region where the security is domiciled and its investment grade. The fund perturbation now depends on the country where the security is issued to catch the risk sources better and on its creditworthiness:

$$L_f = -(-\sum_{i=1}^{I^f} \Delta y_{IR+CR,i} * MD_i * CP_i^f + \sum_{i=1}^{I^f} (\Delta y_{IR+CR,i})^2/2 * C_i * CP_i^f) \tag{3.5}$$

where $\Delta y_{IR+CR,i}$ is the country-specific shock plus the investment-grade-specific credit shock in decimal format.

3.4.3 Modified Duration and Convexity estimation models

Duration is the weighted average of discounted bond's future cash flows: it tells the investors how many years it will take the bond cash flow to repay the initial investment. Usually, to compare the sensitivity to the interest rate risk of bonds (or funds), we use the modified duration, which measures the change in bond price when the interest rate fluctuates. Modified duration well approximates the variation in the bond price when a slight variation in the interest rate level occurs, and it depends on the bond's features. Hence, an ideal model to estimate modified duration when we do not have real value is based on the available variables related to future cash flows. In particular, our dataset offers excellent and almost complete coverage of data related to maturity and coupon specifics. It comes naturally to calibrate a model able to predict modified duration based on those features:

$$MD_i = \alpha + \beta_1 DTM_i + \beta_2 CR_i + \beta_3 CT_i + \varepsilon_i \tag{3.6}$$

where the regression model is calibrated by exploiting data for which we have full coverage on the maturity expressed in days (DTM_i), the type of coupon (CT_i is a categorical variable that considers the zero-coupon bond as the baseline, and the fixed ($CT_{FIX,i}$), floating ($CT_{FLO,i}$), step-up ($CT_{STE,i}$), inflation-linked coupon ($CT_{IDX,i}$) as alternatives) and the coupon rate (CR_i). The estimated coefficients, reported in Table 15 on the LHS, are statistically significant, with the security's maturity positively impacting the modified duration, as the investment recovery positively depends on the residual maturity. Since a zero-coupon bond repays the initial investment when it expires, its duration equals maturity. All the other coupon-type variables significantly reduce the duration value thanks to their periodic payments. As for the coupon rate, the higher it is, the lower the bond sensitivity to market happenings.
The second-order adjustment of convexity catches the curvature that characterizes the price-yield curve. Based on our data availability, we calibrate the convexity model to estimate convexity as follows (estimates in

	MD	C
const	15.10***	-58.22***
	(0.15)	(0.75)
DTM	0.0007***	0.0017***
	(0.00)	(0.00)
CR	-0.24***	-0.54***
	(0.01)	(0.15)
MD		21.88***
		(0.09)
CT_FIX	-10.46***	
	(0.15)	
CT_FLO	-11.19***	
	(0.21)	
CT_IDX	-4.59***	
	(0.53)	
CT_STE	-17.45***	
	(0.16)	
R-squared	0.47	0.76

Table 15: The Table reports the estimated coefficients of the models on modified duration (Eq.(3.6)) and convexity (Eq.(3.7)). Standard errors in parenthesis. The modified duration model is calibrated on 43.673 observations, the convexity one on 25.825.

Table 15 on the RHS):

$$C_i = \alpha + \beta_1 DTM_i + \beta_2 CR_i + \beta_3 MD_i + \varepsilon_i \tag{3.7}$$

Again, maturity positively impacts the convexity estimation, and the coupon-type coefficients are negative. The modified duration positively impacts the convexity value: bonds with high duration tend to have high convexity.

3.5 Results

In this section we report the results of the stress test analysis, specifying the losses suffered by the fund cohorts in absolute and relative terms.

3.5.1 300 bps interest rate shock

This section reports the results of the stress test model we described in section 3.4.1. We simulate a 300bps parallel shock of the yield curve and test how the fund cohorts react to the perturbation. Each fund reacts to the shock with a specific sensitivity, proxied by the duration and convexity of the securities composing its portfolio.

In absolute terms (panel above of Figure 42), the government bond funds are the most affected cohort, followed by the investment grade corporate bond funds and the mixed bond funds. This outcome is driven by the high sensitivities to interest rate risk of the securities contained in their portfolio and the significant mass of these categories. Mixed bond funds and government bond funds are the categories with more elevated portions of public bonds in their strategies. Usually, public securities tend to have longer maturities meaning a higher susceptibility to shift of the yield curve. Investment grade bond funds are instead the largest category. In relative terms (panel below of Figure 42), the mixed bond funds are the category experiencing the biggest loss as percentage to the NAV (slightly above 16%), followed by government bond funds and emerging markets bond funds. On the contrary, corporate bond funds have a majority of private corporate bonds in their portfolios, generally characterized by a shorter maturity and lower sensitivity to interest-rate-related events. When a 300bps shock happens on the yield curve, they are only a glancing hit, with relative losses below 10%.

3.5.2 Interest rate and credit ESMA shock

This section reports the results of the stress test model we described in section 3.4.2. For each security we simulate a shock composed by a country-specific interest rate shock and a investment-grade-dependent credit shock. Hence, each security is perturbed by a specific shock depending on its domicile and creditworthiness. The ratio of this setup is to disrupt the funds and catch the underlying risks the securities are exposed to. The shocks figures, reported in Appendix, are calibrated to capture the risks due geopolitical and macroeconomic events (interest

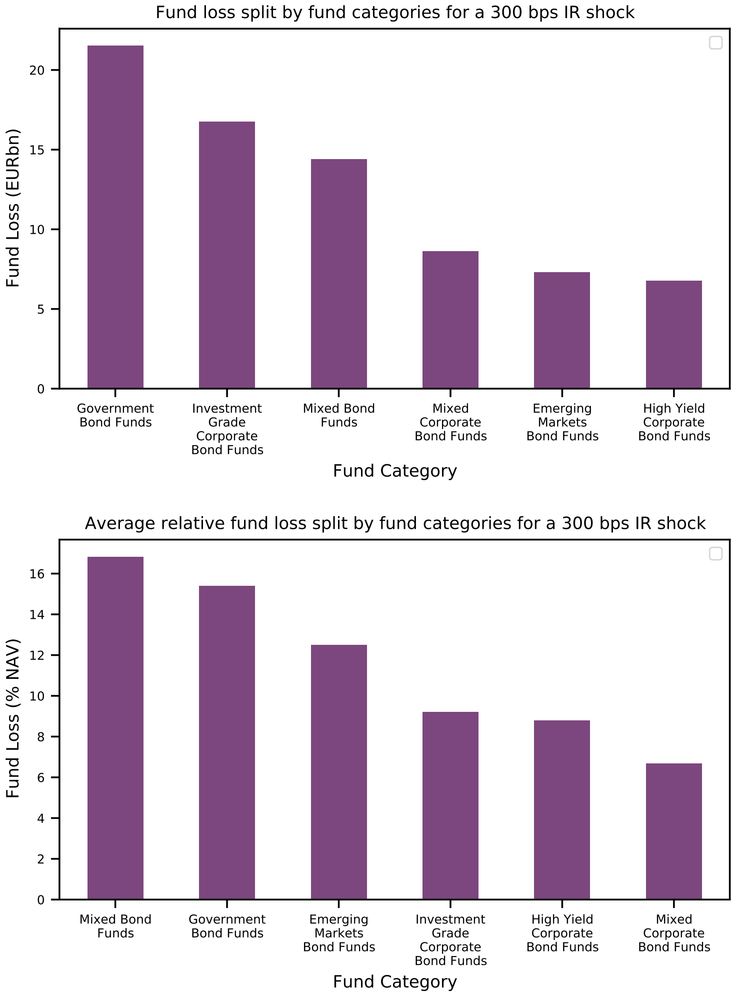
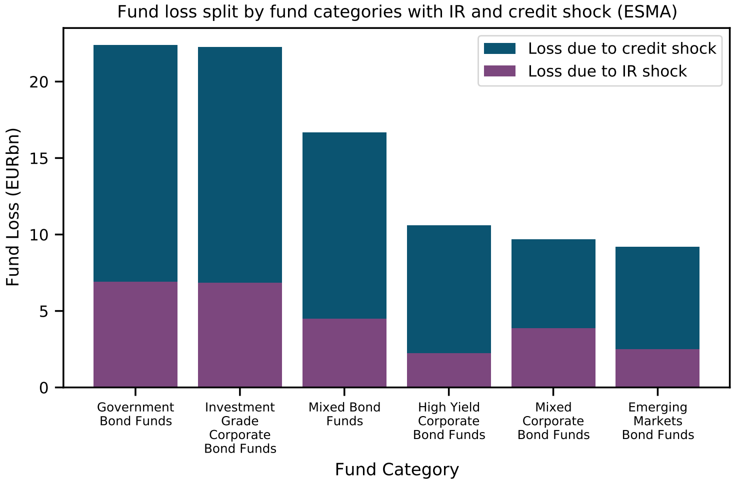


Figure 42: The Figure reports the losses due to a 300 bps parallel shock of the yield curve in absolute (panel above) and relative (panel below) terms split by fund cohorts. The absolute loss for each fund category is computed as the sum of fund losses belonging to that category. The loss in relative terms is given by the absolute loss divided by the total NAV of the category. Source: our elaboration on CBI data.

rate shock) and the risk related to the security investment grade quality (credit shock).
The results in absolute terms, reported in the panel above of Figure 43, are perfectly in line with the previous stress testing exercise: the government bond funds, investment grade bond funds and the mixed bond funds are the most hit categories. More informative is the panel below of Figure 43 that reports the losses in relative terms disentangling the interest rate and credit effect. It is easy to point out how the credit shocks hit the most for the fund cohorts containing an higher portion of speculative bonds. High yield corporate bond funds have in their portfolios more than 95% of speculative bonds determining consistent credit shocks. The same happens for mixed bond funds and emerging markets bond funds, that contain many low quality rating bonds. Credit shock has milder effects for government bond funds and investment grade corporate bond funds, thanks to the high rating bonds they have in their strategies. This type of analysis is of extreme importance because it highlights the potential pockets of vulnerability in the various fund categories.

3.6 Towards to a systemic approach: a network analysis for the Irish-resident bond funds

The previous stress test exercises embrace a microprudential approach, considering each fund in isolation without considering the interconnection among each other. One limitation of this approach (ESMA, 2019) is not taking into account their "collective action" or "potential mitigating effects." The shared reaction to the same shocks generates spillovers and second-round effects. An exogenous market shock (an interest rate rise) generally forces asset managers to liquidate some shares of the funds to satisfy the investors' liquidity requirements. Typically, the reaction is the same across funds, and the sales generate an asset price depression. The knock-on effect provokes a consequent NAV reduction which determines additional fire sales (Fiedor and Katsoulis, 2019).
The shift to a macroprudential approach is of tremendous importance in considering the interdependences among the institutions since the sys-



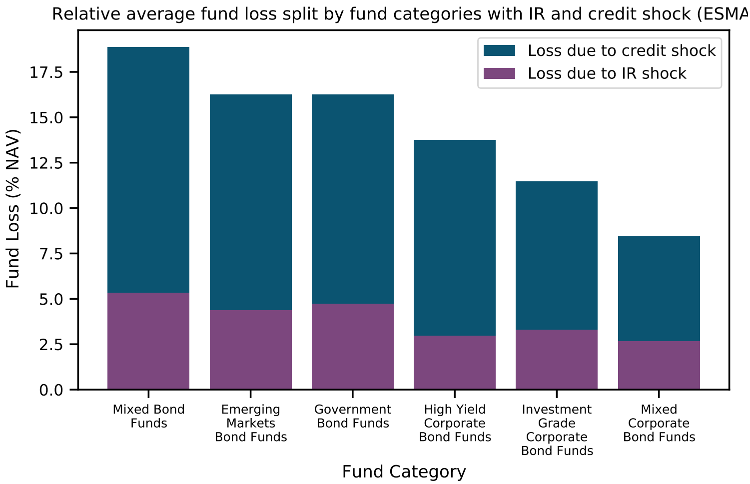


Figure 43: The Figure reports the losses due to a customized parallel shock of the yield curve in absolute (panel above) and relative (panel below) terms split by fund cohorts. The security-specific shock is the sum of an interest rate shock that depends on the security country or region and a credit shock that depends on the security creditworthiness. The total loss for each fund category is computed as the sum of fund losses belonging to that category, considering both the loss due to the interest rate shock and the credit one. The relative loss is given by the total loss divided by the total NAV of the category. Source: our elaboration on CBI data.

tem is not just the sum of its parts.
The paragraph reviews the models for the interconnectedness intended as the linkages between the funds regardless of the market conditions (ESMA, 2020) and proposes a specification as an initial case study, which will be further developed in the future. The ultimate goal is to realistically model the interdependencies among the funds and integrate the network effects within the stress test framework.
The literature presents many frameworks on interbank networks and has developed many novel stress test exercises to assess the systemic risk within the banking system (Battiston et al., 2016, Jerome Henry and Kok, 2013, Caccioli, Barucca, and Kobayashi, 2018, Cont and Schaanning, 2017, Vodenska et al., 2021, Bricco and Xu, 2019).
On the contrary, the analysis of interconnectedness and contagion in the non-banking financial sector is still nascent, and due to the rapid expansions in the last years, academics and researchers in the (inter)national institutions are working to close this gap. They can amplify financial distress, henceforth the necessity to create macroprudential tools to monitor them (Bank for International Settlements, 2021). ESMA Report on Trends, Risks, and Vulnerabilities (ESMA, 2019) analyses the European fund industry's interconnectedness, revealing that funds exposed to less liquid asset classes are more likely to be affected by shocks. Fiedor and Katsoulis, 2019 proposes an architecture for implementing a macroprudential stress test on Irish-resident investment funds, monitoring the systemic risk.
Based on those works and inspired by the methodology proposed in the paper by Sakakibara et al., 2015 we construct an inter-fund exposure network. Investment funds present overlapping portfolios, meaning their portfolios share some assets. To analyze the interdependences among funds, we consider the vulnerability originated by their common holdings. In our schematization, funds are the network nodes, and the edges depend on the common holding two funds share. Let $F = \{f_1, \ldots, f_N\}$ be the set of investment funds and $S = \{s_1, \ldots, s_M\}$ the set of securities in which the funds invest. Each fund f_i invests in a pool of securities $S(f_i)$. We model it as a directed weighted graph where the arrow points

to the fund with shared assets, and the weight ($w_{i,j}$) is computed as the fraction (with respect to NAV) of portfolio fund "i" shared with fund "j." Technically, the weight is computed as $w_{i,j} = \frac{E_{i,[S(f_i) \cap S(f_j)]}}{NAV_i}$, where the $E_{i,[S(f_i) \cap S(f_j)]}$ is the total exposure that fund "i" shares with fund "j". The level of interconnectedness among the funds gives an idea of the systemic risk severity. The more the funds are connected (i.e., they invest in the same securities), the higher the spillovers due to systemic risk in case of an exogenous shock in the market.
In the figures below, we report the network graph of the Irish-resident bond funds analyzed before. The node's color depends on the fund category it belongs to, and precisely:

- Investment Grade Corporate Bond Fund is in pink;
- Government Bond Fund is in light green;
- Mixed Corporate Bond Fund is in light blue;
- High Yield Corporate Bond Fund is in orange;
- Mixed Bond Fund is in dark green;
- Emerging Market Bond Fund is in red.

The node's size depends on the fund's number of shared securities. In the graphical representations, we construct indirect graphs (for visual clarity, we avoid building direct graphs), and the edge between two nodes depends on the presence of common holdings; the edge's color intensity is a function of the number of securities two funds have in common. On top of that, we used an algorithm to compute modularity (a measure of the strength of the division of a network into clusters).[11]
In Figures 44 and 45, we report the network graph for each fund category. The level of interconnection within each category is consistent: the probability of funds belonging to the same category having similar investment strategies (hence investing in similar securities) is high, and the graphs confirm our expectations.

Figure 46 shows a more comprehensive network graph with the in-

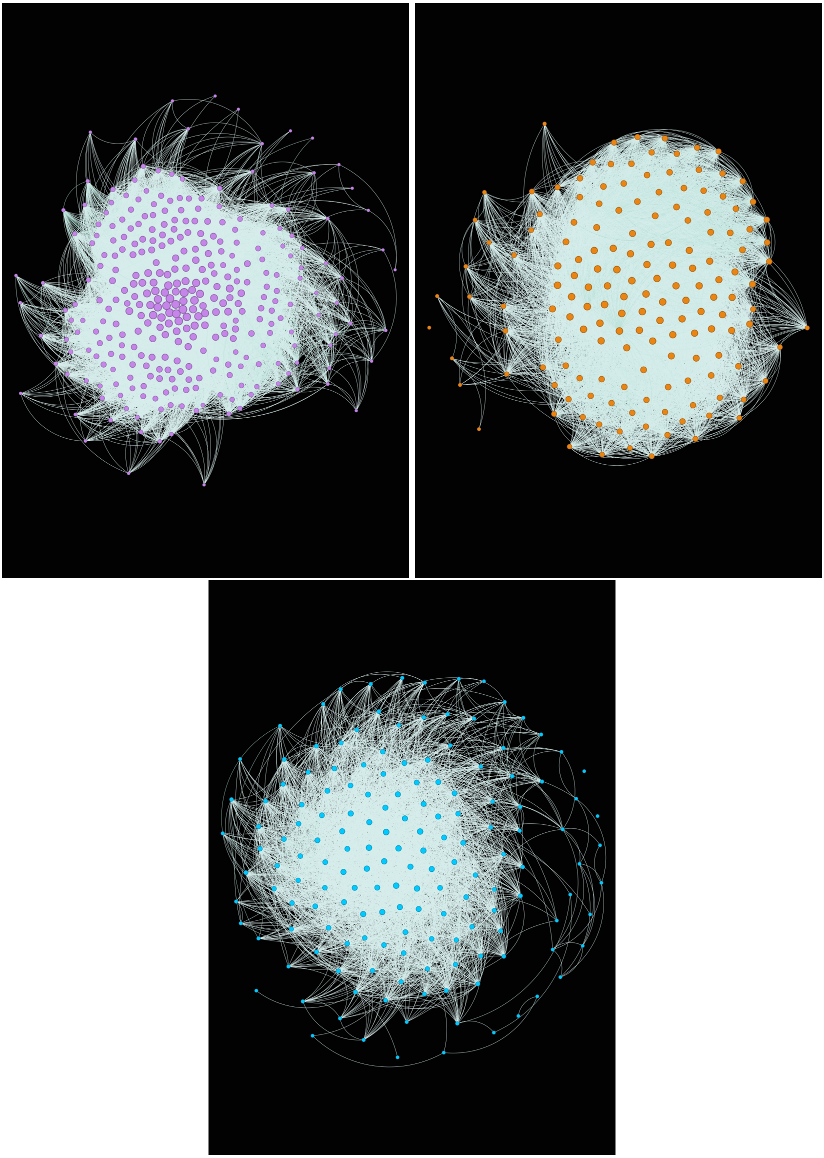


Figure 44: Network graphs of the funds belonging to the following fund categories: Investment Grade Corporate Bond Funds (top-left in pink), High Yield Corporate Bond Funds (top-right in orange), Mixed Corporate Bond Funds (bottom in light blue).

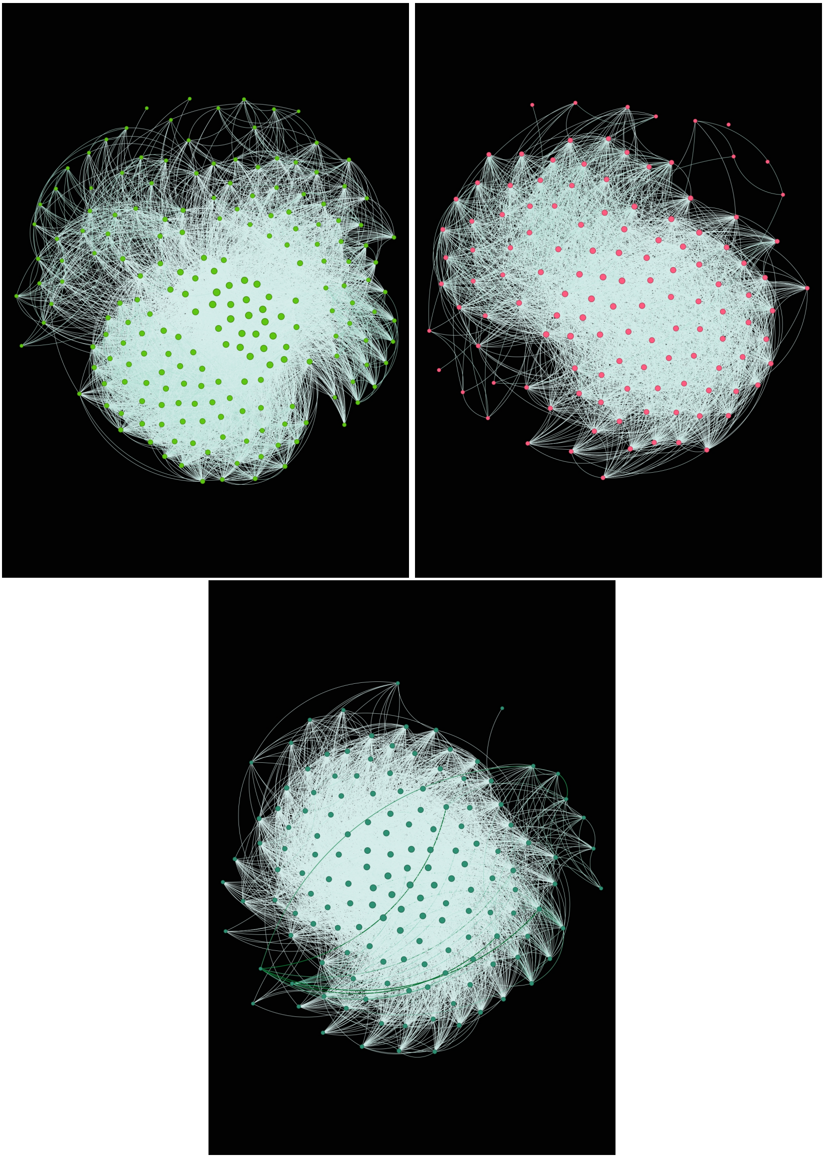


Figure 45: Network graphs of the funds belonging to the following fund categories: Government Bond Funds (top-left in light green), Emerging Market Bond Funds (bottom-left in red), Mixed Bond Funds (bottom-right in dark green).

terconnections of all the funds in the analysis.[12] Funds are clustered for their fund category, and the fund categories present a significant level of interconnectedness among each other. This result is less obvious than the previous one: if an interconnection within the same category was expected because of the investment similarities, an interdependence cross-categories is less straightforward.[13] However, the network demonstrates how the world of funds is strongly interconnected regardless of their fund category. To further corroborate this evidence, we cluster the funds based on the strength of their bonds, regardless of their fund category. Based on modularity, the algorithm detects four more densely connected communities within the network. The peculiarity of each community is that it contains a set of funds strongly interconnected because of the similarity in their investment strategies. As we can appreciate from the network graph in Figure 47, each modularity class not only contains a predominance of funds belonging to the same fund category. All the modularity classes contain more than one fund category, and the four classes are interconnected. The cluster at the top-left of Figure 47 presents a prevalence of high-yield corporate bond funds, mixed corporate bond funds, and investment-grade bond funds. The community at the top-right is populated mainly by investment-grade and mixed corporate bond funds. While government bond funds together with the mixed bond funds populate the group on the bottom-right, the set on the bottom-left shows very heterogeneous categories (emerging market bond funds, mixed corporate bond funds, mixed bond funds and government bond funds). The degree of interconnection among the four modularity classes is not negligible, meaning that funds belonging to different categories and modularity classes are still connected with others belonging to different clusters. At this initial stage of the analysis, whether this network structure might amplify or mitigate the distress

[11] The algorithm to compute modularity is the one proposed in Blondel et al. (2008).

[12] The edges between nodes are colored in different shades depending on the bond's strength.

[13] Note that we define the fund categories based on $> 70\%$ in a particular category, and some categories are mixed. So we expect some overlapping portfolios across categories. Fund categories are not an inherent characteristic of the funds, as by construction we allow them to have certain portion of mixed assets.

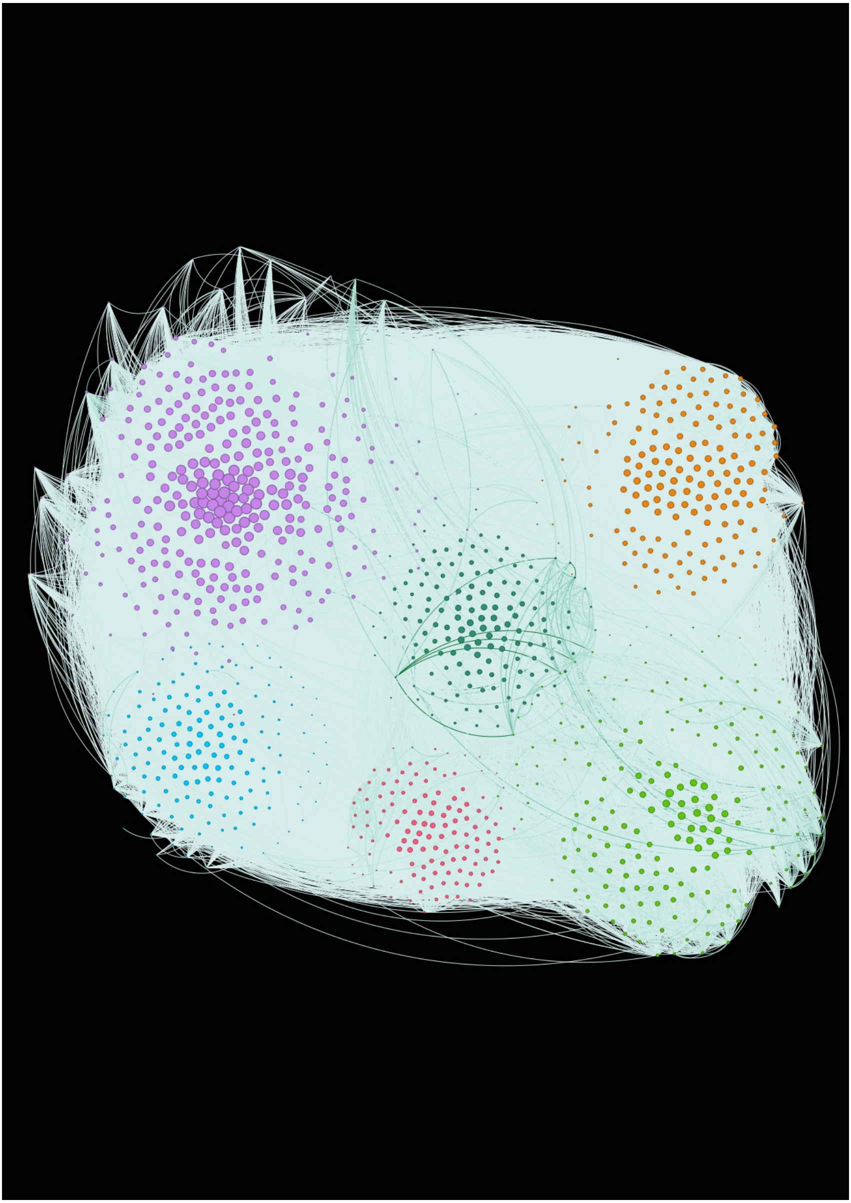


Figure 46: Network graph representing the Irish-domiciled bond funds cluster by their belonging to a fund category.

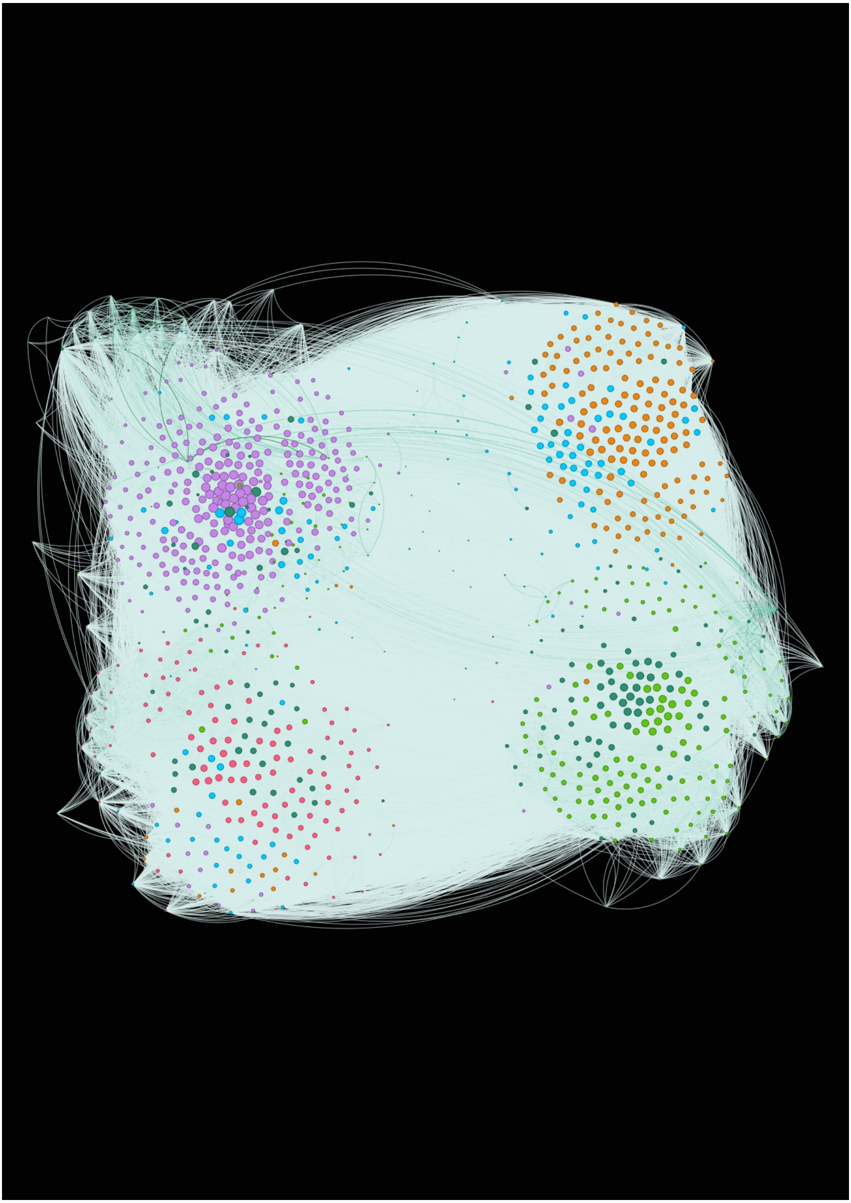


Figure 47: Network graph representing the Irish-domiciled bond funds cluster using modularity measure to detect communities more densely connected.

transmission is unattainable. Diversification is a cornerstone of portfolio theory, but it is hard to say whether a capillary interconnection might favor systemic risk propagation. What we can grasp here is that funds are strongly interconnected with each other, sometimes regardless of their fund category.
This investigation is just a starting point for a relevant - still too few unexplored - topic in the analysis of systemic risk in the world of non-banking institutions.

3.7 Final discussion and policy consideration

This chapter describes the main findings of stress testing the investment funds domiciled in Ireland. We test the resilience of NBFI by generating exogenous shocks to the interest rate (and the credit spread): we note that the most affected categories are the ones containing long-maturity assets, which determines a greater sensitivity to the market shocks.
The work is contextualized in a broader set of analytical tool to study the liquidity imbalances generated by the redemptions due to market shocks. The liquidity mismatch might cause financial instability, and here is the importance of better knowing the bond market structure to set some policy proposals to address systemic risk. The regulators are working to implement new tools to control and contain the threat: the main focus is to reduce and avoid excessive spikes in demand for liquidity by increasing liquidity resilience during the period of stress.
The implemented analysis presents some limitations we may try to overcome in future research. We consider only short-term repricing of bonds; long-term dynamics are not included (e.g., fund managers may rebalance portfolios). In addition, we do not reprice the derivatives used to cover interest rate changes; the availability of the EMIR dataset will be crucial to consider derivatives and fine-tuning the analysis. Also, the stress test framework treats the investment funds independently, without considering the systemic risk due to common holdings. A recent stream of literature is working to tackle this issue by estimating the magnitude of spillover effects on the market. In paragraph 3.6 we propose a network

analysis in the spirit to shed some light on how funds' interconnectedness might influence the systemic risk. Many researchers are currently working to quantify the systemic risk generated by a certain fund or fund category. Once granular data will have a full coverage (meaning that all EU-resident funds disclose their portfolio composition) it would be ideal to develop an indicator which summarize the level of interconnectedness of a fund with the rest of the system in such a way that we can classify the funds according to the systemic risk they might generate and detect where the pockets of vulnerability are. Ideally, the indicator should considers many aspects of the fund: apart from the interlinkages with other funds, the countries where the money flows is relevant, to account for geopolitical risk; finally, also the sector in which the fund invests should be taken into account because some sectors will be more oppressed than other due to the technological change and policies (for example the policies towards a greener economy).

Finally, we do not consider the so-called "amplification mechanism". We apply a specific shock, ignoring other potential happenings on the balance sheet. The interest rate shock impacts both price and volume. The price effect is a reduction in funds' NAVs resulting from the lower valuation of their portfolios; the volume effect is a reduction in funds' NAVs resulting from investor outflows. The second round effect is related to the price impact: the sale of assets reduces their price, further decreasing the fund's net asset value. Given the relationship between investors' flows and returns, the negative performance will trigger additional outflows and require other sales by fund managers.

Conclusion

The present thesis aims to analyze some of the risk sources that impact the financial stability of the economic system. Based on a solid quantitative analysis and statistical models, this work contributes to analyze well-known sources of risk and proposes some tools to detect new ones. It also highlights the importance of keeping up with the latest innovations and changes to anticipate and model new risk sources.
While the first chapter examines an old and famous puzzle in the financial literature, the liquidity risk premium in the stock market, it roots its innovation in the usage of high-frequency data that, if adequately squeezed, can contribute to discover some new evidences and patterns. The study shed light on the sign of liquidity risk premiums, showing how the granularity of data might be impactful in drawing any conclusions.
The second chapter presents a new risk source, a product of innovation: social media and the technological evolution democratize the access to the financial markets of a mass of apparently harmful investors, who demonstrated that, if adequately coordinated, can shake the market stability. The study presents its maximum form of innovation in the data: social media data are still embryonic; they require new technologies and models to become informative and a considerable computational power to be processed. However, they appear very prominent in designing alert tools and nowcasting/forecasting models based on human activity.
Finally, the third chapter proposes an analysis of the current interest rate risk in the Irish-resident bond fund market. Due to the current high infla-

tion level, central banks worldwide are increasing interest rates to cool it down. Of course, this monetary policy presents some drawbacks by negatively impacting the bond funds. Hence the importance to assess how the funds react to the monetary policies by stress testing them.

Apart from the conclusions and potential extensions around every chapter composing this thesis, it is necessary to propose some final recommendations. History has demonstrated that the same problem can occur many times, and even if we are perfectly conscious of it, it might present slightly different appearances. It is of utmost importance to continue exploring and studying the dynamics of the financial market: being informed and proposing new models and techniques is the only way for our research to be up-to-date and functional. In addition, the data are the gold of the new era: data with an incredible informative potential surround us, and if correctly squeezed and combined with the classical sources, they can reveal their usefulness. Finally, the research must not be tight housing. On the contrary, it has to be strictly interconnected with the regulation and the monetary policy and be the main instrument to make conscious choices.

Appendix A

Appendix to Chapter 2

A.1 Data download

For each tree, we have as many rows in the data frame as the number of comments, and each row contains the following information:

- `title`: the textual content of the initial submission;
- `body`: the textual content of the comment;
- `name`: the id of the author of the comment (each id is prefixed by `'t1_'` to specify the author made a comment activity);
- `parent_id`: the author of the parent comment to which the comment in question refers to (the `parent_id` can be prefixed by `'t1_'` if the author of the comment replies to an other comment or it can be prefixed by `'t3_'` if the author of the comment replies to the top-level post, i.e. the submission);
- `author_name`: the name of the author who post the initial submission;
- `depth`: the level of the comment tree at which the comment in question is located (if a tree is composed by the initial submission only, the depth is 0; if the comment refers to the initial submission

the depth is 1; if the comment refers to a comment in the first level, the depth is 2; and so on);

- `score`: the number of up-votes minus the number of down-votes obtained by the comment;
- `score_submission`: the number of up-votes minus the number of down-votes obtained by the initial submission;
- `upvote_ratio`: the percentage of upvotes on the total votes received by a submission;
- `time_submission`: date and time at which the initial submission is published;
- `time_comment`: date and time at which the comment is published;
- `num_comment`: number of comments below the initial submission that compose the tree;
- `flair`: a tag used to categorize the post according to the topic it deals with; they are subreddit specific and in the case of the subreddit *r/WallStreetBets* the users can select among the following ones:
  - YOLO, the acronym for 'You Only Live Once, it can be used for posts presenting extremely aggressive investment strategies with a consistent value at risk;
  - DD, the acronym for Due Diligence, must be applied to post presenting research on a specific company/sector/trade. It should include sources and citations;
  - Discussion, an idea or article that you would like to talk about;
  - Gain, to show off a solid winning trade;
  - Loss, to show off a brutal, crushing loss;
  - Earnings Thread, weekly earnings discussion thread or a specific earnings event;
  - Daily Discussion, daily catch-all thread for discussions;

  - Mods, only for official business.

- `distinguished`: if a bot automatically performs the commenting activity, the variable reports the wording 'Moderator', none otherwise, when a non-automatic user adds the comment.

Note that in the case of submission without comments below, the data frame has a single row with empty values for the comments-related variables.

A.2 The Social Network of Reddit users

In Figure 48, we present an example of network user graph on January 31st, 2021, where the main submissions contain the ticker AMC. There are 15.534 users (represented by the nodes) interacting among them on the platform throughout commenting activity (the 21.032 edges connecting them).

The directed edges point from the comment's author to the author of the main submission. The peripheral nodes in the graph are the less connected users; in the central part of the network, the most connected and central users: the colored ones are the users with the highest in-degree centrality.

Figure 49 presents the network graph for AAPL a and MSFT on two alert dates, respectively June 22nd, 2021 for AAPL and on May 20th, 2021 for MSFT. Compared to the meme-stock case, the non-meme-stocks present a feeble activity on the social network even in extraordinary occasions that determine the alert activation.

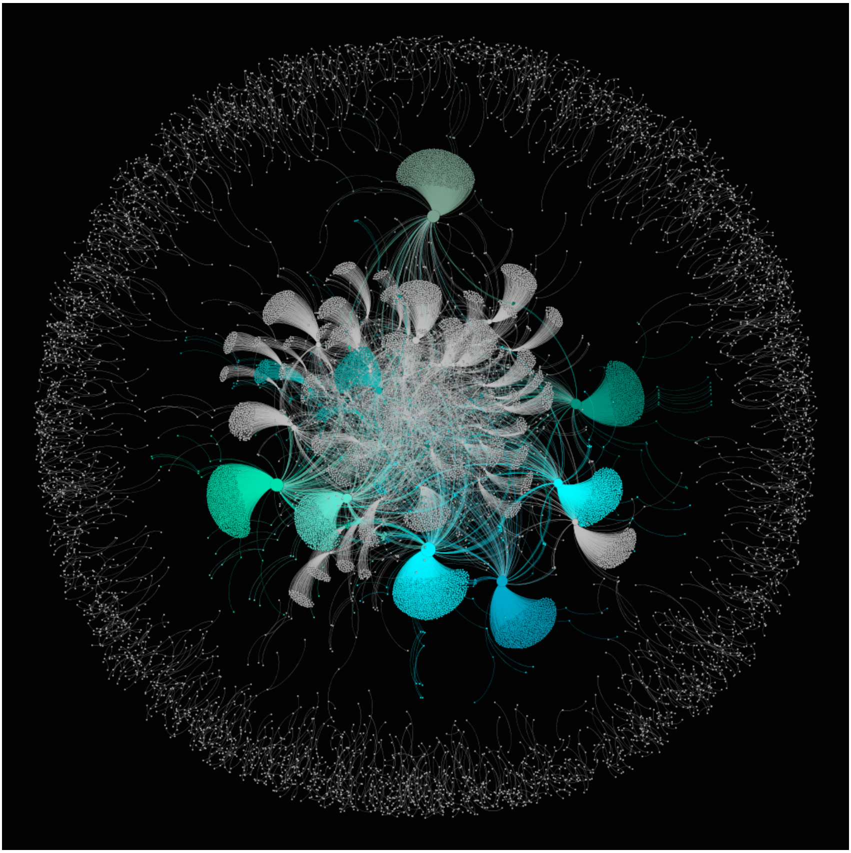
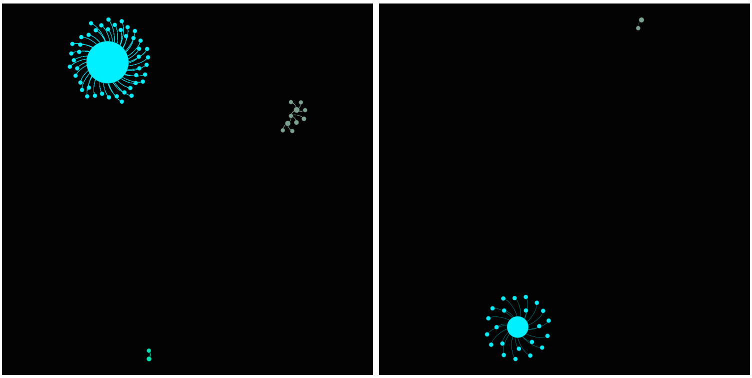


Figure 48: The Figure shows the network of users interacting on Reddit on January 31st, 2021. The network reports the interactions of users posting a submission containing the wording 'AMC'. The graph contains 15.534 nodes and 21.032 edges. The colored part of the network is the nodes involved in the net of the 10 users with the highest in-degree centrality.



(a) (b)

Figure 49: The Figure shows the network of users interacting on Reddit on June 22nd, 2021 for AAPL and on May 20th, 2021 for MSFT. The network shows the interactions of users posting a submission containing the wording 'AAPL' and 'MSFT' respectively.

Appendix B

Appendix to Chapter 3

B.1 Interest rate and credit shock

Every year the European Systemic Risk Board (ESRB) in collaboration with the European Supervisory Authorities calibrate some adverse scenarios for the Money Market Fund (MMF) sector European Securities and Markets Authority, 2022. The scenarios need to assess the resilience of financial institutions to adverse market conditions.
Table 17 reports the shocks in basis points to interest rate they calibrate to stress test the MMF. Table 16 reports the shocks in basis points to credit spread they calibrate to stress test the MMF. We use the shocks in our stress test in the spirit of customizing the stress level we impose to a security to catch the interest rate and credit risk as a function of its exposure to market risk and creditworthiness, respectively.

Investment Grade	**Credit Shock**
AAA	176
AA	201
A	231
BBB	302
BB	349
B	428
<=CCC	507

Table 16: Shocks to credit yields in basis point proposed by the ESRB to conduct stress tests to assess the resilience of financial institutions. The shocks depend on the security creditworthiness and they are split by the investment grade categories considered by the ESRB.

Geographic Area	Country	IR Shock
EU	Euro area	68
EU	Bulgaria	68
EU	Croatia	68
EU	Czech Republic	68
EU	Denmark	68
EU	Hungary	85
EU	Poland	70
EU	Romania	88
EU	Sweden	49
Rest of Europe	United Kingdom	85
Rest of Europe	Iceland	
Rest of Europe	Norway	52
Rest of Europe	Russia	
Rest of Europe	Switzerland	45
Rest of Europe	Turkey	191
North America	Canada	74
North America	United States	97
Australia and Pacific	Australia	75
Australia and Pacific	New Zealand	
South and Central America	Brazil	
South and Central America	Chile	117
South and Central America	Colombia	115
South and Central America	Mexico	114
Asia	China	26
Asia	Hong Kong	92
Asia	India	
Asia	Japan	19
Asia	Korea	
Asia	Malaysia	60
Asia	Singapore	78
Asia	Thailand	
Africa	South Africa	68
EU	All countries	70
Other Adv Econ	All countries	62
Emerging Mkts	All countries	96

Table 17: Shocks to interest rate swap rates in basis point proposed by the ESRB to conduct stress tests to assess the resilience of financial institutions. The shocks depend on the security country or geographic area to catch the various interest rate risk levels.

Bibliography

Abdi, F. and A. Ranaldo (2017). "A Simple Estimation of Bid-Ask Spreads from Daily Close, High, and Low Prices". In: *The Review of Financial Studies* 30, pp. 4437–4480.

Acharya, V.V. and L.H. Pedersen (2005). "Asset pricing with liquidity risk". In: *Journal of Financial Economics* 77.2, pp. 375–410. ISSN: 0304-405X.

Aït-Sahalia, Y and J Jacod (2015). *High Frequency Financial Econometrics*. Princeton University Press.

Akbas, Ferhat, Will J Armstrong, and Ralitsa Petkova (2011). "Idiosyncratic volatility of liquidity and expected stock returns". Working Paper.

Amihud, Y. (2002a). "Illiquidity and stock returns: cross-section and time-series effects". In: *Journal of Financial Markets* 5, pp. 31–56.

— (2002b). "Illiquidity and stock returns: cross-section and time-series effects". In: *Journal of Financial Markets* 5.1, pp. 31–56. ISSN: 1386-4181.

Amihud, Y. and H. Mendelson (1986). "Asset Pricing and the Bid-Ask Spread". In: *Journal of Financial Economics* 17, pp. 223–249.

Anand, Abhinav and Jalaj Pathak (2021). "The role of Reddit in the GameStop short squeeze". In: *Economics Letters*, p. 110249.

Andreou, Panayiotis C et al. (2018). "Differences in options investors? expectations and the cross-section of stock returns". In: *Journal of Banking & Finance* 94, pp. 315–336.

Bali, Turan G et al. (2014). "Liquidity shocks and stock market reactions". In: *The Review of Financial Studies* 27.5, pp. 1434–1485.

Bank for International Settlements (November 2021). *Stress testing – Executive summary*. Financial Stability Institute.

Baranova, Yuliya, Jamie Coen, et al. (July 2017). *Simulating stress across the financial system: the resilience of corporate bond markets and the role of investment funds*. Bank of England, Financial Stability Paper No. 42.

Baranova, Yuliya, Graeme Douglas, and Laura Silvestri (June 2019). *Simulating stress in the UK corporate bond market: investor behaviour and asset fire-sales*. Bank of England, Staff Working Paper No. 803.

Barinov, Alexander (2015). "Why does higher variability of trading activity predict lower expected returns?" In: *Journal of Banking & Finance* 58, pp. 457–470.

Battiston, Stefano et al. (2016). "Leveraging the network: a stress-test framework based on DebtRank". In:

Baudino, Patrizia et al. (2018). *Stress-testing banks – a comparative analysis*. FSI Insights on policy implementation No 12.

Blondel, Vincent D et al. (2008). "Fast unfolding of communities in large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 10, P1000.

Bordino, I. et al. (2012). "Web Search Queries Can Predict Stock Market Volumes". In: *PloS One* 7, e40014.

Bradley, D. et al. (2021). "Place your bets? The market consequences of investment advice on Reddit's Wallstreetbets". In: -.

Brandt, M., P. Santa-Clara, and R. Valkanov (2009). "Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns". In: *The Review of Financial Studies* 22, pp. 3411–3447.

Brennan, M., T. Chordia, and A. Subrahmanyam (1998). "Alternative factor specifications, security characteristics,and the cross-section of expected stock returns". In: *Journal of Financial Economics* 49, pp. 345–373.

Brennan, M. and A. Subrahmanyam (1996). "Market microstructure and asset pricing: On the compensation for illiquidity in stock returns". In: *Journal of Financial Economics* 41, pp. 441–464.

Brennan, Michael J, Tarun Chordia, and Avanidhar Subrahmanyam (1998). "Alternative factor specifications, security characteristics, and the cross-section of expected stock returns". In: *Journal of Financial Economics* 49.3, pp. 345–373.

Bricco, Jana and TengTeng Xu (October 2019). *Interconnectedness and Contagion Analysis: A Practical Framework*. IMF Working Paper.

Caccioli, Fabio, Paolo Barucca, and Teruyoshi Kobayashi (2018). "Network models of financial systemic risk: a review". In: *Journal of Computational Social Science* 1, pp. 81–114.

Caccioli, Fabio, Gerardo Ferrara, and Amanah Ramadiah (February 2022). *Modelling fire sale contagion across banks and non-banks*. Bank of England, Staff Working Paper No. 878.

Carstens, Agustín (December 2021). *Non-bank financial sector: systemic regulation needed*. BIS Quarterly Review.

Central Bank of Ireland (December 2022a). *Financial Stability Review*. Progress report.

— (September 2022b). *Investment Fund Statistics – Q2 2022*. Central Bank of Ireland Report.

Cetorelli, Nicola, Fernando M. Duarte, and Thomas M. Eisenbach (February 2016). *Are Asset Managers Vulnerable to FireSales?* Federal Reserve Bank of New York, Liquidity Series III.

Choi, D. et al. (2015). "Characterizing Conversation Patterns in Reddit: From the Perspectives of Content Properties and User Participation Behaviors". In: *COSN '15: Proceedings of the 2015 ACM on Conference on Online Social Networks*, pp. 233–243. DOI: 10.1145/2817946.2817959.

Chordia, T., A. Subrahmanyam, and V. Anshuman (2001). "Trading activity and expected stock returns". In: *Journal of Financial Economics* 59, pp. 3–32.

Chordia, Tarun, Avanidhar Subrahmanyam, and V Ravi Anshuman (2001). "Trading activity and expected stock returns". In: *Journal of Financial Economics* 59.1, pp. 3–32.

Chung, Kee H and Chairat Chuwonganant (2018). "Market volatility and stock returns: The role of liquidity providers". In: *Journal of Financial Markets* 37, pp. 17–34.

Cima, Simone, Neill Killeen, and Vasileios Madouros (December 2019). *Mapping Market-Based Finance in Ireland*. Central Bank of Ireland, Financial Stability Note.

Constâncio, Vítor (Sept. 2017). *Macroprudential stress-tests and tools for the non-bank sector*. ESRB Annual Conference.

Cont, Rama and Eric Schaanning (February 2017). *Fire sales, indirect contagion and systemic stress testing*. Norges Bank Working Paper.

Corrado, Charles J (1989). "A nonparametric test for abnormal security-price performance in event studies". In: *Journal of financial economics* 23.2, pp. 385–395.

Costola, Michele, Matteo Iacopini, and Carlo RMA Santagiustina (2021). "On the "mementum" of Meme Stocks". In: *Economics Letters* 207, p. 110021.

D'Avolio, G. (2002). "The market for borrowing stock". In: *Journal of Financial Economics* 66, pp. 271–306.

Datar, V., N. Naik, and R. Radcliffe (1998). "Liquidity and stock returns: An alternative test". In: *Journal of Financial Markets* 1, pp. 203–219.

De Long, J. et al. (1990). "Noise Trader Risk in Financial Markets". In: *Journal of Political Economy* 98, pp. 703–738.

Dees, Stéphane, Jérôme Henry, and Reiner Martin (February 2017). *STAMPE: Stress-Test Analytics for Macroprudential Purposes in the euro area*. ECB.

DeMiguel, V. et al. (2020). "A Transaction-Cost Perspective on the ultitude of Firm Characteristics". In: *The Review of Financial Studies* 33, pp. 2180–2222.

Dim, C. (2021). "Should Retail Investors Listen to Social Media Analysts? Evidence from Text-Implied Beliefs". In: -.

Easley, David, Soeren Hvidkjaer, and Maureen O'Hara (2002). "Is information risk a determinant of asset returns?" In: *The journal of finance* 57.5, pp. 2185–2221.

Eaton, G. et al. (2021). "Zero-Commission Individual Investors, High Frequency Traders, and Stock Market Quality". In: -.

ESMA (September 2019). *Stress simulation for investment funds*. ESMA Economic Report.

— (February 2020). *Interconnectedness in the EU fund industry*. ESMA Report on Trends, Risks and Vulnerabilities.

European Central Bank (November 2017). *Financial Stability Review November 2017 – Euro area financial institutions*. Financial Stability Review.

European Securities and Markets Authority (November 2022). *Adverse scenario for the European Securities and Markets Authority's money market fund stress testing guidelines in 2022*.

European Systemic Risk Board (July 2022). *EU Non-bank Financial Intermediation Risk Monitor 2022*. NBFI Monitor No 7.

Fabozzi, Frank J. (1999). *Duration, Convexity, and Other Bond Risk Measures*. New Hope: Frank J. Fabozzi Associates.

Fama, E. and K. French (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of Financial Economics* 33, pp. 3–56.

Fama, E. and J. Macbeth (1973). "Risk, Return and Equilibrium: Empirical Tests". In: *Journal of Political Economy* 81,3, pp. 607–636.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu (2020). "Taming the factor zoo: A test of new factors". In: *The Journal of Finance* 75.3, pp. 1327–1370.

Fiedor, Paweł and Petros Katsoulis (Apr. 2019). *An Lonn Dubh: A Framework For Macroprudential Stress Testing of Investment Funds*. Central Bank of Ireland, Financial Stability Notes Vol. 2019, No. 2.

Financial Stability Board (2021). *Global Monitoring Report on Non-Bank Financial Intermediation*. FSB Report.

— (November 2022a). *Enhancing the Resilience of Non-Bank Financial Intermediation*. Progress report.

— (June 2022b). *Understanding and addressing systemic risks in non-bank financial intermediation*. FSB Conference.

Flora, Maria, Ilaria Gianstefani, and Roberto Reno' (2021). "The liquidity uncertainty premium puzzle". In: *Submitted*.

Fu, Fangjian (2009). "Idiosyncratic risk and the cross-section of expected stock returns". In: *Journal of financial Economics* 91.1, pp. 24–37.

Fusari, N., R. Jarrow, and S. Lamichhane (2021). "Testing for Asset Price Bubbles using Options Data". In: *Johns Hopkins Carey Business School Research Paper* 20-12.

Gerhold, S. et al. (2014). "Transaction costs, trading volume, and the liquidity premium". In: *Finance and Stochastics* 18, pp. 1–37.

Gianstefani, Ilaria, Luigi Longo, and Massimo Riccaboni (2022). *The echo chamber effect resounds on financial markets: a social media alert system for meme stocks*. DOI: 10.48550/ARXIV.2203.13790. URL: https://arxiv.org/abs/2203.13790.

Gianstefani, Ilaria, Naoise Metadjer, and Kitty Moloney (2023). "The interest rate sensitivity of Irish bond funds". In: *Central Bank of Ireland, Financial Stability Notes*.

Giuzio, Margherita et al. (October 2021). *Investment funds, risk-taking, and monetary policy in the euro area*. ECB, Working Paper Series No 2605.

Glosten, L. and P. Milgrom (1985). "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders". In: *Journal of Financial Economics* 14, pp. 71–100.

Goldstein, Itay, Hao Jiang, and David T. Ng (2017). "Investor flows and fragility in corporate bond funds". In: *Journal of Financial Economics* 126.tbc, pp. 592–613.

Green, Jeremiah, John RM Hand, and X Frank Zhang (2017). "The characteristics that provide independent information about average us monthly stock returns". In: *The Review of Financial Studies* 30.12, pp. 4389–4436.

Hasso, T. et al. (2021). "A note on GameStop, short squeezes, and autodidactic herding: An evolution in financial literacy?" In: *Finance Research Letters* In press.

Henry, Jerome and Christoffer Kok (October 2013). *A macro stress testing framework for assessing systemic risks in the banking sector*. ECB Occasional paper series no. 152.

Hou, Kewei, Chen Xue, and Lu Zhang (2020). “Replicating anomalies”. In: *The Review of Financial Studies* 33.5, pp. 2019–2133.

Huang, Jiekun (2018). “The customer knows best: The investment value of consumer opinions”. In: *Journal of Financial Economics* 128.1, pp. 164–182.

International Monetary Fund (October 2022a). *Global Financial Stability Report: Navigating the High-Inflation Environment*. IMF Report.

— (June 2022b). *Ireland, Financial System Stability Assessment*. IMF Country Report No. 22/215.

Jegadeesh, N and S Titman (1993). “Returns on buying winners and selling losers: implications for market efficiency”. In: *Journal of Finance* 48, pp. 65–91.

King, Benjamin and Jamie Semark (April 2022). *Reducing liquidity mismatch in open-ended funds: a cost-benefit analysis*. Bank of England, Staff Working Paper No. 975.

Klein, T. (2021). “A note on GameStop, short squeezes, and autodidactic herding: An evolution in financial literacy?” In: *Finance Research Letters* In press.

Korajczyk, R. and R. Sadka (2004). “Are Momentum Profits Robust to Trading Costs?” In: *Journal of Finance* 59, pp. 1039–1182.

Kyle, A. (1985). “Continuous Auctions and Insider Trading”. In: *Economatrica* 53, pp. 1315–1336.

Lane, Philip R. and Kitty Moloney (2018). “Market-based finance: Ireland as a host for international financial intermediation”. In: *Financial Stability Review, Non-Bank Finance: Trends and Challenges*. Ed. by Banque de France. Banque de France.

Ledoit, Oliver and Michael Wolf (2008). “Robust performance hypothesis testing with the Sharpe ratio”. In: *Journal of Empirical Finance* 15.5, pp. 850–859.

Leys, Christophe et al. (2013). “Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median”. In: *Journal of Experimental Social Psychology* 49, pp. 764–766.

Long, C., B. Lucey, and L. Yarovaya (2021). “”I Just Like the Stock” versus ”Fear and Loathing on Main Street”: The Role of Reddit Sentiment in the GameStop Short Squeeze”. In: *Social Science Research Network* preprint.

MacKinlay, A Craig (1997). “Event studies in economics and finance”. In: *Journal of economic literature* 35.1, pp. 13–39.

Mao Y.and Wei, W., B. Wang, and B. Liu (2012). “Correlating S&P500 stocks with Twitter data.” In: -. Ed. by First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research, pp. -.

Newey, W. K. and K. D. West (1987). “A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix”. In: *Econometrica* 55.3, pp. 703–708.

Novy-Marx, R. and M. Velikov (2016). “A Taxonomy of Anomalies and Their Trading Costs”. In: *The Review of Financial Studies* 29, pp. 104–147.

Nutz, M. and J. Scheinkman (2020). “Shorting in Speculative Markets”. In: *Journal of Finance* 75, pp. 995–1036.

Page, Lawrence et al. (Nov. 1999). *The PageRank Citation Ranking: Bringing Order to the Web.* Technical Report 1999-66. Previous number = SIDL-WP-1999-0120. Stanford InfoLab. URL: `http://ilpubs.stanford.edu:8090/422/`.

Pastor, L. and R.F. Stambaugh (2003). “Liquidity Risk and Expected Stock Retums”. In: *Journal of Political Economy* 111.3, pp. 642–685.

Pedersen, L. (2021). “Game On: Social Networks and Markets”. In: *NYU Stern School of Business Forthcoming*.

Pereira, J. and H. Zhang (2010). “Stock returns and volatility of liquidity”. In: *Journal of Financial and Quantitative Analysis* 45, pp. 1077–1110.

Pereira, Joao and Harold Zhang (2010). “Stock returns and the volatility of liquidity”. In: *Journal of Financial and Quantitative Analysis*, pp. 1077–1110.

Preis, T., H. Susannah Moat, and E. Stanley (2013). “Quantifying Trading Behavior in Financial Markets Using Google Trends”. In: *Scientific Reports* 3, p. 1684.

Quagliarello, Mario (2009). *Stress Testing the Banking System: Methodologies and Applications*. Cambridge: Cambridge University Press.

Renault, T. (2018). “Market manipulation and suspicious stock recommendations on social media”. In: *(Forthcoming)*.

Ríos, S. et al. (2019). “Semantically enhanced network analysis for influencer identification in online social networks”. In: *Neurocomputing* 326, pp. 71–81.

Ruiz, E. et al. (2012). “Correlating Financial Time Series with Micro-Blogging Activity”. In: -. Ed. by First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research.

Sakakibara, Takumasa et al. (2015). “Clustering Mutual Funds Based on Investment Similarity”. In: *Procedia Computer Science* 60, pp. 881–890.

Sydow, Matthias et al. (Aug. 2021). *Shock amplification in an interconnected financial system of banks and investment funds*. ECB, Working Paper Series.

Thukral, Sachin et al. (2018). “Analyzing Behavioral Trends in Community Driven Discussion Platforms Like Reddit”. In: *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 662–669. DOI: 10.1109/ASONAM.2018.8508687.

Umar, Z. et al. (2021). “A tale of company fundamentals vs sentiment driven pricing: The case of GameStop”. In: *Journal of Behavioral and Experimental Finance* 30, pp. 2214–6350.

US Securities and Exchange Commission (2021). “Staff Report on Equity and Options Market Structure Conditions in Early 2021”. In: -.

Vasileiou, E., E. Bartzou, and P. Tzanakis (2021). “Explaining Gamestop Short Squeeze using ntraday Data and Google Searches”. In: -.

Vernimmen, Pierre et al. (2014). *Corporate Finance: theory and practice*. Wiley.

Vodenska, Irena et al. (2021). “From stress testing to systemic stress testing: The importance of macroprudential regulation”. In: *Journal of Financial Stability* 52.

Zheludev, I., R. Smith, and T. Aste (2014). “When Can Social Media Lead Financial Markets?” In: *Scientific Reports* 4, p. 4213.