



HAL
open science

Stock Market Liquidity : Transaction Costs, Crowding and Price Formation Process

Amine Raboun

► **To cite this version:**

Amine Raboun. Stock Market Liquidity : Transaction Costs, Crowding and Price Formation Process. Economics and Finance. Université Paris sciences et lettres, 2021. English. NNT : 2021UPSLD029 . tel-03650904

HAL Id: tel-03650904

<https://theses.hal.science/tel-03650904>

Submitted on 25 Apr 2022

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

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

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à Université Paris-Dauphine

**Stock Market Liquidity: Transaction Costs, Crowding and
Price Formation Process**

Soutenue par

Amine Raboun

Le 02 Mars 2021

École doctorale n°543

Ecole doctorale SDOSE

Spécialité

Sciences économiques

Composition du jury :

Gaëlle Le Fol	Professeur, Université Paris-Dauphine	<i>Président</i>
Robert Kosowski	Professeur, Imperial College	<i>Rapporteur</i>
Frederic Abergel	Professeur, CentraleSupélec	<i>Rapporteur</i>
Thierry Foucault	Professeur, HEC Paris	<i>Examineur</i>
Frank de Jong	Professeur, Tilburg University	<i>Examineur</i>
Marie Brière	Professeur associé, Université Paris-Dauphine	<i>Directeur de thèse</i>
Charles-Albert Lehalle	Professeur associé, Imperial College	<i>Examineur</i>
Fabrice Rahmouni	Head of Index Dept, Euronext	<i>Examineur</i>

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

Les opinions exprimées dans cette thèse sont ceux de l'auteur uniquement et ne représentent pas forcément le point de vue d'Euronext Paris ou d'une quelconque autre institution.

A Abdallah et Soumia ...

Acknowledgments

It was during my master's degree in Finance (El Karoui) that my interest in the financial markets has awakened. The diversity of the lectures, going from hedging of vanilla structured products to high-frequency trading strategies, made me realize how vast is the world of finance. At that moment, I made the wish to become an "expert" of the financial markets and decided to go for a Ph.D. program. A bit more than 3 years later, I am probably not yet an expert, but I had the privilege to deepen my knowledge on different aspects related to finance. The thesis format of CIFRE (industrial partnership) helped a lot in this learning process and allowed me to confront the academic imperative with the constraints of business, and get the views of both the practitioners and academics on the questions I was studying. This adventure is the richest experience so far, both professionally and personally.

First of all, I want to share my enthusiasm regarding the choice of my jury: I admire the work of each member and I sincerely respect the quality of their work in their respective fields. My first acknowledgments go to Frederic Abergel and Robert Kosowski who accepted to be this thesis referee (rapporteurs). I know it is not an easy task, and I thank you for the enormous amount of time you devoted to reading my thesis attentively. I thank Gaelle Le Fol, Thierry Foucault, and Frank de Jong for accepting to evaluate my research, for your constructive and encouraging comments, and your interest in my research in general. I am convinced that your inputs will be very helpful to improve my papers, and I am looking forward to the discussion.

I would like to express all my gratitude and appreciation to my thesis supervisors, Marie Brière and Charles-Albert Lehalle. You are great researchers, of course on the scientific aspect, but also on the human level. I learned a lot from you and discovered new thematic under your guidance. Thank you for your support, your trust, your wise advice, and especially your availability, sometimes even in the middle of your holidays. It is thanks to you that I was able to carry out this work smoothly. The diversity of your research interest, coupled with your incredible career in Finance allowed me to develop

a deep understanding of the financial system as a whole, starting from the finest interactions between market players in central limit order-books, to investment decisions of asset managers.

I would like to warmly thank Gaelle Le Fol and Tamara Nefedova for being there since the beginning of this adventure. You've followed the progress of my thesis closely, including moments of doubts and panic and those of good achievements and glory. Thank you for your unlimited support both on the pedagogical and psychological level. Being part of my CST "Comité de Suivie de Thèse" has created a close interaction between the DRM and Leda research departments. I enjoyed all our meetings, seminars, and conferences we attended together. It was always constructive and fun.

I am infinitely grateful to Euronext, especially the Index department, for allowing me to realize my Ph.D. thesis in the best conditions - not only materially but in all aspects. Indeed, I benefited from complete freedom in the choice of the subjects and the orientation of my research and I always felt supported in the various steps of my Ph.D. No need to mention the top edge equipment, and IT infrastructure I benefited from, including the computation power of the cloud services and the access to the various internal databases which helped me a lot throughout my research. I would like to thank the whole team for fully integrating me into the life of the company, for keeping me informed on the progress of all the projects, even those I couldn't get involved in, and for making me part of all the strategic decisions within the team. My first thanks go to Fabrice. Frankly speaking, this research would not have been possible if not for your unconditional support. For me, you are the living example of an eternal optimist. You always see the glass half full, even in the most delicate situations. I am always amazed by how you approach problems, as they are great opportunities to expand the business. Thank you for being an attentive ear and for always giving me valuable advice. Above all, thank you for sharing your deep knowledge of the financial markets, always with infectious enthusiasm. Geoffrey, thank you for teaching me how to submit a DVA ;) No joke, it was the most valuable lesson I got on the operational/ business mindset. Thank you for always being a business-oriented person, for the team spirit, and for sharing your ideas about good

investment opportunities. Jakub, your presence in the team made the most tedious tasks, fun exercises. Thank you for your scientific curiosity, for the interesting discussions about factor investing, portfolio construction, and liquidity. Thank you for taking repeatedly the time to answer my questions about the trade databases. I know I can be tiring with my questions. Above all, thank you for introducing me to the strategic board games. The parties with Shiva and the rest were fun and I am looking forward to more. Robin, Thank you for always cheering for me during my period of doubt during the thesis. I am happy we got you back on the team, and I am looking forward to working with you. I would like to thank the Index team in Amsterdam, Albert, Harrold, Boudewijn, Fleurisse, and Victor. Our business trips to Amsterdam could not be as fun without you. Thank you the youths of Euronext for our little foolishness at the summer and Christmas parties. And thanks to all the others who made our workplace enjoyable.

I would like to thank Paris-Dauphine University, particularly the economics department for the interesting research life I could benefit from. Thank you to the administrative team for your continuous help. I had the privilege to have very valuable discussions with many researchers in the department. I want to thank all researchers and Ph.D. students I encountered during those 3 fascinating years. I liked the variety of research fields in the Leda department, covering a broad spectrum of societal issues including the environment and global warming, health and aging, development of the third world countries, energy and raw materials, finance, and monetary policies. All nourished by the latest developments in the theoretical reflections on microeconomic foundations, welfare economics, market design, and interactions

I would like to particularly thank my wonderful colleagues with whom I shared the famous P157 office. I will start with Eugénie, cause I will never forget our nice little fights about opening/closing the window and our discussions about plastic bottle usage. Thank you for organizing the dinners at your place early in our Ph.D. period and throwing the Hallowing party. These events, although seldom, created a bond between us and we all felt being part of the same group even if we are working in separate fields. Noémie, thank you for motivating me to work on Saturdays, the days where my papers advanced

the most. Leslie, thank you for assuming the role of the dean, not only for our office but for all Ph.D. students in the Leda department. Clara for the good vibes and positive energy whenever you are at the office. Quentin, since both of us conduct a CIFRE Ph.D. our moments together were brief but very impactful. I enjoyed our discussions about green funds. Thank you all for the bowling party. Finally thanks to our juniors Paul and Mathilde. Take good care of our lovely office !!

I would like to thank the team at Capital Fund Management. Guillaume, Elise, Valentin, Antoine, and José. I won't forget our moments in the desert of Morocco for those who were part of this trip.

Finally, my acknowledgment would not be complete without thanking my family who have supported me throughout my life and who have always been a source of inspiration and motivation. Father, thank you for always being there at the right moment, for giving me invaluable life advice. Mom, thank you for your love and care and for worrying more than us when we have concerns. I hope I will always make you proud. My older brother Oussama, you doing a Ph.D. thesis at the same time in the same university, has created a deeper bond between us. We had a lot of fun during this period and It is what made it a fruitful and successful period for both of us. My sister Chaimae and my brother Ayoub, you don't know how proud I am of you. Chaimae for starting your business at a young age without much help from anyone and Ayoub our professional swimmer, for accumulating victories and good rankings at the national level. I know that lately, with the COVID-19 pandemic, we don't have many opportunities to meet physically, but know that I am always thinking of you.

Thank you to the ones that I forget.

Abstract

This thesis proposes three original contributions, in the form of autonomous papers, to the literature on transaction costs, price formation process, and crowding. The first chapter of the thesis extensively studies the profitability of one of the most implemented strategies in the asset management industry nowadays, namely the factors, sometimes also referred to by Asset Pricing Anomalies or Smart Beta. These strategies are expected to generate above average risk-premia but also involve more portfolio rebalancing, which can generate significant transaction costs. We assess to what extent market frictions constitute a limit to arbitrage for these strategies, and what would be their break-even capacities given the trading behavior of institutional investors. The second chapter highlights the important role of institutional investors' synchronous trading in forecasting transaction costs. However, the congestion of institutional trades is not measurable before the start of the trading session and can only be known ex-post. We provide a methodology to estimate transaction costs in a crowded environment, using a Bayesian network that captures the dependencies of the imbalance of investor order flows and investors' historical trading decisions, to better predict orders transaction cost. Finally, the third chapter models the response of market-makers to the overall liquidity supply and demand in different regimes of uncertainty. We then challenge the model implications during the highly uncertain period of the COVID-19 outbreak.

Résumé

Cette thèse propose trois contributions originales, sous la forme de trois articles autonomes, à la littérature sur les coûts de transaction, le processus de formation des prix et l'encombrement des marchés. Le premier chapitre de la thèse étudie de manière approfondie la rentabilité de l'une des stratégies les plus implémentées aujourd'hui dans le secteur de la gestion d'actifs, à savoir les facteurs, aussi nommés parfois Smart Beta. Ces stratégies sont réputées de générer une prime de risque supérieure à la moyenne, mais elles impliquent également un rebalancement dynamique des titres en portefeuille en fonction de leurs caractéristiques, ce qui amène à des frais de transactions importants. Nous évaluons dans quelle mesure les frictions du marché constituent une limite à l'arbitrage pour ces stratégies, et quelles seraient leurs capacités d'investissement maximales compte tenu des habitudes de négociation des investisseurs institutionnels. Le deuxième chapitre souligne le rôle important de l'encombrement des ordres implémentés par les investisseurs institutionnels dans la prévision des coûts de transaction. Cependant, l'encombrement n'est pas mesurable avant le début de la séance de négociation et ne peut être connu que ex-post. Nous proposons une méthodologie pour estimer les coûts de transaction dans un environnement encombré, en utilisant un réseau bayésien qui saisit les dépendances du déséquilibre des flux des ordres des investisseurs présents sur le marché et le comportement de négociation habituel des investisseurs, afin de mieux prédire le coût de transaction des ordres. Finalement, le troisième chapitre modélise la réponse des teneurs de marché à l'offre et à la demande globale de liquidité sous différents régimes d'incertitude. Nous testons ensuite les implications du modèle pendant la période de l'épidémie de COVID-19 réputé très incertaine.

Contents

Acknowledgments	5
Abstract	9
Résumé	11
List of figures	17
List of tables	20
General Introduction	21
Introduction Générale	37
1 Trading Costs of Asset Pricing Anomalies	55
1.1 Introduction	56
1.2 Data	60
1.3 Methodology:	67
1.3.1 Implementation Shortfall	67
1.3.2 Portfolio implementation cost	70
1.4 Results	74
1.4.1 Non parametric estimation	74
1.4.2 Parametric estimation	78
1.4.3 Break-even capacity	80
1.5 Conclusion	81
References	82

Appendices	85
1.A Data preparation	85
1.B Fama-French Portfolio Construction	86
1.C Definitions / Equations	89
1.C.1 Portfolio's turnover	89
1.C.2 Garman Klass Volatility definition	90
2 Transaction Costs Modelling in a Crowded Environment	91
2.1 Introduction	92
2.2 Related literature	96
2.3 ANcerno database	99
2.4 Transaction cost modelling	101
2.4.1 Order size	101
2.4.2 Order flow imbalance	102
2.4.3 Joint effect of order size and order flow imbalance	104
2.5 Bayesian network modelling with net order flow imbalance as latent variable	108
2.5.1 Bayesian inference	109
2.5.2 Bayesian network modelling	111
2.5.3 Forecasting implementation shortfall	120
2.5.4 Inference of investors order flow imbalance given post-trade cost and market conditions	123
2.6 Conclusion	126
References	128
Appendices	132
2.A Garman Klass volatility definition	132
2.B Bid-Ask spread and volatility distribution dependencies	132
2.C Beta distribution properties	133
2.D Net order flow imbalance properties	133
2.E Implementation shortfall distribution	134
2.F Hasting-Metropolis algorithm	134

3	Liquidity Provision in different Uncertainty Regimes	137
3.1	Introduction	138
3.2	Model and assumptions	142
3.3	Data	146
3.4	Kyle model estimation	153
3.5	Implications on the Price Formation Process	158
3.6	Conclusion	162
	References	163
	Appendices	166
3.A	Model Proof	166
3.B	Micro price	169
3.C	Liquidity metrics	170
	3.C.1 Volatility:	170
	3.C.2 Bid-Ask spread:	170
	General Conclusion	171

List of Figures

1.1	Parametric vs Non-Parametric Estimation of Implementation Shortfall	75
1.2	Break even capacity	80
1.3	Merging referential, market data, fundamental data and trade level data together	87
2.1	A simple graphical model for trading costs modelling	98
2.2	Order size effect on trading costs	102
2.3	Net order flow imbalance effect on trading costs	104
2.4	Joint effect of order size and net order flow imbalance on trading costs	105
2.5	Bayesian inference of a linear regression	110
2.6	Bayesian network for transaction costs modelling	113
2.7	Net order flow imbalance distribution given meta-order side	115
2.8	Net order flow imbalance as a function of participation rate	115
2.9	Inferred net order flow imbalance given the side and the size of the meta-order	118
2.10	Bayesian Inference of Net Order Flow Imbalance	125
2.11	Bid-Ask spread and volatility distribution dependencies	132
2.12	Net order flow imbalance, daily returns correlation	134
2.13	Implementation shortfall marginal distribution	135
3.1	French stock market reaction to COVID-19 Pandemic	148

List of Tables

1.1	Descriptive Statistics of ANcerno Institutional Execution Database	62
1.2	Characteristics of ANcerno Traded Stocks	64
1.3	ANcerno Tickets Characteristics	66
1.4	ANcerno Parent Tickets Average Implementation Shortfall	69
1.5	ANcerno % Coverage of Anomalies Sub-Portfolios	71
1.6	Model calibration on ANcerno tickets	73
1.7	Non-parametric estimation results	77
1.8	Parametric estimation results	79
1.9	ANcerno Variables	88
2.1	Transaction cost model	107
2.2	OLS regression vs Bayesian regression	111
2.3	Bayesian inference: Bid-Ask spread, volatility dependencies	114
2.4	Bayesian inference: Net order flow imbalance dependencies	117
2.5	Bayesian inference: Implementation shortfall dependencies	119
2.6	Performance of the Bayesian network compared to the standard OLS regression	122
2.7	Performance of the Bayesian network given order size	124
2.8	Bayesian inference of net order flow imbalance	126
2.9	Net Order Flow Imbalance auto-correlation	134
3.1	Summary statistics on Market conditions	147
3.2	Summary statistics on Account type trading	151
3.3	Cross-sectional regression of market-makers contribution to the micro-price on the aggregate liquidity demand	157

3.4 Cross-sectional regression of intraday price changes on the daily liquidity demand	161
---	------------

General introduction

Motivation

Transaction costs are a major preoccupation of the asset management industry. Because like any other cost they erode funds' performance, reduce their attractiveness, and decrease the asset manager's revenue. Inversely, the judicious management of the costs supported by the fund creates more opportunities to improve investment returns and to collect more inflows. But unlike any other costs, the transaction costs depend on the fund's characteristics and investment style. Joenväärä et al., 2014 show how the rebalancing frequency and fund size impact the performance persistence of hedge fund portfolios through transaction costs. For example, choosing to over-weight small-capitalization firms in an equity portfolio may result in improving the portfolio gross return but it also increases the cost of implementing the strategy. Similarly, increasing the portfolio rebalancing frequency may improve the market timing and increase on-paper performance, but the resulting profit could be offset by the subsequent higher turnover. Therefore, transaction costs are important drivers of investment decisions and should be considered during the process of investment. Otherwise, a strategy that may seem profitable at first glance and able to generate a significant risk premium may no longer be profitable after accounting for transaction costs.

Two main reasons have made the question of transaction cost fairly important during the last decade. The first reason is regulatory. Regulatory bodies all over the world have introduced a series of new measures to improve the execution quality and to increase the transparency of the brokerage industry. On one hand, they prohibited broker-dealers from

selling offers that bundle research and execution services, to avoid what the US refers to as “soft dollars”. On the other hand, they obliged broker-dealers to produce detailed and standardized transaction costs reports to prove they have taken all the necessary steps to obtain the best possible execution cost for their clients. The second reason is related to the rapid transformation of the asset management industry. The assets managed by the industry worldwide has grown fairly rapidly in dollar value, with AUM (Asset under Management) increasing threefold since the beginning of the 21st century, going from 37.3 trillion U.S. dollars in 2004 to reach 111.2 trillion U.S. dollars in 2020 (PWC, 2020). Besides, the industry has concentrated around a few large players managing a substantial amount of AUM. The world’s top 10 largest asset managers account for more than 30% of the global AuM (Haldane et al., 2014), and this trend is bound to continue in the near future as the biggest firms have also the biggest share in new inflows. In 2018, the top ten US players captured 81% of net mutual fund inflows (Joe Carrubba, 2019). Therefore the size of the orders submitted by institutional investors is getting larger and larger. Thirdly, the move into passive and tracking strategies has increased the potential for investor herding and correlated market movements (Bolognesi and Andrea Zuccheri, 2008). Furthermore, passive investment has seen the emergence of new systematic strategies that departs from the traditional free-float market cap weighting scheme, such as Smart Beta, and Factor-based investing, which raised more questions about their implementability and maximum capacity.

The transaction cost of large orders is dominated by the market impact¹, estimating the transaction cost involves estimating the most probable price path during the execution. Therefore, the transaction costs depend on several parameters other than order characteristics (order side and size) and asset liquidity (bid-ask spread and volatility). It depends for instance, on the execution aggressivity, duration, and trade scheduling (Almgren and Chriss, 2001), on market conditions, on the fluctuations of the aggregate supply and demand (Cardaliaguet and Lehalle, 2018), on the information content of the trade (Grossman and Stiglitz, 1980), and on the anticipation of market players to this information. Hasbrouck, 1988 argues that the impact of trades depends on their predictability, e.g. the

¹The market impact is the price movement due to the execution of the order

highly predictable trades have little impact. Kyle, 1985 argues that trades convey a signal about private information, and market-makers set their price according to their risk aversion to adverse selection.

Transaction costs are the bridge between the asset-pricing literature, which is interested in the efficient price set by risk-return considerations, and the literature on the microstructure of financial markets, which is interested in price changes at the level of order-book events. In this thesis, we present theoretical, methodological, and empirical contributions to two aspects of this literature. In the first chapter, we analyze the profitability of the asset pricing anomalies, also known as factor-based investment, after accounting for transaction costs. In the second chapter, we investigate the impact of institutional investors' synchronous trading on transaction costs and provide a methodology to estimate transaction costs in a crowded environment. Finally, we model the market-maker response function in different market conditions with different uncertainty regimes.

Chapter I: Stock Market Liquidity and Trading Costs of Asset Pricing Anomalies

Following the first empirical tests of Capital Asset Pricing Model (CAPM), highlighting the existence and the statistical significance of the market beta in explaining the cross-section of asset price returns, hundreds of academic papers and practitioner researches attempted to identify other risk factors underlying the equity market, starting with the pioneering work of Fama and French, 1993 size and value and Carhart, 1997 momentum factors. Today, we count more than 300 factors (Harvey et al., 2016) the literature refers to them as "Asset Pricing Anomalies". This remarkable development of the literature has led to the emergence of factor investing as a new investment paradigm (Brière, Szafarz, et al., 2016) extensively used in all lines of the industry (for example by sovereign wealth fund Goetzmann et al., 2014, or by pension funds and mutual funds Broeders and Jansen, 2019). However, factor-based portfolios depart from the traditional market-cap weighted scheme and involve high turnover which generates significant transaction costs. Their attractiveness may thus strongly depend on the trading costs associated

with the factor's replication.

In this chapter, we try to answer the following questions. How large are asset pricing anomalies' trading costs? Do they remain profitable after accounting for transaction costs? and What is the maximum capacity of these strategies? A few recent papers propose to estimate the trading costs of the asset pricing anomalies, with different results depending on the market impact model and database considered. We can distinguish between three main approaches. First, papers measuring the transaction cost using proxies of the bid-ask spreads derived from End-Of-Day prices (Hasbrouck, 2009, Corwin and Schultz, 2012, Abdi and Ranaldo, 2017 or Dayri and Rosenbaum, 2015). For instance, Novy-Marx and Velikov, 2015 and Chen and Velikov, 2018 compute the trading costs of a large panel of anomalies using the average of the 4 above mentioned end-of-day estimators of the bid-ask spread. They find that the implementation of size, value, and momentum would generate respectively 48bps, 60bps, and 780bps annual transaction costs. They conclude then that the trading cost dramatically reduces market anomalies profitability and only cost mitigation strategies allow to generate positive net returns. Papers based on end-of-day data have the advantage of not being restricted in terms of the portfolio universe and back-testing period, allowing to study the robustness of the risk premia after transaction costs in different market conditions (Recession vs Growth) and over different geographical zones. However, they can only provide proxies of the effective quoted bid-ask spread and do not account for the additional market impact cost incurred by trading large orders. The second category of contributions estimates transaction costs on tick-by-tick databases. For instance, Korajczyk and Sadka, 2004 apply various price impact models (Glosten and Harris, 1988, Breen et al., 2002) to TAQ data, to measure the trading capacity of Jegadeesh and Titman, 1993 momentum strategies, and find that momentum could only be profitable for relatively small investors holding less than 2 billion US-dollars. These papers have the advantage of deriving the cost from effectively traded orders in the intraday session. However, the estimates provide the cost of a randomly selected trade in the market, including informed trades, retail initiated trades, and aggressive liquidity demanders trades. Thus, these estimates do not account for the trading specificities of asset-managers executing algorithms splitting the trades all along the

execution window. Finally, the third category of papers proposes to estimate the cost effectively paid by institutional investors in implementing factor-based strategies based on an asset manager executions database. For example, Frazzini et al., 2012 uses AQR Capital Management's proprietary execution database and finds a lower transaction cost for the momentum anomaly of 354bps, enabling it to withstand market frictions. They claim that institutional investors pay lower fees in practice than what is documented in the literature. However, it is debatable to what extent the results of a particular fund could be generalizable to the rest of the industry and are not biased by fund-specific attributes such as size or sophistication of execution.

Our approach consists of using a database composed of large institutional investors' executions in the US, including pension funds, mutual funds, and asset managers trades, named ANcerno. This database represents 10% of institutional trades and roughly 8% of market traded volume (Puckett and Yan, 2011). Our contribution is twofold. First, we measure the average cost paid by institutional investors for implementing the asset pricing anomalies based on the average ticket size submitted by ANcerno's clients. We then search in the Ancerno database for meta-orders similar to those simulated by the strategies paper portfolios (same day, same stock, and same direction) and we assume the cost will also be similar. We show that the average investor pays an annual transaction cost of 16bps for size, 23bps for value, 31bps for investment and profitability, and 222bps for momentum. These estimates give the order of magnitude of the asset pricing anomalies implementation cost but do not take into account the additional cost that could be induced by trading large amounts. Consequently, the second method aims to estimate the transaction costs for portfolios of different sizes. We consider the square root market impact model to account for the order size effect and measure the capacity of each strategy, i.e. the maximum portfolio size that can be reached before the transaction cost cancels the expected profit. The break-even capacities in terms of fund size are \$ 184 billion for size, \$ 38 billion for value, \$ 17 billion for profitability, \$ 14 billion for investment, and \$ 410 million for momentum. We find that the asset pricing anomalies implementation costs are 60 % cheaper than what is estimated in the literature, based on transaction cost proxies such as bid-ask spread or order book models which by construction does not

replicate the trading behavior of institutional investors. We find that institutional investors pay on average a bit less than half a spread. Therefore, the full bid-ask spread traditionally considered as the lower bound cost is rather a conservative measure. We conclude that once the "Market impact" is accounted for, transaction costs significantly reduce the profitability of large portfolio anomalies. However, these anomalies remain profitable for medium-sized portfolios.

Chapter II: Modeling Transaction Costs in a Crowded Environment

Traditional models in the transaction cost literature have mainly focused on the order size effect and the characteristics of the security traded such as the bid-ask spread, price return volatility, and average trading volume (Almgren et al., 2005, Bacry et al., 2015). Similarly, the subsequent answers formulated for the optimal liquidation strategies have for a long time considered the impact of the trader's order only, taking the case of one asset manager executing a large number of shares, aware that his trading pressure detrimentally moves his own price (see Almgren and Chriss, 2001 mean-variance framework, or Gueant and Lehalle, 2015 liquidity-driven optimal control). It is until recently that the literature considered the effect of the crowd as a significant factor in explaining the magnitude of the transaction costs. Bucci et al., 2020 argues that price impact is a function of the aggregate net volume generated by all market participants. Therefore, shared indiscriminately between all traders, whatever the size of their order. A small-sized order will cost nearly as much as a large order if executed in the same direction during the same time frame. Capponi and Cont, 2019 argues that transaction costs depend not only on one's order but also on the behavior of the rest of the market participants. They compared the effect of the order size and the order flow imbalance and concluded that investors should focus on modeling the aggregate dynamics of market pressure during the execution period, rather than focusing on optimizing market impact at a trade-by-trade level. To this increased awareness of the simultaneous trading effect, optimal execution strategies also took that effect into consideration. Cardaliaguet and Lehalle, 2018 formulated a Mean Field Game optimal liquidation strategy where the trader executes strategically while dealing with price

changes generated by other similar market participants.

However, in practice, the crowding cannot be observed by market participants in real-time. Brokers and market-makers can have a broad view of the imbalance of their clients' flows before the trading session starts, and the asset management dealing desk can only observe their own trading instructions. In this chapter, we propose to shed light on the following questions: What is the impact of institutional investors' synchronous trading on transaction costs? How does it compare to the order size effect? How could we model and estimate transaction costs in a **crowded environment** using partially observable variables?

To answer these questions, we quantify the level of Crowding in the market using the imbalance of large investors meta-orders (Capponi and Cont, 2019). This variable translates as a pressure on the market that either adds an additional cost when the sign of the asset manager's order feeds the market pressure or reduces the cost when the manager's order relief the pressure by providing liquidity to the market. Since the order flow imbalance is not observable (a latent variable), we propose a transaction costs model based on Bayesian networks. This type of model, called graphical models, has interesting features that make them suitable for this analysis. They can give, by probabilistic inference, an estimate of their latent variables once all the other variables are observed, and in an iterative approach improve the estimate as soon as new information is revealed. Besides, unlike many other Machine Learning models, Bayesian networks are not black boxes: one can explicitly model the probabilistic dependencies between variables while taking into account the specificities of each of them by incorporating prior information in the model. In practice, the model can be taught on a database provided ex-post by brokers or stock market exchanges, that allow measuring the level of crowding of market participants (the imbalance) and then implement a learning transfer by using a database in which the imbalance can not be observed.

We find that institutional investors' daily order flow imbalance is a good predictor of transaction costs, and we confirm Capponi and Cont, 2019 findings that the dominant variable for implementation shortfall forecast is indeed the order flow imbalance and not the order size. Interestingly, because investors' trading tends to be crowded, the fund

manager, knowing his average correlation with the aggregate flow, can use his own meta-order to infer the order flow imbalance of the market, and then use it to improve his forecast of the transaction costs. We find that this estimation is more accurate when the size of the meta-order is large. Besides, we disclose evidence that a sell order is more informative on imbalance distribution than a buy order, probably because a crowded selling context is more informative about specific market conditions than a crowded buying context. Moreover, the accuracy of transaction costs and market impact estimates are generally relatively low (Bacry et al., 2015). Practitioners have long suspected that the difficulty of estimating orders transaction costs is due to the variance of price innovations that is hardly predictable. Our bayesian forecasting framework confirms this is true. The Bayesian network explicitly models the dependencies between the variance of the residuals and the rest of the network nodes. We find that the dominant variable in modeling the heteroscedastic noise of transaction cost is, indeed, the price volatility. This allows an investor to assess how confident he could be on each prediction given his meta-order and stock characteristics.

Chapter III: Liquidity Provision and Market-Making in different Regimes of Uncertainty

After studying, in the first chapter, the cost of implementing the asset pricing anomalies, strategies rumored to be crowded, and modeling, in the second chapter, the transaction costs in a crowded trading environment, we study, in the third chapter, the market-makers' behavior in response to the crowd. One of the pioneering models explaining the complex interaction between market agents is the Kyle, 1985 model. This model provides an explanation of the behavior of market participants in a game-theoretic approach, making the link between the informational content of prices, the liquidity characteristics of the asset, and the value of private information that some traders may hold. The model assumes the existence of three types of traders: an informed trader, also referred to as an insider, holds private information on the realization of the future payoff of the asset, random or noise traders, buying and selling securities for various reasons (hedging risk,

liquidity constraints, ...), and risk-neutral competitive market makers. In this set-up, the market maker defines the price optimally after observing the overall liquidity demand requested by all traders, informed or not. In contrast, the informed trader trades the maximal possible quantity at an advantageous price, without alerting the market maker risk aversion. His strategy consists then in hiding his volume in the flow generated by noise traders.

A few recent academic articles have documented a change in market makers trading behavior during periods of uncertainty. For example, Megarbane et al., 2017 analyzed the 2015 European Central Bank announcement of a new monetary policy and the 2016 Brexit vote on financial markets and reported a change in HFT trading behavior during these two events. The authors also point out that HFTs are the main market makers on the limit order book, contributing more than 80% of market depth under normal market conditions. However, during these two events, HFTs significantly reduced their liquidity provision, which was taken over by the rest of the market participants. The same behavior was documented by Kirilenko et al., 2017 on the Flash Crash event of May 6, 2010, on the E-mini S&P 500 index futures market. These empirical findings are supported by theoretical literature explaining market makers' quotes by the degree of information asymmetry in the market. Atiase and Bamber, 1994, Bamber et al., 1997, and Tung, 2000 show that bid-ask spreads are increasing with the level of information asymmetry. When important financial results are published or macroeconomic events occur, bid-ask spreads get larger because market-makers' risk aversion to adversarial selection increases significantly during these periods.

In this chapter, we explore whether market-makers react solely to market flows, or take into account exogenous information in their pricing rule. To do so, we propose an extension to the Kyle, 1985 model with two distinct regimes of uncertainty in the market: A low uncertainty regime, where no major announcements are made neither on the economy nor on the firm's idiosyncratic variables, where the market-maker have no reason to fear abnormal informative prices and a regime of high uncertainty that could be triggered by micro- and macro-events². Our model assume that the variance of the risky

²the results are easily generalizable to a model with more than two levels of uncertainty (for example,

asset is time-dependent and a function of the uncertainty regime. The informed trader observes the realization of the risky asset and chooses his traded amount accordingly to maximize his profit, while the market-maker holds an estimate of the level of uncertainty in the market and sets his price accordingly and in reaction to market flows. All the remaining assumptions are kept identical to Kyle's original model. In this configuration, the market maker is better informed, but he remains in a situation of asymmetric information because he only knows the variance and not the value of the risky asset and may be mistaken about the level of uncertainty. However, this level of information is optimal from an information cost perspective. If the informed trader needs to acquire complete information on the risky asset in order to define her optimal traded amount, the market maker needs only a binary signal to choose whether or not to provide liquidity. During periods of high uncertainty, he could increase the risk premium for liquidity provision and quote more aggressive during low regimes of uncertainty.

We provide empirical validation of the model based on the period marked by the Covid-19 health crisis. Baker et al., 2020 argue that no other infectious disease, including the Spanish flu, has ever impacted the stock market as powerfully as the COVID-19 pandemic. As it could be seen in the unprecedented high levels of the VIX index, markets have been taken by surprise and were unable to anticipate a shock of this nature. The crisis provides, therefore, a unique opportunity to test theories about the price formation process and the behavior of market participants during different regimes of uncertainty. We test the model on the 40 stocks composing the CAC40 index using Euronext tick-by-tick database from January 1st to April 1st, 2020. We propose a methodology to empirically estimate Kyle's lambda on central limit order books. It consists of deriving the price set by market makers from their activity in the order book. Namely, insertions, and cancellations of volume in the best bid and best ask limits. Then, estimating the elasticity of this price to the overall market liquidity demand as specified by Kyle's model. We then distinguish the impact of the uncertainty due to the emergence of the pandemic on market-makers pricing rule. We note that market-makers quotes are indeed a function

low, medium and high) or a model that stipulate a different price reaction for different sources of uncertainty (micro vs macro events)

of the net volume signed by liquidity consumers' orders. The results are significant at 1% and robust to all controls on firms' characteristics and day fixed effects. We find that the elasticity of market-makers' prices to market liquidity demand have increased during the Covid-19 highly uncertain period by 39bps for every 1% increase in net traded volume, suggesting that market-makers indeed take into account the exogenous information in their pricing rule. Finally, we note the rest of market participants increase their share in liquidity provision during periods of high uncertainty. Nevertheless, only the liquidity demand addressed by market makers has a notable effect on the daily price return. Consequently, price reaction to flows is a characteristic to market-participants with an intermediary role.

Bibliography

- Abdi, F., & Rinaldo, A. (2017). A simple estimation of bid-ask spreads from daily close, high, and low prices. *The Review of Financial Studies*, 30(12), 4437–4480.
- Almgren, R., & Chriss, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk*, 3, 5–40.
- Almgren, R., Thum, C., Hauptmann, E., & Li, H. (2005). Direct estimation of equity market impact. *Risk*, 18(7), 58–62.
- Atiase, R. K., & Bamber, L. S. (1994). Trading volume reactions to annual accounting earnings announcements: The incremental role of predisclosure information asymmetry. *Journal of accounting and economics*, 17(3), 309–329.
- Bacry, E., Iuga, A., Lasnier, M., & Lehalle, C.-A. (2015). Market impacts and the life cycle of investors orders. *Market Microstructure and Liquidity*, 1(02), 1550009.
- Baker, S., Bloom, N., Davis, S. J., Kost, K., Sammon, M., & Viratyosin, T. (2020). The unprecedented stock market reaction to covid-19. *Covid Economics: Vetted and Real-Time Papers*, 1(3).
- Bamber, L. S., Barron, O. E., & Stober, T. L. (1997). Trading volume and different aspects of disagreement coincident with earnings announcements. *Accounting Review*, 575–597.
- Bolognesi, E., & Andrea Zuccheri, C. (2008). On the efficiency of benchmarks composition: A behavioral perspective. *University of Bologna*.
- Breen, W. J., Hodrick, L. S., & Korajczyk, R. A. (2002). Predicting equity liquidity. *Management Science*, 48(4), 470–483.

- Brière, M., Szafarz, A. Et al. (2016). *Factor-based v. industry-based asset allocation: The contest* (tech. rep.). Amundi Working Paper.
- Broeders, D., & Jansen, K. A. (2019). Do institutional investors manage factor exposures strategically? *SSRN Working Paper 3446239*. <https://ssrn.com/abstract=3446239>
- Bucci, F., Mastromatteo, I., Eisler, Z., Lillo, F., Bouchaud, J.-P., & Lehalle, C.-A. (2020). Co-impact: Crowding effects in institutional trading activity. *Quantitative Finance*, 20(2), 193–205.
- Capponi, F., & Cont, R. (2019). Trade duration, volatility and market impact. *Volatility and Market Impact (March 14, 2019)*.
- Cardaliaguet, P., & Lehalle, C.-A. (2018). Mean field game of controls and an application to trade crowding. *Mathematics and Financial Economics*, 12(3), 335–363.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57–82.
- Chen, A. Y., & Velikov, M. (2018). Accounting for the anomaly zoo: A trading cost perspective. *SSRN Working Paper 3073681*. <https://ssrn.com/abstract=3073681>
- Corwin, S. A., & Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *The Journal of Finance*, 67(2), 719–760.
- Dayri, K., & Rosenbaum, M. (2015). Large tick assets: Implicit spread and optimal tick size. *Market Microstructure and Liquidity*, 1(01), 1550003.
- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3–56.
- Frazzini, A., Israel, R., & Moskowitz, T. (2012). Trading costs of asset pricing anomalies.
- Glosten, L. R., & Harris, L. E. (1988). Estimating the components of the bid/ask spread. *Journal of Financial Economics*, 21(1), 123–142.
- Goetzmann, W. N., Sorenson, O., & Ang, A. (2014). *Strategy for norway's pension fund global*. Yale School of Management.
- Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, 70(3), 393–408.
- Gueant, O., & Lehalle, C.-A. (2015). General intensity shapes in optimal liquidation. *Mathematical Finance*, 25(3), 457–495.

- Haldane, A. Et al. (2014). The age of asset management. *speech at the London Business School*, 4(4).
- Harvey, C. R., Liu, Y., & Zhu, H. (2016). and the cross-section of expected returns. *The Review of Financial Studies*, 29(1), 5–68.
- Hasbrouck, J. (1988). Trades, quotes, inventories, and information. *Journal of financial economics*, 22(2), 229–252.
- Hasbrouck, J. (2009). Trading costs and returns for us equities: Estimating effective costs from daily data. *The Journal of Finance*, 64(3), 1445–1477.
- Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), 65–91.
- Joe Carrubba, B. M., Dean Frankle. (2019). How asset managers can win in a winner-takes-all world. <https://www.bcg.com/publications/2019/asset-managers-winner-takes-all>
- Joenväärä, J., Kosowski, R., & Tolonen, P. (2014). The effect of investment constraints on hedge fund investor returns.
- Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The flash crash: High-frequency trading in an electronic market. *The Journal of Finance*, 72(3), 967–998.
- Korajczyk, R. A., & Sadka, R. (2004). Are momentum profits robust to trading costs? *The Journal of Finance*, 59(3), 1039–1082.
- Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315–1335.
- Megarbane, N., Saliba, P., Lehalle, C.-A., & Rosenbaum, M. (2017). The behavior of high-frequency traders under different market stress scenarios. *Market Microstructure and Liquidity*, 3(03n04), 1850005.
- Novy-Marx, R., & Velikov, M. (2015). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(1), 104–147.
- Puckett, A., & Yan, X. (2011). The interim trading skills of institutional investors. *The Journal of Finance*, 66(2), 601–633.

- PWC. (2020). Asset & wealth management revolution: Embracing exponential change.
<https://www.pwc.com/gx/en/asset-management/asset-management-insights/assets/awm-revolution-full-report-final.pdf>
- Tung, S. (2000). The effect of information asymmetry on bid-ask spreads around earnings announcements by nasdaq firms. *Review of Pacific Basin Financial Markets and Policies*, 3(03), 331–346.

Introduction Générale

Motivation

L'optimisation des coûts de transaction représente depuis longtemps un centre d'intérêt majeur dans le secteur de la gestion d'actifs, vu qu'ils détériorent la performance des fonds, réduisent leur attractivité et diminuent les revenus du gestionnaire d'actifs. À la différence des autres types de coûts, les coûts de transaction sont partiellement définis par la stratégie d'investissement. Par exemple, le choix de sur-pondérer les entreprises à petite capitalisation dans un portefeuille d'actions permet potentiellement de générer un rendement excédentaire brut, le surcroît de rentabilité peut, quant à lui, être compensé par le coût de mise en œuvre de la stratégie. De même, choisir d'augmenter la fréquence de révision de la composition du portefeuille, doit en théorie optimiser le moment de l'investissement et donc la rentabilité du portefeuille. Cependant, l'accumulation des coûts de transaction à chaque révision peut rendre la stratégie peu attrayante. Par conséquent, le coût d'implémentation est un facteur important dans la décision d'investissement et doit être pris en compte, au même degré que le risque et le rendement au cours du processus d'investissement. Sinon, une stratégie qui peut sembler au premier abord rentable peut être perdante après avoir comptabilisé les coûts de transaction.

Deux raisons principales ont placé la question des coûts de transaction au centre des débats sur la gestion des actifs au cours de la dernière décennie: La première raison est d'ordre réglementaire. Les régulateurs des marchés financiers ont introduit une série de réformes visant à rendre le service de courtage plus transparent et améliorer la qualité d'exécution. D'une part, suite à MIFID II, il est interdit aux courtiers de ven-

dre des offres qui proposent à la fois des services d'exécution et offrants la recherche dans le même produit commercial. Ceci permet d'éviter ce que les Américains appellent "soft dollars". D'autre part, ils ont rendu indispensable la publication de rapports détaillés et standardisés sur les coûts de transaction et le prix obtenu après chaque exécution afin de prouver que le courtier a pris toutes les mesures nécessaires pour obtenir le meilleur coût d'exécution possible pour son client. Ceci a soulevé la question de ce qui constitue une bonne exécution. La deuxième raison est liée à la transformation rapide du secteur de la gestion d'actifs. D'une part le montant des fonds gérés par l'industrie n'a cessé d'augmenter d'une année à l'autre avec une vitesse phénoménale. Les actifs sous gestion (AuM: Assets under Management) ont triplé depuis le début du 21^{ème} siècle, passant de 37,3 milliards de dollars US en 2004 à 111,2 milliards de dollars US en 2020 (PWC, 2020). D'autre part, les actifs sous gestion se sont concentrés chez un nombre infime de grands investisseurs institutionnels gérant une part importante des actifs mondiaux : D'après Haldane et al., 2014, les dix plus grands investisseurs institutionnels gèrent plus de 30% des actifs mondiaux. Cette tendance serait amenée à se poursuivre dans les prochaines années, car les plus grands gestionnaires sont également ceux qui reçoivent la plus grande part des nouveaux flux. En 2018, les dix premiers acteurs américains dans l'industrie ont gagné 81% des flux positifs des fonds communs de placement (Joe Carrubba, 2019). Finalement, l'investissement passif a accru le risque d'encombrement des investisseurs sur des stratégies similaires et par conséquent a augmenté le risque systémique de corrélation des mouvements du marché (Bolognesi and Andrea Zuccheri, 2008). En outre, l'investissement passif a vu l'émergence de nouvelles stratégies systématiques qui s'écartent de la pondération traditionnelle basée sur la capitalisation boursière, telles que les Smart Beta, l'investissement factoriel, ou l'investissement prenant compte des critères environnementaux, sociétaux ou de gouvernance ESG. Ce qui a soulevé davantage de questions sur le coût d'implémentation de ces stratégies et de leurs capacités maximales.

Étant donné que le coût d'impact du marché³, domine le coût de transaction des ordres à volume important, l'estimation des coûts d'exécution implique d'habitude la

³L'impact du marché est le mouvement de prix dû à l'exécution de l'ordre

prédiction de la trajectoire de prix la plus probable. Mise à part les caractéristiques de l'ordre ⁴ et la liquidité de l'actif⁵, les coûts de transaction dépendent également de plusieurs paramètres comme :

- L'agressivité, la durée et la planification de l'exécution dans le temps (Almgren and Chriss, 2001)
- Des fluctuations entre l'offre et la demande globale et des conditions du marché (Cardaliaguet and Lehalle, 2018)
- Du contenu informatif de la transaction et de l'anticipation des teneurs du marché à cette information (Grossman and Stiglitz, 1980).

Hasbrouck, 1988 affirme que l'impact de transaction dépend de sa prédictibilité, c'est-à-dire les transactions hautement prévisibles ont peu d'impact. Kyle, 1985 mets en relief le fait que les transactions transmettent un signal sur l'information privée détenu par les traders informés, et que les teneurs de marché par rapport à leur aversion au risque de sélection adverse définissent un prix en anticipation à cette information

Les coûts de transaction constituent le pont entre la littérature sur l'évaluation des actifs et la littérature sur la microstructure des marchés financiers. La première s'intéresse à l'efficacité du prix par rapport à des considérations rendement-risque. Quant à la deuxième, elle s'intéresse aux variations de prix au niveau du carnet d'ordres. Dans cette thèse, nous présentons des contributions théoriques, méthodologiques et empiriques aux deux aspects de cette littérature. Dans le premier chapitre, nous analysons la rentabilité nette, après coûts de transaction, des stratégies à vocation de tirer profit des anomalies du marché, connues sous le nom d'investissement factoriel. Dans le deuxième chapitre, nous étudions l'impact du trading synchrone des investisseurs institutionnels sur les coûts de transaction. Une méthodologie est ensuite fournie dans le but d'estimer ces coûts dans un environnement encombré. Enfin, nous modélisons la fonction de réponse des teneurs de marché dans différentes conditions de marché avec différents régimes d'incertitude.

⁴Par exemple la taille de l'ordre et son signe : achat/vente

⁵La liquidité peut être mesuré par plusieurs indicateurs comme la volatilité, l'écart entre prix d'achat et prix de vente...

Chapter I: Liquidité des marchés boursiers et coûts de transaction des anomalies de prix des actifs

Les premiers tests empiriques du modèle d'évaluation des actifs financiers (CAPM) ont permis de mettre en évidence le bêta du marché comme facteur explicatif des rendements des prix des actifs. Suite à ces tests, plusieurs recherches ont été menées dans le but d'identifier d'autres facteurs de risque sous-jacents au marché des actions, en commençant par les travaux pionniers de Fama et French sur les facteurs de taille, de valeur et de tendance "Momentum" (Fama and French, 1993 Carhart, 1997). Nous comptons, aujourd'hui, plus de 300 facteurs (Harvey et al., 2016) que la littérature identifie comme anomalies des actifs "Asset Pricing Anomalies". Ce développement remarquable de la littérature a conduit à l'émergence de l'investissement factoriel comme nouveau paradigme d'investissement (Brière, Szafarz, et al., 2016), largement répandu dans le secteur de la gestion d'actifs (fonds souverains Goetzmann et al., 2014, fonds de pension Broeders and Jansen, 2019, fonds communs de placement et fonds spéculatifs). Toutefois, l'investissement factoriel s'écarte du schéma de pondération traditionnel par rapport la capitalisation boursière. Il implique donc un taux de rotation de portefeuille élevé, et des coûts de transaction importants. L'attrait de ces stratégies d'investissement peut donc dépendre fortement des coûts de transaction associés à la réplique du facteur.

Dans ce chapitre, nous essayons de répondre aux questions suivantes: Quelle est l'importance des coûts de transaction des stratégies basées sur les facteurs d'investissement? Ces stratégies demeurent-elles rentables après prise en compte des coûts de transaction? Quelle est la capacité maximale de ces stratégies?

De récentes recherches proposent d'estimer les coûts de transaction des stratégies d'investissement factoriel. Toutefois, les résultats divergent selon le modèle d'impact sur le prix considéré et/ou la base de données utilisée. On peut distinguer entre trois approches principales :

- L'approximation des écarts entre les cours d'achat et de vente "Bid-Ask spread" dérivés des prix journaliers d'ouverture, de fermeture, le plus élevé et le plus bas (Hasbrouck, 2009, Corwin and Schultz, 2012, Abdi and Ranaldo, 2017 Dayri and

Rosenbaum, 2015). Par exemple, Novy-Marx and Velikov, 2015 et Chen and Velikov, 2018 calculent les coûts de transaction d'un large panel d'anomalies en utilisant l'estimation du Bid-Ask spread par l'échantillonneur de Gibbs du modèle de Roll modifié. Ils concluent que le coût de transaction réduit considérablement la rentabilité des anomalies du marché. Seules les méthodes d'atténuation des coûts permettent aux stratégies d'investissement de générer un rendement net positif. L'avantage de cette approche est qu'elle peut être appliquée à tous les univers d'investissement sur toutes les périodes de test de la significativité du rendement. Cela permet d'étudier la robustesse des primes de risque après coûts de transaction dans différentes conditions de marché (récession vs croissance) et sur différentes zones géographiques. Toutefois, elle ne peut fournir que des approximations de l'écart effectif entre les cours d'achat et de vente, et ne tient donc pas compte du coût d'impact sur le prix qu'entraîne l'exécution d'ordres à volume importants.

- L'estimation du coût de transaction des anomalies de prix des actifs sur les bases de données "tick-by-tick". Par exemple, Korajczyk and Sadka, 2004 mesure la capacité d'investissement des stratégies "Momentum" à travers différents modèles d'impact sur les prix, sur la base de données TAQ (Glosten and Harris, 1988, Breen et al., 2002, Jegadeesh and Titman, 1993). Il conclue que le "Momentum" ne pourrait être rentable que pour un montant d'investissement relativement bas à moins de 2 milliards de dollars US. Cette approche a l'avantage de fournir le coût des ordres effectivement payé en séance de négociation. Toutefois, les estimations fournissent le coût d'une transaction choisie au hasard sur le marché, y compris les transactions initiées par les particuliers et les transactions agressives des demandeurs de liquidité. Ainsi, elles ne tiennent pas compte des spécificités de négociation des gestionnaires d'actifs qui exécutent avec des algorithmes répartissant les transactions tout au long de la fenêtre d'exécution.
- Finalement, l'estimation des coûts de transactions à partir d'une base de données d'exécution d'un gestionnaire d'actifs implémentant les stratégies factorielles. Par exemple, Frazzini et al., 2012 utilise la base de données d'exécution propriétaire

d'AQR Capital Management et trouve un coût de transaction nettement plus faible pour l'anomalie Momentum, lui permettant de résister aux frictions du marché. Il affirme que les investisseurs institutionnels paient en pratique des frais d'exécution moins élevés que ce qui est documenté dans la littérature. Toutefois, on peut se demander dans quelle mesure les résultats d'un fonds particulier pourraient être généralisables au reste du secteur et ne sont pas biaisés par des attributs spécifiques au fonds tels que la taille ou la sophistication de l'exécution.

Notre approche consiste à utiliser une base de données composée des exécutions des grands investisseurs institutionnels aux États-Unis, contenant les transactions des fonds de pension, des fonds communs de placement et des gestionnaires d'actifs, appelée ANcerno. Cette base de données représente 10 % du volume traité par les institutionnelles et environ 8 % du volume négocié sur le marché américain (Puckett and Yan, 2011). Nous proposons de répondre à cette question en deux temps. D'abord, en mesurant le coût de transaction moyen payé par les investisseurs institutionnels pour implémenter ces stratégies sur la base de la taille moyenne des ordres traités par les clients d'ANcerno. Nous recherchons ensuite dans cette base de données des méta-ordres similaires aux rebalancements des portefeuilles stratégiques (même jour, même titre, et même direction) et nous supposons que le coût sera également le même. Cette estimation donne l'ordre de grandeur du coût d'implémentation des facteurs. Par contre, elle ne prend pas en compte le coût d'impact sur le prix qui pourrait être induit par la négociation de montants importants. Par conséquent, la deuxième méthode vise à estimer les coûts de transaction pour des portefeuilles de tailles différentes. Nous utilisons le modèle à racine carrée de l'impact sur le prix pour tenir compte de l'effet de la taille des ordres et nous mesurons la capacité de chaque stratégie ; c'est-à-dire la taille maximale du portefeuille qui peut être atteinte avant que le coût de transaction n'annule le bénéfice attendu. Nous constatons que les coûts de mise en œuvre des stratégies factorielles sont 60 % moins élevés que ce qui est estimé dans la littérature, que ça soit par les méthodes basées sur l'approximation du bid-ask spread ou sur les modèles de carnet d'ordres. Car elles ne tiennent pas compte des caractéristiques d'exécution des investisseurs institutionnels. Nous constatons, par exem-

ple, que les investisseurs institutionnels paient en moyenne un peu moins de la moitié d'un bid-ask spread. Ce qui fait que le bid-ask spread considéré traditionnellement comme le coût de transaction minimal, est en fait une mesure plutôt prudente. Nous concluons qu'une fois l'impact sur le prix est pris en compte, les coûts de transaction réduisent considérablement la rentabilité des stratégies factorielles étudiées. Toutefois, ces stratégies restent rentables pour les portefeuilles de taille moyenne.

Chapter II: Modélisation des coûts de transaction dans un environnement encombré

Les premiers modèles estimant les coûts de transaction se sont concentrés principalement sur l'effet de la taille de l'ordre traité et les caractéristiques du titre négocié, telles que l'écart entre les cours d'achat et de vente, la volatilité et le volume moyen des transactions (Almgren et al., 2005, Bacry et al., 2015). De même, les réponses formulées au problème d'exécution optimale ont longtemps considéré l'impact de l'ordre du négociant uniquement. Par exemple, le modèle moyenne-variance de Almgren and Chriss, 2001, ou le modèle en contrôle optimal sur la liquidité Gueant and Lehalle, 2015, prenant le cas d'un gestionnaire d'actifs exécutant un grand nombre d'actions et conscient que la pression qu'il exerce sur le marché, fait évoluer son propre prix à son désavantage. Ce n'est que récemment que la littérature s'est intéressé à l'impact de la négociation des autres participants au marché "La foule" sur les coûts de transaction. Bucci et al., 2020 mets en évidence que l'impact sur les prix est une fonction du volume net global généré par tous les acteurs du marché à un instant donné et qu'il est partagé entre tous les négociateurs. En effet, un ordre de petite taille coûtera presque autant qu'un ordre de grande taille si les deux ordres sont exécutés au même moment, dans la même direction. Capponi and Cont, 2019 quant à eux affirment que les coûts de transaction ne dépendent pas uniquement des caractéristiques de l'ordre, mais également du comportement des autres participants au marché. Ils ont comparé l'effet de la taille des ordres et le déséquilibre des flux d'ordres des investisseurs institutionnels sur l'estimation des couts de transaction, et ont conclu que les investisseurs devraient se concentrer sur la modélisation de la pression du marché

pendant la période d'exécution, plutôt que de se concentrer sur l'optimisation de l'impact de chaque transaction. À cette prise de conscience de l'effet de l'exécution simultanée, les solutions d'exécution optimales ont également tenu compte de cet effet. Cardaliaguet and Lehalle, 2018 ont formulé une stratégie de liquidation optimale Mean Field Game où le négociateur exécute stratégiquement son ordre tout en tenant compte des changements de prix dû au trading des autres participants.

En pratique, les participants au marché ne peuvent pas observer le comportement de la foule en temps réel. Les courtiers et les teneurs de marché peuvent avoir une vue d'ensemble du déséquilibre des flux de leurs clients avant le début de la séance de négociation, et la salle des marchés de gestionnaires d'actifs ne peut observer que leurs propres instructions. Dans ce chapitre, nous proposons d'éclairer les questions suivantes : Quel est l'impact de la négociation synchrone des investisseurs institutionnels sur les coûts de transaction ? Comment se compare-t-elle à l'effet de la taille des ordres ? Et Comment pourrions-nous modéliser et estimer les coûts de transaction dans un **environnement encombré** en utilisant des variables partiellement observables ?

Pour répondre à ces questions, nous quantifions le niveau de l'encombrement "crowding" sur le marché en utilisant le déséquilibre des méta-ordres des grands investisseurs, appelé dans Capponi and Cont, 2019 "Order Flow Imbalance". Cette variable se traduit par une pression sur le marché qui soit ajoute un coût supplémentaire lorsque l'ordre du gestionnaire alimente la pression du marché, soit réduit le coût lorsque l'ordre fournit la liquidité au marché. Comme le déséquilibre du flux des ordres n'est pas observable (une variable latente dans le jargon bayésien), nous proposons un modèle de coûts de transaction basé sur les réseaux bayésiens. Ce type de modèle, appelé modèle graphique, présente des caractéristiques intéressantes et adapté à cette problématique. Ils permettent, par inférence probabiliste, d'estimer les variables latentes en fonction des variables observables, et par une approche itérative, améliorer la prédiction chaque fois que des nouvelles informations sont révélées. En outre, contrairement à de nombreux modèles d'apprentissage automatique, les réseaux bayésiens ne sont pas des boîtes noires : On peut modéliser explicitement les dépendances probabilistes entre les variables tout en tenant compte des spécificités de chacune d'entre elles et en incorporant des informa-

tions préalables dans le modèle. En pratique, le modèle peut être appris sur une base de données permettant de mesurer le niveau d'encombrement des participants au marché (le déséquilibre des meta-ordres). Ce type de base de donnée est fournie désormais par les courtiers ou les bourses après la séance de trading. Ensuite, transférer l'apprentissage à une base de données qui ne contient pas l'information sur le déséquilibre des flux. Nous constatons que le déséquilibre des flux des investisseurs institutionnels est un bon indicateur des coûts de transaction. Nous confirmons donc les conclusions de Capponi and Cont, 2019 selon lesquelles la variable dominante dans la prédiction du coût d'exécution est le déséquilibre du flux d'ordres du marché et non la taille de l'ordre. Il est intéressant de noter que, comme les investisseurs institutionnels ont tendance à implémenter des stratégies similaires, le gestionnaire de fonds, connaissant sa corrélation moyenne avec le flux global, peut utiliser son propre méta-ordre pour déduire une première estimation du déséquilibre du flux des ordres dans le marché. Ensuite utiliser cette prédiction a priori pour mieux prévoir le coût de son ordre. Nous trouvons que l'estimation du coût d'exécution de l'ordre est d'autant plus précise que la taille de l'ordre est importante. En outre, un ordre de vente est plus informatif sur la distribution du déséquilibre des flux du marché qu'un ordre d'achat. Ceci est expliqué par le fait qu'une pression vendeuse élevée correspond souvent à un contexte de crise dans le marché qui s'applique à tous les participants au marché. Alors qu'un flux acheteur est moins parlant sur les conditions du marché.

Enfin, les praticiens ont longtemps soupçonné que la difficulté d'estimer les coûts de transaction émane de la variance des innovations de prix qui est difficilement prévisible. Grâce à la structure de notre réseau bayésien, nous arrivons à modéliser explicitement les dépendances entre la variance des résidus et le reste des variables. Nous démontrons qu'effectivement la variable dominante dans la modélisation du bruit hétéroscédastique du coût de transaction est la volatilité des prix. Cela permet à un investisseur d'évaluer le degré de confiance qu'il pourrait avoir dans chaque estimation compte tenu de son méta-ordre et des caractéristiques du titre traité.

Chapitre III : Règle de tarification des teneurs de marché dans différents régimes d'incertitude

Après avoir étudié, dans le premier chapitre, les coûts de transaction de quelques stratégies tendanciennes, à savoir les stratégies factorielles, et puis avoir modélisé dans le deuxième chapitre, les coûts de transaction dans un environnement de négociation encombré, nous étudions, dans le troisième chapitre, sur le comportement des teneurs de marché face à la pression de marché générée par tous les négociateurs “la foule”. L’un des modèles pionniers expliquant l’interaction complexe entre les agents du marché est le modèle de Kyle, 1985. Ce modèle fournit une explication du comportement des agents du marché avec une approche de théorie des jeux, en faisant le rapport entre le contenu informationnel des prix, les caractéristiques de liquidité de l’actif et la valeur de l’information privée que peut détenir un trader initié. Le modèle se base sur trois types d’agents : un agent informé, également appelé agent initié car il détient une information privée sur la valeur future de l’actif, des agents aléatoires, négociant des titres pour diverses raisons (couverture de risque, contraintes de liquidité ...), et des teneurs de marché compétitifs et risque neutres. Le teneur de marché fixe le prix de manière optimale après avoir observé la demande globale de liquidité soumise par tous les types de négociateurs, informés ou pas. En contrepartie, le trader informé maximise son profit en négociant la quantité maximale possible à un prix avantageux, sans alerter le teneur du marché. Sa stratégie consiste alors de cacher son volume dans le flux généré par les transactions des traders aléatoires.

Quelques articles universitaires récents ont documenté un changement dans le comportement des teneurs de marché lors de périodes d’incertitude. Par exemple, Megarbane et al., 2017 a analysé l’impact de l’annonce de la Banque centrale européenne de 2015 sur l’adoption d’une nouvelle politique monétaire et l’annonce du résultat du Brexit en 2016 sur le mouvement des prix sur les marchés financiers. Ils ont signalé un changement dans le comportement de négociation des HFT pendant ces deux événements. Les auteurs soulignent également que les HFT sont les principaux teneurs de marché sur le carnet d’ordres à cours limite, contribuant à plus de 80 % de la profondeur du carnet dans

les conditions normales de négociation. Cependant, au cours de ces deux événements, les HFT ont réduit largement leur apport de liquidité, et le reste des participants au marché ont pris le relais. Le même comportement a été documenté par Kirilenko et al., 2017 lors de l'événement Flash Crash du 6 mai 2010 sur le marché à terme de l'indice E-mini S&P 500. Ces recherches empiriques sur le changement de comportement des teneurs de marché pendant les périodes de crises, de publication de résultats ou de hautes incertitudes de manière générale, sont épaulées par une littérature théorique expliquant le comportement des teneurs de marché par le degré d'asymétrie d'information. Atiase and Bamber, 1994, Bamber et al., 1997, et Tung, 2000 montrent que les écarts entre les cours d'achat et de vente "bid-ask spread" sont une fonction croissante du niveau d'asymétrie d'information et s'élargissent lors de la publication de résultats financiers important ou lors de l'annonce de nouvelles mesures macroéconomiques, car l'aversion au risque de sélection adverse des teneurs de marché est élevée pendant ces périodes.

Dans ce chapitre, nous examinons si les teneurs de marché réagissent uniquement aux flux du marché ou s'ils tiennent également compte des informations exogènes dans leur règle de tarification. Pour ce faire, nous proposons une extension du modèle Kyle, 1985 avec deux régimes distincts d'incertitude sur le marché : Un régime de faible incertitude, où aucune annonce majeure n'est faite ni sur l'économie ni sur les variables idiosyncrasiques de l'entreprise, où le teneur de marché n'a aucune raison de craindre un niveau de prix informatifs anormal, et un régime de forte incertitude qui pourrait être déclenché par des événements de type micro ou macro ⁶. Notre modèle suppose que la variance de l'actif risqué dépend du temps et est fonction du régime d'incertitude. Le trader informé observe la réalisation de l'actif risqué et choisit son montant négocié en conséquence pour maximiser son profit, tandis que le teneur de marché détient une estimation du niveau d'incertitude du marché et fixe son prix en conséquence et en réaction aux flux du marché. Toutes les autres hypothèses sont maintenues identiques au modèle original de Kyle. Dans cette configuration, le teneur de marché est mieux informé, mais il reste dans une situation d'asymétrie d'information car il ne connaît que la variance et

⁶Les résultats sont facilement généralisables à un modèle comportant plus de deux niveaux d'incertitude (par exemple, faible, moyen et élevé) ou à un modèle qui stipule une réaction différente des teneurs de marché selon la source d'incertitude (événement micro- ou macro)

non la valeur de l'actif risqué et peut se tromper sur le niveau d'incertitude. Cependant, ce niveau d'information peut être optimal du point de vue du coût de l'information. Si le négociant avisé doit acquérir l'information complète sur l'actif risqué afin de définir le montant optimal à négocier, le teneur de marché n'a besoin que d'un signal binaire pour choisir de fournir ou non la liquidité. En période de forte incertitude, il pourrait augmenter la prime de risque pour la fourniture de liquidité et coter de manière plus agressive en période de faible incertitude.

Nous complétons ensuite la modélisation par une validation empirique sur la période marquée par la crise sanitaire Covid-19. Baker et al., 2020 dénonce qu'aucune autre épidémie de maladie infectieuse, y compris la grippe espagnole, n'a eu un impact aussi fort sur le marché boursier que la pandémie COVID-19. Toutes les mesures habituellement utilisées pour quantifier l'incertitude comme l'indice VIX sont montées en flèche. La crise nous offre alors une occasion unique pour tester les théories sur le processus de formation des prix et le comportement des participants au marché pendant les régimes d'incertitudes. Nous testons le modèle sur les 40 valeurs composant l'indice du marché Français CAC40 en utilisant la base de données tick-by-tick d'Euronext du 1er janvier au 1er avril 2020. Nous proposons une méthodologie pour estimer empiriquement le lambda de Kyle sur les données du carnet d'ordres central à cours limité. Elle consiste à dériver le prix fixé par les teneurs de marché à partir de leur activité dans le carnet d'ordres, les insertions et les annulations de volume dans les meilleures limites d'offre et de demande, et à estimer ensuite l'élasticité de ce prix à la demande de liquidité globale du marché telle que spécifiée par le modèle de Kyle. Nous distinguons par la suite l'impact de l'incertitude dû à l'apparition de la pandémie sur les cotations des teneurs de marché. Nous constatons que les cotations de ces derniers sont en effet fonction du volume net signé par l'ordre des consommateurs de liquidité. Les résultats sont significatifs à 1% et robustes aux contrôles sur les caractéristiques des entreprises et les effets fixes du jour. Nous prouvons qu'une augmentation de 1 % du volume net par rapport au volume négocié entraîne un déplacement des prix des teneurs de marché de 14,45 points de base avant la crise sanitaire. Cette réaction des teneurs de marché a été augmentée de 38,86 points de base pendant le régime de forte incertitude Covid pour atteindre 53,31 points de base.

Enfin, nous constatons qu'en période de forte incertitude, les autres participants du marché participent davantage au rôle d'apport de liquidités. Nous estimons l'impact de la liquidité sur le rendement quotidien des prix en distinguons entre la liquidité fournie par les teneurs de marché et celle fournie par les autres agents qui peuvent la fournir pour diverses raisons, sans que ça ne constitue leur activité principale. Nous trouvons que seule la demande de liquidité adressée par les teneurs de marché a un effet notable sur le rendement quotidien des prix. Par conséquent, la réaction des prix aux flux est une caractéristique des participants au marché ayant un rôle d'intermédiation.

Bibliography

- Abdi, F., & Rinaldo, A. (2017). A simple estimation of bid-ask spreads from daily close, high, and low prices. *The Review of Financial Studies*, 30(12), 4437–4480.
- Almgren, R., & Chriss, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk*, 3, 5–40.
- Almgren, R., Thum, C., Hauptmann, E., & Li, H. (2005). Direct estimation of equity market impact. *Risk*, 18(7), 58–62.
- Atiase, R. K., & Bamber, L. S. (1994). Trading volume reactions to annual accounting earnings announcements: The incremental role of predisclosure information asymmetry. *Journal of accounting and economics*, 17(3), 309–329.
- Bacry, E., Iuga, A., Lasnier, M., & Lehalle, C.-A. (2015). Market impacts and the life cycle of investors orders. *Market Microstructure and Liquidity*, 1(02), 1550009.
- Baker, S., Bloom, N., Davis, S. J., Kost, K., Sammon, M., & Viratyosin, T. (2020). The unprecedented stock market reaction to covid-19. *Covid Economics: Vetted and Real-Time Papers*, 1(3).
- Bamber, L. S., Barron, O. E., & Stober, T. L. (1997). Trading volume and different aspects of disagreement coincident with earnings announcements. *Accounting Review*, 575–597.
- Bolognesi, E., & Andrea Zuccheri, C. (2008). On the efficiency of benchmarks composition: A behavioral perspective. *University of Bologna*.
- Breen, W. J., Hodrick, L. S., & Korajczyk, R. A. (2002). Predicting equity liquidity. *Management Science*, 48(4), 470–483.

- Brière, M., Szafarz, A. Et al. (2016). *Factor-based v. industry-based asset allocation: The contest* (tech. rep.). Amundi Working Paper.
- Broeders, D., & Jansen, K. A. (2019). Do institutional investors manage factor exposures strategically? *SSRN Working Paper 3446239*. <https://ssrn.com/abstract=3446239>
- Bucci, F., Mastromatteo, I., Eisler, Z., Lillo, F., Bouchaud, J.-P., & Lehalle, C.-A. (2020). Co-impact: Crowding effects in institutional trading activity. *Quantitative Finance*, 20(2), 193–205.
- Capponi, F., & Cont, R. (2019). Trade duration, volatility and market impact. *Volatility and Market Impact (March 14, 2019)*.
- Cardaliaguet, P., & Lehalle, C.-A. (2018). Mean field game of controls and an application to trade crowding. *Mathematics and Financial Economics*, 12(3), 335–363.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57–82.
- Chen, A. Y., & Velikov, M. (2018). Accounting for the anomaly zoo: A trading cost perspective. *SSRN Working Paper 3073681*. <https://ssrn.com/abstract=3073681>
- Corwin, S. A., & Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *The Journal of Finance*, 67(2), 719–760.
- Dayri, K., & Rosenbaum, M. (2015). Large tick assets: Implicit spread and optimal tick size. *Market Microstructure and Liquidity*, 1(01), 1550003.
- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3–56.
- Frazzini, A., Israel, R., & Moskowitz, T. (2012). Trading costs of asset pricing anomalies.
- Glosten, L. R., & Harris, L. E. (1988). Estimating the components of the bid/ask spread. *Journal of Financial Economics*, 21(1), 123–142.
- Goetzmann, W. N., Sorenson, O., & Ang, A. (2014). *Strategy for norway's pension fund global*. Yale School of Management.
- Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, 70(3), 393–408.
- Gueant, O., & Lehalle, C.-A. (2015). General intensity shapes in optimal liquidation. *Mathematical Finance*, 25(3), 457–495.

- Haldane, A. Et al. (2014). The age of asset management. *speech at the London Business School*, 4(4).
- Harvey, C. R., Liu, Y., & Zhu, H. (2016). and the cross-section of expected returns. *The Review of Financial Studies*, 29(1), 5–68.
- Hasbrouck, J. (1988). Trades, quotes, inventories, and information. *Journal of financial economics*, 22(2), 229–252.
- Hasbrouck, J. (2009). Trading costs and returns for us equities: Estimating effective costs from daily data. *The Journal of Finance*, 64(3), 1445–1477.
- Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), 65–91.
- Joe Carrubba, B. M., Dean Frankle. (2019). How asset managers can win in a winner-takes-all world. <https://www.bcg.com/publications/2019/asset-managers-winner-takes-all>
- Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The flash crash: High-frequency trading in an electronic market. *The Journal of Finance*, 72(3), 967–998.
- Korajczyk, R. A., & Sadka, R. (2004). Are momentum profits robust to trading costs? *The Journal of Finance*, 59(3), 1039–1082.
- Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315–1335.
- Megarbane, N., Saliba, P., Lehalle, C.-A., & Rosenbaum, M. (2017). The behavior of high-frequency traders under different market stress scenarios. *Market Microstructure and Liquidity*, 3(03n04), 1850005.
- Novy-Marx, R., & Velikov, M. (2015). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(1), 104–147.
- Puckett, A., & Yan, X. (2011). The interim trading skills of institutional investors. *The Journal of Finance*, 66(2), 601–633.

- PWC. (2020). Asset & wealth management revolution: Embracing exponential change. <https://www.pwc.com/gx/en/asset-management/asset-management-insights/assets/awm-revolution-full-report-final.pdf>
- Tung, S. (2000). The effect of information asymmetry on bid-ask spreads around earnings announcements by nasdaq firms. *Review of Pacific Basin Financial Markets and Policies*, 3(03), 331–346.

Chapter 1

Stock Market Liquidity and the Trading Costs of Asset Pricing Anomalies

About this chapter

The authors would like to thank Andrew Chen, Gaëlle Le Fol, Augustin Landier, Giuliano De Rossi, and Paul Schneider for their useful comments that helped to shape this chapter at its early stages. My thank you goes also to the participants of the following conferences for their constructive comments and remarks: the 12th Financial Risks International Forum (Paris, 2018), 2018 Quant Vision Summit in Paris, 12th Annual Hedge Fund Research Conference (Paris, 2020), and the CFA Society France Amundi Conference (Paris, 2020)

Abstract

Using a large database of the US institutional investors' trades, this paper sheds new light on the question of anomalies-based portfolio transaction costs. We find that the real costs paid by large investors to implement the well-identified Fama-French anomalies (size, value, investment, and profitability) and Carhart momentum are significantly lower than documented in the previous studies. We show that the average investor pays an annual transaction cost of 16bps for size, 23bps for value, 31bps for investment and profitability, and 222bps for momentum. The five strategies generate statistically significant net returns after accounting for transaction costs of respectively 4.29%, 1.98%, 4.45%, 2.69%, and 2.86%. When the market impact is taken into account, transaction costs reduce substantially the profitability of the well-known anomalies for large portfolios, however, these anomalies remain profitable for average size portfolios. The break-even capacities in terms of fund size are \$ 184 billion for size, \$ 38 billion for value, \$ 17 billion for profitability, \$ 14 billion for investment, and \$ 410 million for momentum.

Keywords: Trading Costs, Market Impact, Liquidity, Anomalies-based Investments.

1.1 Introduction

Fama and French, 1993 Size and Value, Carhart, 1997 Momentum and Fama and French, 2015 Investment and Profitability factors (also known as “asset pricing anomalies”) are widely recognized as a source of rewarded risk. Nowadays, they are extensively used by investors to build new portfolios that aim to over-perform traditional market cap-weighted portfolios. Broeders and Jansen, 2019 show that pension funds manage factor exposures strategically on their equity portfolios. However, the success of anomalies-driven investments raises important questions regarding the liquidity and scalability of such strategies. Because they involve high turnover and potentially generate significant transaction costs, their attractiveness for portfolio construction may strongly depend on portfolio size and market impact. In this paper, we estimate the cost of trading the well-

identified anomalies based on execution data of institutional investors. Using ANcerno database, composed of large investors' trades including pension funds, mutual funds, and asset managers, we find that the average investor from ANcerno pays an annual transaction cost of 16bps for size, 23bps for value, 31bps for both investment and profitability anomalies and 222bps for momentum. This is the cost of implementing the strategies for an average ticket size submitted by ANcerno institutional clients and corresponds to a total portfolio size of approximately \$ 1 million. This average estimation does not account for the potential additional cost induced by the market impact of large portfolios' holders. Therefore, we explore how robust these asset pricing anomalies are to market impact by estimating the dependence of trading costs to the size of the trades following Kyle and Obizhaeva, 2018 framework. We derive the break-even capacity for each strategy, i.e., the maximum attainable fund size before price impact eliminates profits. We estimate \$ 184 billion for size, \$ 38 billion for value, \$ 17 billion for profitability, \$ 14 billion for investment, and \$ 410 million for momentum, the most frequently rebalanced strategy.

Several authors studied the limits to arbitrage of asset pricing anomalies. For instance, Novy-Marx and Velikov, 2015 estimate trading costs of a large panel of anomalies including Fama-French size and value and Carhart momentum. They find that the implementation of size, value, and momentum would generate respectively 48bps, 60bps, and 780bps annual transaction costs. These costs dramatically reduce market anomalies profitability. However, their trading cost measure, based on Hasbrouck, 2009 Gibbs sampler estimation of the effective spread, is a proxy of the bid-ask spread derived from end-of-day data and can be quite different from real transaction costs incurred during the intraday session. In Chen and Velikov, 2018, the authors use tick-by-tick databases such as NYSE Trades and Quotes (TAQ) to study the post-publication trading costs of 120 stock market anomalies. They compute the average of 4 end-of-day estimates of the bid-ask spread and find an average cost per anomaly of 100bps, corresponding to an average negative net return of -3bps. According to this study, only cost mitigation strategies are able to generate positive net returns. Their estimation of intraday trading costs is however limited to fixed cost and does not account for market impact. Korajczyk and Sadka, 2004 apply various price impact models (Glosten and Harris, 1988, Breen et al., 2002) to TAQ data,

to measure market impact of Jegadeesh and Titman, 1993 momentum strategies. They conclude that the abnormal returns (Fama French 3-factor alpha) are reduced respectively by 204bps and 192bps for the 11/1/3¹ and 5/1/6 momentum strategies. Momentum could only be profitable for relatively small investors holding less than \$2 billion of AUM. Similarly, Lesmond et al., 2004 investigate the profitability of relative strength portfolios, including Jegadeesh and Titman, 1993 6/1/6 momentum. They confront the strategies to a battery of trading cost estimates such as quoted and effective spreads and find that stocks that generate the largest momentum returns are precisely the ones with the highest transaction costs. Depending on the trading cost measure, the net alpha of the strategy is reduced by 544bps to 937bps. Using a dynamic trading model a la Garleanu and Pedersen (2013), Bonelli et al., 2019 develop a closed formula to estimate the capacity of a trading strategy with respect to its gross performance, the liquidity of the underlying securities, and the dynamics of the signal on which the strategy is based, that they apply to four well-known anomalies including value, momentum, and operating profitability. They find that even if the value signal is slow, the strategy has limited capacity, non-existent in the recent period between 2002-2017 because the pure Sharpe ratio is low. Momentum has a low capacity as well, between \$64 and 73 million, because of the quick mean-reverting signal. While operating profitability has a large capacity of \$43 billion for large-caps and \$14 billion for mid-caps because the pure Sharpe ratio is high and the signal is slow. Contrary to the previous studies, Frazzini et al., 2012 base their estimation on the proprietary database of AQR Capital Management's executions and find a lower transaction cost for the momentum anomaly of 354bps. They argue that TAQ database estimates are higher than what institutional investors pay in practice, for two main reasons. On one hand, the models employed are too conservative. On the other hand, TAQ database approximates the average trade, including informed traders, retail traders, liquidity demanders, and those facing high price impact costs. Patton and Weller, 2019 rely on US-based mutual funds returns to estimate anomalies implementation costs. Using Fama and MacBeth, 1973 framework they assess the gap in factor-mimicking portfolio performance for each particular factor

¹Jegadeesh and Titman, 1993 define momentum strategy using three parameters J/S/K, where J is the length of the period over which past returns are calculated in months, K is the holding period in months, and S is the waiting period. "The skip" is also in months

and find that the difference in compensation per unit of market exposure between paper portfolios returns and mutual funds returns are respectively -0.97%, 2.09%, and 5.04% per year for size, value and momentum anomalies. Patton and Weller, 2019 analysis differs from ours in a sense they are interested in estimating the cost of all potential sources of limits to arbitrage, including regulatory constraints, investability and borrowing costs, while our analysis assesses the transaction cost when the trades are possible. However, while their approach gives average estimates of implementation costs, it does not account for market impact related to fund size. Therefore, Patton and Weller, 2019 result should be seen as a lower bound cost of anomalies implementation cost.

Our results on ANcerno database of executions by institutional investors generalizes Frazzini et al., 2012 findings. We compute portfolios' trading costs in two ways. The first method is a non-parametric approach. It consists of averaging, for a given stock and rebalancing day, the costs of ANcerno reported tickets executed during the same day, in the same direction that the simulated portfolio for each anomaly, regardless of the size of the tickets. This method accounts for the precise transaction costs borne on the rebalancing days of the strategy. The second method focuses specifically on capturing the transaction costs dependence on the volume treated due to market impact. Following Kyle's theoretical model, a series of empirical studies demonstrated the concave relation between the implementation shortfall and the order size (Torre and Ferrari, 1999, Moro et al., 2009, Gomes and Waelbroeck, 2015, Bacry et al., 2015, Briere et al., 2019). We estimate a similar model on ANcerno tickets using as explanatory variables the ratio of the ticket size with respect to an average daily turnover, bid-ask spread, and the price returns volatility. Our estimations are lower than those documented from daily or intraday data for two main reasons. First, trade level databases usually referenced in the literature, such as TAQ, do not link single market trades to their originating parent orders. The resulting price deviation is shared by all investors participating in the same trading session and are not linked to the investor originating the trade. Therefore, these databases are more suited to study the price formation process resulting from investors' orders, than to estimate the trading costs paid by a single investor. Second, when we inflict to the strategy the cost of the spread plus the margin, we indirectly assume that large investors do not mitigate trans-

action costs and send only aggressive market orders that consume liquidity. Yet, large investors split their orders over several hours, or even days, depending on the size of the orders and the relevance of the trading signal. They can alternate between limit orders, executed only at the limit price or better, and market orders, which demand immediate execution at the best available price. These market practices can save institutional investors a significant amount of transaction costs.

The paper is organized as follows. In the first section, we describe the databases we use for our study. In the second section, we define the trading cost measure, explain the parametric and non-parametric estimation methods, and compare the results. Finally, we discuss the profitability of five of the most recognized anomalies, Fama-French size, value, investment, and profitability anomalies, and Carhart momentum after accounting for transaction costs. We also assess the break-even capacity of each strategy.

1.2 Data

We obtain institutional trading data for the period from January 1st 1999 to June 30th 2015 from ANcerno Ltd. ANcerno, formerly Abel Noser Corporation is one of the leading consulting companies in providing Transaction Cost Analysis (TCA) in the US. It provides equity trading costs analysis for more than 500 global institutional investors, including pension funds, insurance companies, and asset managers. This database was largely used by academics to investigate institutional investors market practices (see for example Anand et al., 2011, Puckett and Yan, 2011 and Eisele et al., 2019). ANcerno clients contribute to the database by sending batches of their equity trades in order to monitor their execution quality. Therefore, costs estimated on ANcerno are representative of what institutional investors effectively pay for their executions. Previous research confirmed that ANcerno database is free from any survivorship or backfill bias (see Puckett and Yan, 2011), constitute approximately 8% of the total CRSP daily dollar volume (Anand et al., 2013), and 10% of total institutional activity (Puckett and Yan, 2011).

For each execution, ANcerno reports information on the CUSIP and ticker of the stock, the execution time at minute precision, the execution date, execution price, side

(i.e., buy or sell), number of shares traded, commissions paid, whether the trade is part of a larger order, and a number of trade-level benchmarks to evaluate the quality of the execution. For a limited period of time (until 2011), ANcerno database contained client identifiers allowing to link parent orders to institutions executing them. An institution could be either a large mutual fund, a group of funds, or a single fund subscribing to Abel Noser's analytical service. Each institution could have one or several accounts. Table 1.1 provides descriptive statistics of ANcerno trades. In our sample, we successfully track the activity of 1078 institutions with 149 thousand accounts, responsible for 51.3 trillion dollars of transactions, and using the service of 1488 different brokerage firms. Compared to the market volume reported in CRSP, ANcerno accounts for an average of 5.44% over the whole period. However, this proportion varies in time. We observe an increase from 2.28% in 1999 to 8.28% at the end of 2004, then a steady drop from 2005 to 2011, then a revamp of volume after 2011. The sharp decrease in ANcerno volume as a percentage of CRSP after 2005 may be the direct result of the US market fragmentation happening after Reg NMS² regulation, while the high percentage volume after 2011 could be explained by the increase of passive investing. On the contrary, the traded dollar volume varies between \$2060 and \$4506 billion without any visible monotonicity. Part of the volume reported in ANcerno is executed outside the traditional market venues, thus is not reported in CRSP. The traded amount reported in ANcerno is over a trillion dollars every year and is, therefore, large enough to be relevant.

²Reg NMS (Regulation National Market System) is a financial regulation in the US that came into force in 2005 to modernize and strengthen the National Market System for equity securities. One of its major rules concerns market fragmentation.

Table 1.1. Descriptive Statistics of ANcerno Institutional Execution Database

The table gives descriptive statistics on ANcerno trading data for each year of our sample period (From January 1999 until June 2015). The number of institutions refers to the number of unique clientcodes. The number of Funds, Managers or Accounts is the number of unique clientmgrcodes. The number of brokers corresponds to the number of unique broker identifiers from BrokerXref file where the couple clientcode-clientbkrcode is present in ANcerno. The amount traded in \$ is the sum of the dollar volume executed by ANcerno institutions in the sample. The amount traded in % of market volume is the ratio of ANcerno reported volumes w.r.t CRSP daily turnover

	# Institu- tions	# Funds, Managers or Accounts	# Brokers	# Stocks	Amount traded (\$ billions)	Amount traded (% of market volume)
Full Sample	1078	148621	1488	10044	51310	5.44
1999	381	6153	657	6291	2060	2.28
2000	374	6390	669	6239	3181	2.46
2001	401	13654	716	5396	3026	3.37
2002	428	16847	765	4935	3096	5.00
2003	405	26861	751	4930	2667	6.03
2004	408	23112	716	4126	4122	8.28
2005	379	18928	761	4912	3930	6.17
2006	403	22081	753	4773	4232	5.98
2007	381	28999	738	4941	4506	5.35
2008	338	26600	701	4507	4187	3.95
2009	303	41848	650	4207	2875	3.70
2010	258	43227	632	3951	2508	3.69
2011	-	-	675	3884	1828	5.96
2012	-	-	723	3715	2596	8.30
2013	-	-	647	3755	2714	9.39
2014	-	-	531	3809	2742	7.12
June 2015	-	-	392	2753	1033	3.64

ANcerno has the advantage of not being restricted to a single trading venue. It covers the three main US historical venues that compose CRSP and Compustat universe, namely, New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and NASDAQ. Table 1.2 Panel A shows the percentage coverage in terms of number of stocks and market capitalization. By construction, ANcerno is composed of only investible assets as all reported stocks were physically held by US institutional funds. Henceforth, it is slightly biased towards large-capitalization stocks. ANcerno encompasses 92% of NYSE stocks, but slightly more than one-third of AMEX companies. Yet, this difference in coverage should not bias the estimation of the studied asset pricing anomalies implementation cost, as 98% of CRSP overall market capitalization is present in ANcerno and Fama-French methodology is based on market-cap-weighted portfolios. ANcerno comprises a fair amount of small companies as well. In Panel B of Table 1.2, half of the companies are smaller than \$ 226 million worth. While the average firm size is \$ 2.4 billion, ANcerno also covers a broad range of value and momentum stocks, as shown by the spectrum of book-to-equity ratios and preceding 12 months returns covered.

Table 1.2. Characteristics of ANcerno Traded Stocks

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1st, 1999 to June 30th, 2011. The left part of panel A gives the ratio of common stocks (those with a sharecode of 10 or 11) present in ANcerno relative to CRSP database. The right part of panel A displays the ratio on market-capitalization coverage. It is computed as the sum of market capitalization of stocks present in ANcerno divided by CRSP universe total market capitalization. Panel B reports descriptive statistics for stocks traded by ANcerno institutions. We obtain price data (prices, traded volume, outstanding shares) from CRSP database, Book-to-Market from Compustat. GK volatility is the Garman Klass (1980) estimation of the volatility. Turnover is the average percentage of outstanding shares traded on a single day

Panel A: ANcerno Coverage								
	% Number of Stock				% Market Capitalization			
	ALL	NYSE	AMX	NASDAQ	ALL	NYSE	AMX	NASDAQ
Full Sample	0.72	0.92	0.36	0.66	0.98	0.99	0.73	0.96
1999	0.63	0.92	0.35	0.55	0.99	0.99	0.83	0.97
2000	0.67	0.92	0.32	0.62	0.99	0.99	0.83	0.99
2001	0.66	0.93	0.28	0.61	0.99	0.99	0.85	0.98
2002	0.71	0.94	0.33	0.66	0.99	0.99	0.88	0.98
2003	0.77	0.96	0.40	0.73	0.99	1.00	0.88	0.98
2004	0.84	0.97	0.52	0.83	0.99	1.00	0.90	0.98
2005	0.84	0.98	0.51	0.83	0.99	1.00	0.86	0.98
2006	0.85	0.98	0.52	0.84	0.99	1.00	0.81	0.98
2007	0.86	0.98	0.56	0.85	0.99	1.00	0.81	0.98
2008	0.83	0.98	0.43	0.82	0.99	1.00	0.73	0.98
2009	0.83	0.97	0.43	0.80	0.99	1.00	0.68	0.98
2010	0.83	0.97	0.48	0.79	0.99	1.00	0.84	0.98
2011	0.83	0.97	0.50	0.79	0.99	1.00	0.87	0.98
2012	0.43	0.75	0.02	0.29	0.96	0.98	0.21	0.91
2013	0.44	0.74	0.03	0.30	0.94	0.97	0.31	0.89
2014	0.43	0.69	0.03	0.30	0.92	0.95	0.34	0.85
June 2015	0.37	0.59	0.02	0.27	0.86	0.89	0.33	0.80

Panel B: Stock Characteristics				
	mean	25%	50%	75%
Market Capitalization (\$ 100 billion)	24.0	0.57	2.26	9.33
Average Traded Volume (\$ million)	23.7	0.16	1.08	7.72
Book-to-Market Ratio	1.05	0.32	0.57	0.93
Lagged 12-month Return (%)	0.16	-0.36	-0.04	0.42
Turnover (%)	1.25	0.21	0.48	1.03
GK Volatility (%)	29.4	13.0	21.8	37.4

ANcerno database has several advantages over any other trades-level database such as TAQ or TRTH (Thomson Reuters Tick History), which are both abundantly referenced in the academic literature (see, for example, Glosten and Harris (1988), Hasbrouck (1991)). First, ANcerno mainly reports the activity of institutional investors, which are the most likely to implement anomalies-based portfolios, which is the focus of this paper. Second, ANcerno is not restricted to a single stock exchange. In the US, it covers all trading venues present in CRSP. Third, ANcerno links child tickets to their corresponding parent tickets, whether the execution was split into several days or executed in one swoop. While other transaction databases tend to list the entire amount of their orders placed in the market or effectively matched, without putting any link between related trades. This may result in false estimations of portfolios trading costs, polluted by short term opportunistic investors and high-frequency traders (Frazzini et al., 2018). Finally, ANcerno provides more information on the fixed costs born by institutional investors such as broker commissions and trading fees. Table 1.3 gives an overview of ANcerno tickets' characteristics. By a parent-ticket, we mean a buy or sell order sent by an individual fund or manager on a single stock, whether the trading firm chose to split the order across brokers or days. ANcerno provides an identifier per parent ticket, with the corresponding intended volume and execution period, which allow us to track the related child tickets. We observe that both, market conditions and institutional investors' trading behavior changed through time. We note that the number of parent-tickets has increased significantly during the studied period, starting from \$1.92 million in 1999 to attain \$11.66 million in 2008. Inversely, the average parent-ticket size and average participation rate have both constantly been dropping during the period. From \$ 1.12 million to \$ 0.39 million average size, and from 6.42 % to 0.58% participation rate. The execution period has also shrunk from almost 3 days on average to about half 1.5 days at the end of the studied period. It suggests that institutional investors changed their trading scheme from sending relatively few very large orders (i.e., parent-tickets) to their preferred brokers, each of which should be split into several days by the broker, to more controlled algorithms, where the asset manager takes care of the daily execution scheduling, sends relatively smaller orders and asks the broker to implement them in one or two days maximum. Note that this smooth change

in market practices happened at the same time as the increase in automated trading, the decrease of over the counter trades, and the multiplication of the trading venues resulting from post-RegNMS market fragmentation (Laruelle and Lehalle, 2018). Large investors adjusted their trading behavior as a response. Commissions and fixed fees increased from 11.52bps in 1999 to 18.85bps in 2003, then dropped quickly afterward to reach 3.69bps in 2015.

Table 1.3. ANcerno Tickets Characteristics

Institutional trading data are obtained form ANcerno Ltd on the period ranging from January 1st, 1999 to June 30th, 2015. Parent Ticket is an order sent by an institutional investor. It could be split into multiple child tickets. Participation rate is the size of the parent ticket compared to the 12 months average daily volume. Duration is the execution period calculated in days. Commissions are computed as the ratio of Commissions per share divided by the open price of the day the ticket is issued.

	# Parent Tickets (million)	Av Parent Ticket size (\$ million)	% of buy	# Child Tickets	Particip. rate (%)	Duration (days)	Commissions (bps)
Full Sample	6.99	0.62	50.81	2.80	2.27	1.73	9.27
1999	1.92	1.12	53.73	2.66	6.42	2.85	11.52
2000	3.02	1.09	54.64	2.54	4.43	2.24	12.47
2001	3.71	0.87	55.83	2.45	3.89	2.03	14.62
2002	4.30	0.80	54.87	2.66	3.82	2.15	18.85
2003	4.99	0.62	53.62	2.90	3.53	1.96	17.08
2004	6.18	0.92	52.86	3.25	3.38	1.96	13.00
2005	6.73	0.80	51.35	3.26	2.74	1.79	9.99
2006	9.47	0.56	49.75	3.66	1.74	1.56	9.73
2007	10.23	0.55	49.81	4.13	1.48	1.50	7.07
2008	11.66	0.48	48.94	2.71	0.90	1.40	8.60
2009	10.31	0.39	47.64	2.29	0.94	1.47	8.05
2010	10.62	0.37	47.22	2.11	0.84	1.35	4.04
2011	8.20	0.24	48.09	2.25	0.42	1.22	2.89
2012	10.93	0.26	47.82	2.40	0.44	1.26	3.13
2013	7.69	0.37	48.79	2.66	0.65	1.38	3.62
2014	5.93	0.49	48.02	2.93	0.73	1.54	3.69
June 2015	2.82	0.39	49.50	3.59	0.58	1.56	3.69

1.3 Methodology:

To estimate anomalies portfolios trading costs, we calculate the cost of rebalancing each stock belonging to the underlying anomaly portfolio and sum them up with their respective weights. The trading cost of each stock is measured as the sum of the implementation shortfall and fixed costs, including commissions, taxes, and fees.

1.3.1 Implementation Shortfall

Implementation shortfall, as defined in Perold (1988), measures the difference between a theoretical or benchmark price (in our case, the closing price at the time the strategy's desired holdings are generated, i.e., prior day) and an actual traded price, in percent of the benchmark price. Implementation shortfall measures the total amount of slippage a strategy might experience from its theoretical returns. In essence, our cost estimates measure how much of the theoretical returns to a strategy can actually be achieved in practice.

For a parent-ticket m of size $Q_k(m)$ and side $s_k(m)$ ($= 1$ for buy tickets and -1 for sell tickets) executed at date d with $Ntrades$ child tickets, the implementation shortfall is calculated as follows:

$$IS_k(m, d) = \frac{s_k(m)}{P_k^{ref}(d)} \left(\sum_{i=1}^{Ntrades} \frac{v_{k,m}(i)}{Q_k(m)} \times P_k(i) - P_k^{ref}(d) \right) \quad (1.1)$$

where $Q_k(m) = \sum_{i=1}^{Ntrades} v_{k,m}(i)$, $v_{k,m}(i)$ is the volume of each child ticket i related to the parent-ticket m and P_k^{ref} is the closing price of stock k at the review date $d - 1$ of the back-tested strategy. All $P_k(i)$ are happening after the open of day d .

Table 1.4 shows the distribution of ANcerno parent tickets implementation shortfall. Interestingly, we find that the implementation shortfall can be negative. The market movements could be either favorable or detrimental to the trade. In fact, during bullish periods, buy tickets are more expensive than sell tickets and vice-versa. For instance, in 2009, when the market daily average return was 22.79bps, the average implementation shortfall of buy order was 31.41bps, almost twice the bid-ask spread (16.37bps), whereas

sell tickets benefited from this buying pressure with a low 5.84bps average cost. Similarly, in 2008 during the Subprime crisis, sell tickets had an excessive implementation shortfall of 50.72bps. Buy tickets, on the contrary, experienced a price improvement, i.e., negative implementation shortfall, of -16.95bps, the moment the bid-ask spread amounted for 17.27bps. The direction of the ticket compared to the direction of market movements (indicating potential market pressure) is thus an important factor in transaction costs analysis. In addition, Table 1.4 also highlights the dependence of transaction costs to firm size. As expected, large companies are cheaper to trade than smaller ones (8.69bps against 14.35bps). Transaction costs are increasing with the stocks' bid-ask spread and volatility. In 1999, when the bid-ask spreads were the highest on our sample (52.62bps), implementation shortfall was also the highest (42.03bps). Inversely in 2007, the bid-ask spread and implementation shortfall were at their lowest levels (9.19bps and 3.34bps respectively). More volatile periods are also associated with larger bid-ask spreads and higher implementation shortfall. During the above average volatility periods, market-makers aware of adverse selection, revise their limits farther from the mid-price (Sandaas, 2001).

Table 1.4. ANcerno Parent Tickets Average Implementation Shortfall

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1st, 1999 to June 30th, 2015. Our sample include only common stocks (those with a share code of 10 or 11 in CRSP). The split by trading venue is based on CRSP. The split between large and small caps is based on the NYSE median capitalization in December of each year. Buy (respectively Sell) correspond to the average implementation shortfall for buy (respectively sell) tickets. R^M is the average daily return of the equally weighted basket composed of CRSP stocks. σ^{GK} is the average Garman Klass Volatility of CRSP stocks computed over 1 year rolling window. Spread is the monthly average quoted Bid-Ask spread relative to the mid price obtained from TRTH trade database

	25%	50%	Mean	75%	Buy	Sell	R^M	Spread	σ^{GK}
	(bps)	(bps)	(bps)	(bps)	(bps)	(bps)	(bps)	(bps)	(%)
Full Sample	-103.8	7.31	15.05	127.78	9.86	20.74	5.13	17.58	22.78
Large Cap	-64.25	4.21	8.69	79.67	8.54	8.98	7.72	14.35	9.69
Small Cap	-92.60	8.32	14.35	117.81	13.08	15.83	3.79	21.33	19.63
NYSE	-64.69	3.32	6.96	77.42	6.30	7.72	4.53	14.96	10.49
AMEX	-118.3	7.11	11.38	136.37	1.45	20.41	5.33	27.18	18.41
NASDAQ	-77.22	8.55	16.19	105.35	16.56	15.93	6.16	17.73	13.54
1999	-126.2	18.60	43.81	184.33	48.29	38.56	12.57	52.62	26.16
2000	-183.9	15.36	32.48	232.42	22.01	45.01	-3.95	51.73	33.06
2001	-148.2	15.38	25.32	191.28	8.25	46.67	12.17	33.03	32.42
2002	-138.2	11.32	23.36	173.83	-15.0	69.70	-3.93	25.30	26.11
2003	-92.29	11.41	20.00	125.51	30.71	7.52	23.76	15.62	23.74
2004	-79.01	7.21	11.23	100.64	12.42	9.88	8.46	11.66	17.08
2005	-72.67	6.49	8.66	91.10	9.98	7.23	2.01	10.73	15.51
2006	-76.93	4.60	6.30	91.18	7.29	5.22	6.56	9.49	15.67
2007	-88.32	3.23	3.34	98.88	1.74	5.07	-2.27	9.19	16.12
2008	-161.7	5.86	16.23	182.76	-17.0	50.72	-21.0	17.27	25.45
2009	-134.4	8.67	18.88	162.38	31.41	5.84	22.79	16.37	36.70
2010	-90.16	6.67	10.98	110.98	18.99	3.22	11.10	10.52	21.84
2011	-99.19	6.50	12.28	120.03	5.84	18.94	-3.02	10.90	18.03
2012	-69.74	0.00	1.53	72.71	4.27	-1.21	8.54	17.91	9.35
2013	-51.66	4.47	7.39	64.97	9.49	5.28	15.92	13.04	7.40
2014	-59.32	2.23	3.84	66.47	0.52	7.15	2.82	12.53	7.60
June 2015	-59.71	0.00	0.75	61.16	1.41	0.09	-1.30	12.46	7.35

1.3.2 Portfolio implementation cost

We propose two ways to assess the trading cost of anomalies portfolios. The first approach (non-parametric estimation) consists of averaging, for each stock rebalanced on a given day, the transaction costs of all trades for that stock happening that day in the same direction in the ANcerno database. The second approach estimates a model capturing the dependency of trading costs with respect to the traded volume, and applies the estimation results to the backtested anomalies (parametric estimation).

Non Parametric Estimation

Fama-French and Carhart Momentum anomalies are based on the largest possible universe of all stocks listed in the 3 major US trading venues (NYSE, AMEX, and NASDAQ). The paper strategies do not take into account any trading or liquidity constraints. Our non-parametric approach implies restricting the universe to stocks traded by institutional investors and present in ANcerno database. Table 1.5 Panel A shows ANcerno average coverage of the Fama-French six sub-portfolios composing the size, value, investment, and profitability anomalies, while panel B gives the coverage of the four sub-portfolios composing Carhart momentum anomaly (see details about the sub-portfolios construction in Appendix B) over the back-tested period. Large capitalization stocks are only marginally impacted by the universe restriction. More than 97% of the stocks (98% of the total market capitalization) in the original Fama-French and Carhart portfolios are traded in ANcerno database. The universe restriction impacts more heavily the small-cap portfolios. For example, only 37.9% of the stocks traded in the Fama-French Small-High sub-portfolio are traded in ANcerno and 40.9% for the Small-Down sub-portfolio.

When the new composition of the anomalies portfolio becomes effective after the close of day $d - 1$, we extract the weight of each stock k that needs to be rebalanced $\delta w_k(d)$. We consider all ANcerno tickets submitted in day d , on stock k , executed in the direction $s_k(m)$ as the rebalancing trade ($s_k(m) = \text{sign}(\delta w_k(d))$) and average the cost of ANcerno tickets that meet these criteria to obtain an estimation of trading cost $\hat{\xi}_k(d)$ for each stock k (equation 2.2). Finally, we compute portfolio trading cost as the

Table 1.5. ANcerno % Coverage of Anomalies Sub-Portfolios

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1st, 1999 to June 30th, 2015. Prices data are obtained from CRSP database while fundamental data are derived from Compustat. We implement the size, value, profitability and investment anomalies based on Fama-French 6 sub-portfolios methodology (described in the Appendix B) and momentum anomaly following Carhart. The table below shows the percentage number of stocks in each sub-portfolio present in ANcerno and how much it represent in terms of market value.

Panel A: Fama-French Anomalies							
Sub-portfolios	Big High	Big Low	Big Medium	Small High	Small Low	Small Medium	Average
Size, Value							
Number Stocks (%)	97.7	98.4	97.6	37.9	57.3	54.0	73.82
Market Cap (%)	98.88	99.76	98.95	75.45	82.77	80.23	89.3
Investment							
Number Stocks (%)	96.5	99.1	97.5	40.0	56.8	56.2	74.3
Market Cap (%)	99.5	99.1	98.8	77.3	81.9	79.9	89.4
Profitability							
Number Stocks (%)	98.2	98.8	98.0	45.3	51.5	55.9	74.6
Market Cap (%)	99.1	99.3	97.9	80.6	82.5	77.2	89.4
Panel B: Carhart Momentum Anomaly							
Sub-portfolios	Big Down	Big Up	Small Down	Small Up	Average		
Number Stocks (%)	97.5	97.7	40.9	51.1	71.80		
Market Cap (%)	99.12	99.26	76.25	78.96	88.4		

weighted sum of its $Nsec$ components' trading costs multiplied by the size of the portfolio, $AUM(d - 1)$, at the review date.

$$\hat{\xi}_k(d) = \frac{1}{\sum_m \mathbb{I}_{\delta W_k(d) s_k(m) > 0}} \sum_m \mathbb{I}_{\delta W_k(d) s_k(m) > 0} (IS_k(m, d) + fixedcost_k(m, d)) \quad (1.2)$$

$$Prtf Cost (d) = AUM(d) \times \sum_{k=1}^{Nsec} |\delta W_k(d)| \times \hat{\xi}_k(d) \quad (1.3)$$

where $\delta W_k(d)$ is the delta weight of stock k at the review date $d - 1$. $s_k(m)$ is the side of ticket m on stock k . $IS_k(m, d)$ and $fixedcost_k(m, d)$ are respectively the implementation shortfall and the fixed cost of ticket m on stock k . $AUM(d - 1)$ is the portfolio size at the

end of day $d - 1$. N_{sec} is the number of stocks in the portfolio.

Parametric Estimation

One of the drawbacks of the non-parametric estimation is that it only considers stocks that are traded in ANcerno database. Also, this estimation of average transaction costs based on all trades executed in the database does not account for the well-known dependence of the trading costs on the size of the trades. Bacry et al., 2015 model, also used by Frazzini et al., 2018, postulates a squared root relationship between the implementation shortfall and the size of the trade, measured as the fraction of daily volume traded in a stock (see equation 2.3 below).

$$IS_k(m, d) = \alpha \times \psi_k(d) + \beta \times \sigma_k^{GK}(d) \times \sqrt{\frac{Q_k(m)}{ADV_k(d)}} + \varepsilon_k(m, d) \quad (1.4)$$

where $IS_k(m, d)$ is the implementation shortfall of ticket m submitted on stock k at day d . $\psi_k(d)$ is the quoted intraday bid-ask spread of stock k averaged on the month, $Q_k(m)$ is the ticket size, $ADV_k(d)$ is the daily traded volume averaged on a 12 months rolling window, and $\frac{Q_k(m)}{ADV_k(d)}$ is the participation rate, $\sigma_k^{GK}(d)$ is the Garman Klass intraday volatility of stock k estimated on a 12 month rolling window, α , β are model's parameters and $\varepsilon_k(m, d)$ is the error.

To calibrate the model, we consider all tickets reported in ANcerno database in the 12 months preceding the review date $d - 1$ of participation rate $\frac{Q_k(m)}{ADV_k(d)}$ higher or equal to 0.01%. We find that below that threshold, the ticket size has a limited effect on trading cost because of intraday volatility noise. We then form 1000 bins based on the tickets' participation rate and estimate the model on the average implementation shortfall of each bin. The left side of Table 1.6 describes the result of the calibration made at end of each month (152 regressions). We note that the average coefficient of the bid-ask spread is 0.4. When the market impact generated by the ticket is small ($Q_k(m)$ close to 0), arbitrageurs on average do not pay the full bid-ask spread but only 40% of it. Therefore, for ANcerno institutional investors, the bid-ask spread is a conservative estimation of trading costs. The right side of table 1.6 shows the result of 10 thousand bootstraps. In each iteration,

we withdraw randomly 1/12 of our ANcerno sample, constitute the bins and estimate the model. The bootstrap average coefficients 5% confidence interval confirm the rolling window approach. We use the model parameters 5% confidence interval to compute the portfolios' trading costs confidence interval.

Table 1.6. Model calibration on ANcerno tickets

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1st, 1999 to June 30th, 2015. Quoted intraday bid-ask spreads are obtained from TRTH database averaged over the month. σ_k^{GK} and ADV are respectively the Garman Klass volatility and the average trading volume computed from CRSP database over a 12 months rolling window. The coefficients of the rolling regressions are estimated at the end of each month on 1000 bins based on tickets participation rate $\frac{Q_k}{ADV_k}$ where only tickets submitted by institutional investors on the latest 12 months are considered. The bootstrap draws randomly 1/12 of the database tickets. 10 000 regressions are estimated on the 1000 corresponding bins

	1Y Rolling Window		10 ⁴ Bootstrap	
	Ψ_k	$\sigma_k^{GK} \sqrt{\frac{Q_k}{ADV_k}}$	Ψ_k	$\sigma_k^{GK} \sqrt{\frac{Q_k}{ADV_k}}$
Av coef	0.40	0.77	0.42	0.75
2.5% quantile	0.35	0.68	0.37	0.70
97.5% quantile	0.45	0.76	0.46	0.79
std err	0.02	0.001	0.02	0.001
t_{stat}	17.68	37.34	20.11	39.67
$P > t $	0.01	0.00	0.00	0.00
Average R_{adj}^2	0.88		0.91	
Nbr of regressions	152		10000	
Nbr of observations per regression	1000		1000	

In the parametric estimation, we consider the full universe of all CRSP stocks. The estimated model parameters $\hat{\alpha}$ and $\hat{\beta}$ are used to estimate stocks trading costs as shown in equation 2.4. The intraday bid-ask spreads are retrieved from TRTH when available. For missing data, usually small capitalization, we use Abdi and Ranaldo, 2017 CRSP daily spread estimates. It consists of averaging the spread of the daily last available bid and last available ask over the month. For fixed costs, we take the average of commissions and

taxes per ticket computed on the month preceding the review date.

$$\hat{\xi}_k(d) = \hat{\alpha} \times \psi_k(d) + \hat{\beta} \times \sigma_k^{GK}(d) \times \sqrt{\frac{Q_k(d)}{ADV_k(d)}} + \frac{1}{N} \sum_{k,m}^{N} fixedcost_k(m,d) \quad (1.5)$$

Where $Q_k(d)$ is the size of the rebalancing trade. It derives explicitly from the size the portfolio AUM(d-1) such as $Q_k(d) = \frac{AUM(d-1) \times \delta W_k(d)}{P_k^{close}(d-1)}$. Finally, we sum stock trading cost $\hat{\xi}_k(d)$ at portfolio level as in equation 1.3).

In Figure 2.2, we regroup ANcerno tickets in 1000 bins based on participation rate $\frac{Q}{ADV}$ and plot the average implementation shortfall of the tickets in each of the buckets (blue dots) and the non-parametric estimation (black dot). ANcerno tickets show a concave relation between implementation shortfall and ticket size relative to daily traded volume. We observe a sharp increase in the costs from -4bps to 20bps when ticket size increases from 0.01% to 2% of the ADV. The slope decays afterward. For instance, a ticket with a 40% participation rate has an 80bps trading cost. The parametric estimation captures well the dependence to volume and confirms the relevance of the square root dependence of trading costs to traded volume. The non-parametric estimate does not capture volume dependence but represents the average cost paid by institutional investors.

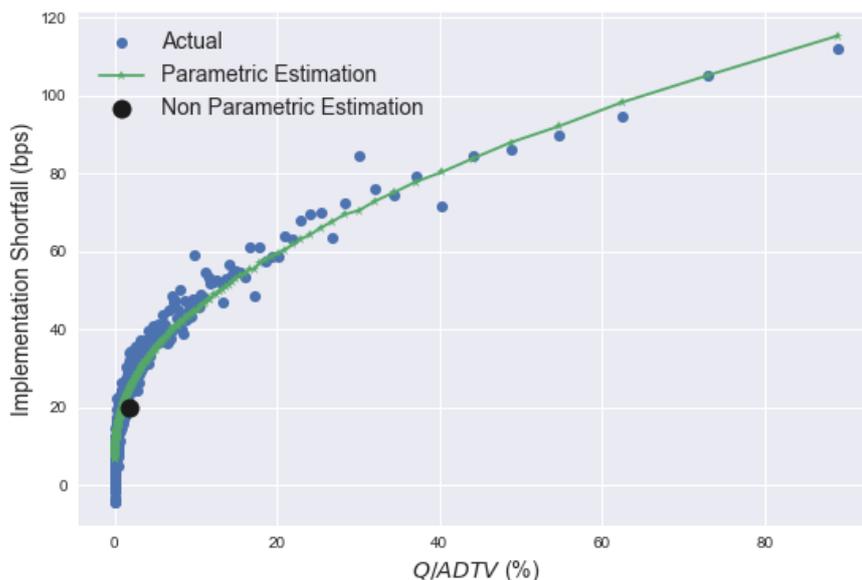
1.4 Results

1.4.1 Non parametric estimation

In this section, we discuss the profitability of Fama-French anomalies size, value, profitability, and investment and Carhart momentum anomaly from the perspective of the average trading costs experienced by ANcerno institutional investors. For each anomaly, we compare (1) the gross performance of the anomaly, when constructed on CRSP universe of all US-listed stocks following Fama-French methodology, (2) the gross performance after restricting to ANcerno universe of stocks traded by institutional investors, and (3) the net performance after accounting for transaction costs. Table 1.7 reports the result

Fig. 1.1. Parametric vs Non-Parametric Estimation of Implementation Shortfall

This figure plots the average implementation shortfall for ANcerno tickets. We sort all trades into 1000 bins based on their participation rate $\frac{Q}{ADV}$. The blue dots represent the average implementation shortfall of each of the bins. The green line is the corresponding predictive value following equation 2.3 averaged by a bin. Finally, the black marker indicates the average tickets participation rate and the average non-parametric estimation



of the backtest starting on the 30th of June 1999 and ending the 30th of June 2015. Hou, Xue, and Zhang (2017) report a significant reduction of most anomalies returns' when restricting the exposure to small and micro-capitalization stocks. Similarly, we find that the performance of size and value anomalies is reduced substantially by the restriction of the investment universe to ANcerno traded stocks, which are not well covered by ANcerno database (from 4.84% to 4.46% for size and from 2.94% to 2.22% for value). For instance, the long leg of size anomaly composed solely of small companies loses 0.42% in performance while the short leg gains 0.05%. For the same reason, both the long and short legs of the value anomaly, which both contain small-capitalization stocks, experienced a substantial loss in performance (0.55% and 0.16% respectively). The momentum and investment anomalies are robust to the universe change (experience a small decay of 0.05%) while the profitability anomaly is impacted positively (4.20% on CRSP vs 4.77% on ANcerno). Israel and Moskowitz (2013) document a similar pattern for momentum,

showing that it is equally strong among large and small capitalizations, while value is stronger among small caps.

Transaction costs significantly reduce the performance of the anomalies. Fama-French anomalies, rebalanced only once per year, have a low turnover and moderate trading costs, 16bps for size, 23bps for value, and 31bps for investment and profitability. Momentum strategy is much more costly to implement. It is rebalanced at the end of each month (60% monthly turnover) and has an average trading cost of 222bps. Trading costs account for 1/3 of the strategy's gross performance. Note however that our estimates are around 0.4 times (two times and a half lower than) Novy-Marx and Velikov, 2015 transaction costs estimates based on daily effective bid-ask spreads (48, 60, and 780bps for size, value, and momentum respectively). This 0.4 is also the bid-ask spread coefficient of the parametric model after calibration. Therefore investors pay around 40% of the bid-ask spread to trade small portfolios, and the full bid-ask spread is rather a conservative measure of trading costs. Moreover, Chen and Velikov, 2018 argue that measures based on end-of-day data tend to over-estimate trading costs. Our estimates are more in line with those of Frazzini et al., 2012 for the momentum anomaly 351bps.

Table 1.7. Non-parametric estimation results

The back history is run from 30th June 1999 to 30th June 2015. Ann Gross return is the strategy annualized average gross return in %. Volatility is the annualized monthly standard deviation in %. Ann Net return is the annualized average return after trading costs are deducted. Ann turnover is the annual turnover computed as in Appendix B.2.

	Ann Gross Returns (%)	Volatility (%)	Ann Net Returns (%)	Ann Turnover	Trading Costs (bps)
Panel A: Size Anomaly (SMB)					
CRSP Univ Portfolio	4.84	12.41	-	0.45	-
CRSP Univ Long Leg	11.99	20.61	-	0.25	-
CRSP Univ Short Leg	-7.16	15.41	-	0.20	-
Ancerno Univ Portfolio	4.46	11.11	4.29	0.54	15.69
Ancerno Univ Long Leg	11.57	20.35	11.42	0.35	13.65
Ancerno Univ Short Leg	-7.11	15.37	-7.13	0.19	2.04
Panel B: Value Anomaly (HML)					
CRSP Univ Portfolio	2.94	12.05	-	0.76	-
CRSP Univ Long Leg	10.60	18.35	-	0.44	-
CRSP Univ Short Leg	-7.67	18.93	-	0.32	-
Ancerno Univ Portfolio	2.22	12.16	1.98	0.76	22.50
Ancerno Univ Long Leg	10.05	18.84	10.02	0.44	2.56
Ancerno Univ Short Leg	-7.83	18.36	-8.03	0.32	19.95
Panel D: Profitability Anomaly (RMW)					
CRSP Univ Portfolio	4.20	10.80	-	0.71	-
CRSP Univ Long Leg	9.80	15.47	-	0.31	-
CRSP Univ Short Leg	-5.60	21.82	-	0.40	-
Ancerno Univ Portfolio	4.77	10.38	4.45	0.76	30.61
Ancerno Univ Long Leg	10.57	15.40	10.42	0.33	13.90
Ancerno Univ Short Leg	-5.80	21.66	-5.96	0.42	16.71
Panel E: Investment Anomaly (CMA)					
CRSP Univ Portfolio	3.08	6.35	-	1.25	-
CRSP Univ Long Leg	9.96	17.77	-	0.68	-
CRSP Univ Short Leg	-6.89	18.69	-	0.57	-
Ancerno Univ Portfolio	3.02	7.00	2.69	1.19	31.39
Ancerno Univ Long Leg	10.47	17.61	10.34	0.64	12.44
Ancerno Univ Short Leg	-7.46	18.63	-7.65	0.55	18.95
Panel C: Momentum Anomaly (UMD)					
CRSP Univ Portfolio	5.15	18.22	-	7.32	-
CRSP Univ Long Leg	11.71	18.46	-	3.28	-
CRSP Univ Short Leg	-6.56	24.52	-	4.04	-
Ancerno Univ Portfolio	5.09	18.52	2.86	8.34	222.49
Ancerno Univ Long Leg	11.71	18.65	10.78	3.84	93.30
Ancerno Univ Short Leg	-6.62	24.81	-9.91	4.50	129.19

1.4.2 Parametric estimation

The parametric method estimates the trading costs of all US-listed stocks, accounting for the dependence of trading costs on the volume traded. We back-test the anomalies on the wide universe of all CRSP stocks over the same sample period than the non-parametric estimation. Table 1.8, reports anomalies net performance for various levels of portfolio size: \$ 1 million, \$ 100 million, and \$ 1 billion of AUM. For comparison, BlackRock US Basic Value Fund, one of the biggest, has a total AUM of \$ 647 million. The table also reports the 5% confidence interval of portfolio trading costs derived from the 2.5% and 97.5% quantiles of model parameters bootstrapped distribution presented in Table 1.6. We note that the \$ 1 million portfolios generate small-sized rebalancing orders that barely meets the threshold of 0.005% participation rate. Hence the orders have limited market impact. The trading costs estimates are of the same magnitude as the non-parametric estimations: 19bps for size, 31bps for value, 43bps for profitability, 44bps for investment, and 253bps for momentum anomaly. The net returns remain significant (4.46% for size, 2.11% for value, 4.24% for profitability, 3.78% for investment and 2.46% for momentum). The mid-sized portfolios of \$ 100 million generate rebalancing orders of the size of 1% to 2% of daily turnover. The trading costs reduce the performance of the anomalies (by 30, 54, 77, 81, and 417 bps respectively) but the net returns still remain attractive for the Fama-French anomalies (4.35%, 1.88%, 3.89%, and 3.40% respectively). For large portfolios of \$ 1 billion, transaction costs are twice as big (56, 108, 156, 167, and 741 bps respectively) and thus significantly reduce the net returns of the anomalies. While size remains largely profitable over the sample period (4.09% net return), value, profitability, and investment are more heavily impacted (1.33%, 3.10%, 2.55% net return), and momentum loses all appeal (net return are significantly negative [-2.35%, -1.69%]) as the trading costs exceed the gross return of 5.15%.

Table 1.8. Parametric estimation results

The back history is run from 30th June 1999 to 30th June 2015. Ann. Net return is the annualized average return after trading costs are deduced. Volatility is the annualized monthly standard deviation (in %). Annual turnover is computed as in Appendix B.2. Av Participation Rate is the average ticket size w.r.t daily turnover rebalanced by the anomalies. The intervals between brackets correspond to the 5% confidence interval derived from the 2.5% and 97.5% quantiles of bootstrapped model parameters

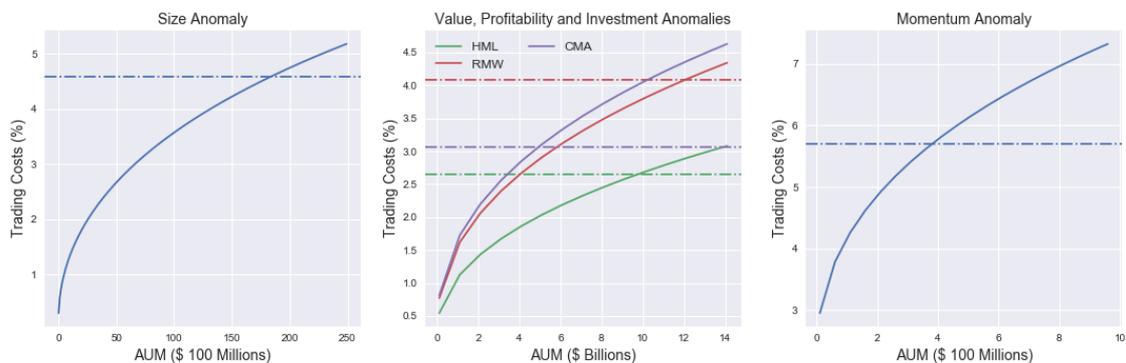
AUM	Ann. Net Return (%)	Volatility (%)	Trading Costs (bps)	Turnover (Monthly)	Av Particip Rate (%)
Panel A: Size Anomaly (SMB)					
\$1 million	4.46 [4.45 4.48]	12.32	19.17 [17.86 20.48]	0.04	0.00
\$100 million	4.35 [4.33 4.37]	12.32	30.11 [28.14 32.07]	0.04	0.47
\$1 billion	4.09 [4.06 4.13]	12.32	56.04 [52.53 59.54]	0.04	4.57
Panel B: Value Anomaly (HML)					
\$1 million	2.11 [2.09 2.14]	11.61	30.84 [28.71 32.98]	0.06	0.01
\$100 million	1.88 [1.85 1.91]	11.61	54.22 [50.75 57.68]	0.06	1.28
\$1 billion	1.33 [1.27 1.40]	11.61	108.77 [102.27 115.24]	0.06	12.21
Panel D: Profitability Anomaly (RMW)					
\$1 million	4.24 [4.21 4.27]	11.21	42.56 [39.77 45.3]	0.06	0.02
\$100 million	3.89 [3.85 3.94]	11.21	77.05 [72.35 81.74]	0.06	1.82
\$1 billion	3.10 [3.01 3.19]	11.21	156.44 [147.54 165.31]	0.06	17.00
Panel E: Investment Anomaly (CMA)					
\$1 million	3.78 [3.75 3.80]	7.54	43.76 [40.88 46.63]	0.10	0.02
\$100 million	3.40 [3.35 3.45]	7.54	81.07 [76.11 86.02]	0.10	1.97
\$1 billion	2.55 [2.45 2.64]	7.54	166.64 [157.12 176.12]	0.10	18.31
Panel C: Momentum Anomaly (UMD)					
\$1 million	2.46 [2.30 2.62]	18.20	253.45 [236.95 269.93]	0.610	0.01
\$100 million	0.93 [0.70 1.16]	18.19	417.43 [393.59 441.11]	0.610	0.49
\$1 billion	-2.03 [-2.35 -1.69]	18.17	741.13 [705.26 776.40]	0.610	3.61

1.4.3 Break-even capacity

In this section, we study how robust the asset pricing anomalies are to market impact by assessing the break-even capacity of each strategy. Figure 2.3 document the break-even capacities of each strategy on the period ranging from the 30th June 1999 to the 30th June 2015. The continuous lines plot anomalies trading costs with respect to portfolio size, while dashed lines express anomalies returns. The intersection between dashed and continuous lines points to the break-even capacity of each anomaly, which is the maximal fund size attainable before price impacts eliminate profits. We find that size is the most capacitive, with \$ 184 billion break-even capacity corresponding to 4.84% average cost. Followed by value with \$ 38 billion capacity and 2.94% trading costs, then investment and profitability with respectively \$14 and \$17 billion. Finally, the momentum is the most challenged by trading cost \$ 410 million and 5.15% costs. The limited capacity is partially due to the bad performance of the momentum strategy during our sample period, as it suffered from the 2001 Internet burst and 2007 financial crisis drawdown.

Fig. 1.2. Break even capacity

The left-hand figure plots the trading costs of the low turnover anomalies with respect to portfolio size in \$ billion. The green line for size and the blue lines for value). The right-hand figure shows the result for the monthly rebalanced momentum anomaly. The Dashed lines represent the annualized average return of the anomalies on their respective back-testing period



1.5 Conclusion

Using a broad database of investors' trades, accounting for 10% of institutional investors activity and 8% of total market volume in the US stock market, we estimate the cost of trading of five of the most recognized anomalies in the literature: Size, Value, Investment, Profitability, and Momentum. We find that the average cost paid by investors to trade small orders represent less than 40% of the effective quoted bid-ask spread. This means that institutional investors are able to mitigate their transaction costs in an efficient way. Quoted bid-ask spreads, used in previous studies as a proxy for transaction costs estimates (Novy-Marx and Velikov, 2015, Chen and Velikov, 2018) are rather a conservative measure of trading costs as it assumes that all institutional orders are liquidity consuming. For average-sized portfolios (approximately \$ 1 million), corresponding to average volumes traded in ANcerno database, we find low trading costs for Fama-French (16bps for size, 23bps for value, and 31bps for investment and profitability), because of their low annual turnover. Momentum is more heavily impacted by trading costs (222bps, one-third of the gross returns). Our estimates for the momentum anomaly trading costs are more in line with Frazzini et al., 2012 estimates. After accounting for market impact, estimating the dependence of trading costs to the traded volume using the square root model, we compare the trading costs at three different levels of portfolio size and seek the break-even capacity of each strategy. We conclude that the momentum anomaly is the most expensive to trade. It is monthly rebalanced and involves a high turnover (around 60%). It reaches its capacity at \$410 million. The Fama-French are rebalanced once per year and thus, are more robust to trading costs. Size has an estimated capacity of \$ 184 billion, while the estimated capacity of value, investment, and profitability are respectively \$38, \$14, and \$17 billion. The trading costs estimated on ANcerno database present the great advantage that they are representative of what institutional investors effectively pay to execute their orders. However, the exact cost of implementing the long-short Fama-French portfolios should take into account in addition the costs of short-selling (stocks borrowing costs, and leverage funding costs). Thus, the estimates should be considered as an upper bound on the capacity of these strategies, an extension that we leave for future work.

References

- Abdi, F., & Rinaldo, A. (2017). A simple estimation of bid-ask spreads from daily close, high, and low prices. *The Review of Financial Studies*, 30(12), 4437–4480.
- Anand, A., Irvine, P., Puckett, A., & Venkataraman, K. (2011). Performance of institutional trading desks: An analysis of persistence in trading costs. *The Review of Financial Studies*, 25(2), 557–598.
- Anand, A., Irvine, P., Puckett, A., & Venkataraman, K. (2013). Institutional trading and stock resiliency: Evidence from the 2007–2009 financial crisis. *Journal of Financial Economics*, 108(3), 773–797.
- Bacry, E., Iuga, A., Lasnier, M., & Lehalle, C.-A. (2015). Market impacts and the life cycle of investors orders. *Market Microstructure and Liquidity*, 1(02), 1550009.
- Bonelli, M., Landier, A., Simon, G., & Thesmar, D. (2019). The Capacity of Trading Strategies. *HEC Paris Research Paper No. FIN-2015-1089*. <https://doi.org/http://dx.doi.org/10.2139/ssrn.2585399>
- Breen, W. J., Hodrick, L. S., & Korajczyk, R. A. (2002). Predicting equity liquidity. *Management Science*, 48(4), 470–483.
- Briere, M., Lehalle, C.-A., Nefedova, T., & Raboun, A. (2019). Stock market liquidity and the trading costs of asset pricing anomalies. *Available at SSRN 3380239*.
- Broeders, D., & Jansen, K. A. (2019). Do institutional investors manage factor exposures strategically? *SSRN Working Paper 3446239*. <https://ssrn.com/abstract=3446239>
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57–82.
- Chen, A. Y., & Velikov, M. (2018). Accounting for the anomaly zoo: A trading cost perspective. *SSRN Working Paper 3073681*. <https://ssrn.com/abstract=3073681>
- Eisele, A., Nefedova, T., Parise, G., & Peijnenburg, K. (2019). Trading out of sight: An analysis of cross-trading in mutual fund families. *Journal of Financial Economics*. <https://doi.org/https://doi.org/10.1016/j.jfineco.2018.12.005>
- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3–56.

- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1–22.
- Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3), 607–636.
- Frazzini, A., Israel, R., & Moskowitz, T. (2012). Trading costs of asset pricing anomalies.
- Frazzini, A., Israel, R., & Moskowitz, T. J. (2018). Trading costs.
- Glosten, L. R., & Harris, L. E. (1988). Estimating the components of the bid/ask spread. *Journal of Financial Economics*, 21(1), 123–142.
- Gomes, C., & Waelbroeck, H. (2015). Is market impact a measure of the information value of trades? market response to liquidity vs. informed metaorders. *Quantitative Finance*, 15(5), 773–793.
- Hasbrouck, J. (2009). Trading costs and returns for us equities: Estimating effective costs from daily data. *The Journal of Finance*, 64(3), 1445–1477.
- Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), 65–91.
- Korajczyk, R. A., & Sadka, R. (2004). Are momentum profits robust to trading costs? *The Journal of Finance*, 59(3), 1039–1082.
- Kyle, A. S., & Obizhaeva, A. A. (2018). The market impact puzzle. *Anna A., The Market Impact Puzzle (February 4, 2018)*.
- Laruelle, S., & Lehalle, C.-a. (2018). *Market microstructure in practice*. World Scientific.
- Lesmond, D. A., Schill, M. J., & Zhou, C. (2004). The illusory nature of momentum profits. *Journal of Financial Economics*, 71(2), 349–380.
- Moro, E., Vicente, J., Moyano, L. G., Gerig, A., Farmer, J. D., Vaglica, G., Lillo, F., & Mantegna, R. N. (2009). Market impact and trading profile of large trading orders in stock markets.
- Novy-Marx, R., & Velikov, M. (2015). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(1), 104–147.
- Patton, A. J., & Weller, B. (2019). What you see is not what you get: The costs of trading market anomalies. *Journal of Financial Economics (JFE)*, Forthcoming.

Puckett, A., & Yan, X. (2011). The interim trading skills of institutional investors. *The Journal of Finance*, 66(2), 601–633.

Sandaas, P. (2001). Adverse selection and competitive market making: Empirical evidence from a limit order market. *The Review of Financial Studies*, 14(3), 705–734.

Torre, N., & Ferrari, M. (1999). The market impact model tm. barra research insights. 1999 barra. Inc.

Appendices

1.A Data preparation

This section of the appendix provides a detailed description of databases content, and how we link them together to get the final data source. The main ANcerno dataset is a sample of institutional transaction-level data directly submitted by ANcerno's clients. The data are submitted in batches that include all trades executed by a client during the interval of time covered by the batch. The exact length covered by each batch, however, is not predefined and can range from a few trading days to several months of trades. A large variety of clients rely on ANcerno's monitoring services. The dataset includes transactions reported by several of the main mutual fund families domiciled in the United States, a small number of hedge funds, several pension plan sponsors, and a multitude of asset managers. A client from ANcerno's perspective is any entity that submits trades, which generally consists of an individual mutual fund, a group of funds, or a fund manager. ANcerno assigns unique codes to the clients (variable `clientcode`) and the corresponding institution as reported by the client (variable `clientmgrcode`). However, the exact identity of the client is always anonymized.

For a limited period of time, ANcerno also provided a file ("MasterManagerXref") including the list of the overarching institutions to which the individual clients were affiliated (i.e., the fund families in the case of mutual funds). This additional file includes the name of the institution (variable `manager`), e.g., PIMCO, and a number identifying the institution (variable `managercode`), e.g., 10. We match this file to another file ("ManagerXref") (that includes both the numbers identifying the institutions (variable `managercode`) and the client codes (variable `clientcode`). In this way, we are able to match the main institution name with the original ANcerno trade data via client codes (as the variables `clientcode` and `clientmgrcode` are included both in the "ManagerXref" file and in the main ANcerno file)—see Figure 1.3. Similarly, we link broker identifiers to trades using a fourth file obtained by ANcerno, which is called "BrokerXref". All this information is necessary to better define a ticket and thus related trades, whether the originating house

chose to split their trading on time or across different brokers. The main variables that we use from the ANcerno database are reported in Table 1.9.

As far as stock market data are concerned, we use CRSP daily and monthly stock files provided by WRDS (Wharton Research Data Services), from which we retrieve stock prices (open high low close), daily traded volume, outstanding shares, exchange code, and share code. Stocks balance sheets and fundamental information are obtained from Compustat Annual fundamental files, also provided by WRDS. We first match these two databases using CRSP-COMPUSTAT historical link table that maps Compustat GVKEY stock identifier to CRSP (PERMNO, PERMCO) couple, then we use the resulting table to cross with ANcerno main dataset using the CUSIP. Finally, we check that ANcerno reported prices fall within the range of CRSP daily low-high prices. We drop the few tickets that do not fulfill this condition.

1.B Fama-French Portfolio Construction

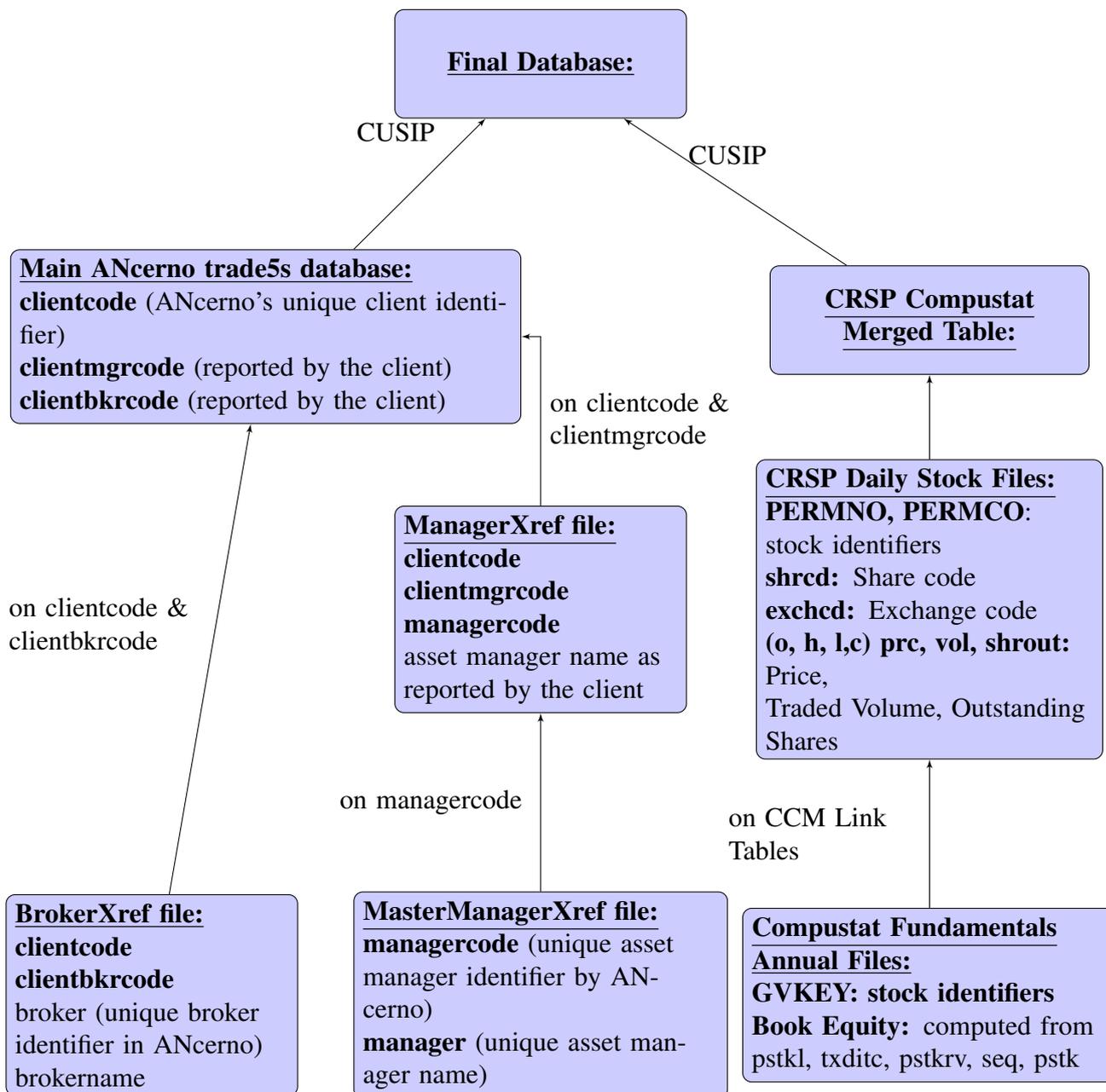
We reproduce portfolio construction described in Fama and French, 2015 paper and Ken French’s webpage.

Size, Value: The Fama/French Size and Value anomalies are constructed using the 6 value-weighted portfolios formed on size and book-to-market. The portfolios, which are formed at the end of each June, are the intersection of 2 portfolios formed on size with respect to NYSE median market equity (Small vs Big) and 3 portfolios formed with respect to the 30% and 40% NYSE quantiles of book to market BE/ME (Value, Neutral and Growth).

$$\begin{aligned}
 SMB = & \frac{1}{2} (Small\ Value + Small\ Neutral + Small\ Growth) \\
 & - \frac{1}{2} (Big\ Value + Big\ Neutral + Big\ Growth)
 \end{aligned} \tag{1.6}$$

$$HML = \frac{1}{2} (Small\ Value + Big\ Value) - \frac{1}{2} (Small\ Growth + Big\ Growth) \tag{1.7}$$

Fig. 1.3. Merging referential, market data, fundamental data and trade level data together



Investment: The Fama/French investment anomaly is constructed using the 6 value-weighted portfolios formed on size and investment. The portfolios, which are formed at the end of each June, are the intersection of 2 portfolios formed on size with respect to NYSE median market equity (Small vs Big) and 3 portfolios formed with respect to the

Table 1.9. ANcerno Variables

The table describes the main ANcerno variables used to compute the results of this paper

Variables	Source File	Description
Client/ Family/ Broker identification Variables		
clientcode	Main dataset	unique client identifier by ANcerno
clientmgrcode	Main dataset	trading investment manager code as reported by the client
clientbkrcode	Main dataset	code of the broker executing the trade as reported by the client
managercode	ManagerXref	unique trading investment manager code attributed by ANcerno
manager	Master ManagerXref	unique investment manager name by ANcerno
broker	BrokerXref	unique numeric broker identifier by ANcerno
Order identification Variables		
cusip	Main dataset	stock cusip
stockey	Main dataset	ANcerno stock identifier
onumber	Main dataset	Ticket indentifier for a single stock, side and date
lognumber	Main dataset	ANcerno identifier for data source
odtOrderDate	Main dataset	Date where the broker receives the ticket
odtLastDate	Main dataset	Last date allowed to liquidate the order
ov	Main dataset	Ticket size
Side	Main dataset	buy or sell (1 = Buy; -1 = Sell)
Trade identification Variables		
tradedate	Main dataset	date of the trade
xdtX	Main dataset	Execution time
Price	Main dataset	price per share
Volume	Main dataset	number of traded shares
Commission USD	Main dataset	per trade commission in USD

30% and 40% NYSE quantiles of investemnt Inv (Conservative, Neutral and Aggressive).

$$\begin{aligned}
 CMA = & \frac{1}{2} (Small\ Conservative + Big\ Conservative) \\
 & - \frac{1}{2} (Small\ Aggressive + Big\ Aggressive)
 \end{aligned}
 \tag{1.8}$$

Profitability: The Fama/French profitability anomaly is constructed using the 6 value-weighted portfolios formed on size and operating profitability. The portfolios, which are formed at the end of each June, are the intersection of 2 portfolios formed on size with respect to NYSE median market equity (Small vs Big) and 3 portfolios formed

with respect to the 30% and 40% NYSE quantiles of operating profitability OP (Robust, Neutral and Weak).

$$RMW = \frac{1}{2} (Small\ Robust + Big\ Robust) - \frac{1}{2} (Small\ Weak + Big\ Weak) \quad (1.9)$$

Momentum: Fama-French implementation of momentum anomaly is based on the 6 value-weighted portfolios formed on size and prior (t-2- ζ t-13) returns. The portfolios, which are formed monthly, are the intersections of 2 portfolios formed on size (Small vs Big with respect to NYSE median market equity) and 3 portfolios formed on prior (t-2- ζ t-13) returns (High, Neutral, Low compared to NYSE stocks quantiles).

$$Mom = \frac{1}{2} (Small\ High + Big\ High) - \frac{1}{2} (Small\ Low + Big\ Low) \quad (1.10)$$

1.C Definitions / Equations

1.C.1 Portfolio's turnover

Portfolio turnover at the review date t is computed as follows :

$$Turnover_t = \sum_{s=1}^{Nsec} \left| \frac{AUM_t w_t^s - AUM_{t-1} w_{t-1}^s (1 + sign(w_{t-1}^s) r_{t-1,t}^s)}{AUM_t} \right| \quad (1.11)$$

$$AUM_t = AUM_{t-1} \sum_{s=1}^{Nsec} w_{t-1}^s (1 + sign(w_{t-1}^s) r_{t-1,t}^s) \quad (1.12)$$

$$Turnover_t = \sum_{i=1}^{Nsec} \left| w_t^s - \frac{w_{t-1}^s (1 + sign(w_{t-1}^s) r_{t-1,t}^s)}{\sum_{s=1}^N w_{t-1}^s (1 + sign(w_{t-1}^s) r_{t-1,t}^s)} \right| \quad (1.13)$$

Where AUM_t and AUM_{t-1} are the portfolio size respectively at review date t and $t - 1$. $Nsec$ is the number of securities composing the portfolio. w_t^s is the weight of the stock s after the review date t . $r_{t-1,t}^s$ is the return of stock s on the period between review date

$t - 1$ and t .

1.C.2 Garman Klass Volatility definition

Garman-Klass estimate of the volatility uses the open, high, low and close prices of the day. This estimate is robust and very close in practice to more sophisticated ones. The formula is given by:

$$\sigma_k^{\text{GK}}(d) = \sqrt{\frac{1}{N} \sum_{t=1}^N \frac{1}{2} \log \left(\frac{H_{d-t}^k}{L_{d-t}^k} \right)^2 - (2 \log(2) - 1) \log \left(\frac{C_{d-t}^k}{O_{d-t}^k} \right)^2} \quad (1.14)$$

where the indexation k refers to the stock. d to the calculation day. N is the length of the rolling window in day. In our case 252 trading days. $O_t^k, H_t^k, L_t^k, C_t^k$ are respectively the open, high, low, close prices of stock k at day t .

Chapter 2

Modelling Transaction Costs when Trades May Be Crowded: A Bayesian Network Using Partially Observable Orders Imbalance

About this chapter

The article present in this chapter has been published in the following book Machine Learning for Asset Management: New Developments and Financial Applications. The authors would like to thank the participants of the following conferences for their constructive comments and remarks: 2nd QFFE International Conference (Marseille, 2018), the 13th Financial Risks International Forum (Paris, 2020), International Conference on Fintech & Financial Data Science (University College Dublin UCD, Ireland, 2019) and the Fixed Income leaders Summit 2020

Abstract

Using a large database of US institutional investors' trades in the equity market, this paper explores the effect of simultaneous executions on trading costs. We design a Bayesian network modelling the inter-dependencies between investors' transaction costs, stock characteristics (bid-ask spread, turnover, and volatility), meta-order attributes (side and size of the trade), and market pressure during execution, measured by the net order flow imbalance of investors meta-orders. Unlike standard machine learning algorithms, Bayesian networks are able to account for explicit inter-dependencies between variables. They also prove to be robust to missing values, as they are able to restore their most probable value given the state of the world. Order flow imbalance being only partially observable (on a subset of trades or with a delay), we show how to design a Bayesian network to infer its distribution and how to use this information to estimate transaction costs. Our model provides better predictions than standard (OLS) models. The forecasting error is smaller and decreases with the investors' order size, as large orders are more informative on the aggregate order flow imbalance (R^2 increases out-of-sample from -0.17% to 2.39% for the smallest to the largest decile of order size). Finally, we show that the accuracy of transaction costs forecasts depends heavily on stock volatility, with a coefficient of 0.78.

Keywords: Trading Costs, Liquidity, Crowding, Bayesian Networks.

2.1 Introduction

Transaction costs became of primary importance after the financial crisis. On the one hand, investment banks turned to more standardized products, switching from a high margin, inventory driven business to a low margin, flow business, where transactions costs have to be minimized. On the other hand, the asset management industry concentrated (Haldane et al., 2014). A common practice has been to organize the execution of large orders around one well-structured dealing desk. In 2007, the first Markets in Financial

Instruments European directive (MiFID) introduced the concept of “best execution” as a new requirement for market participants. The European best practices, including among others execution reviews, transaction costs analysis, and adequate split of large orders, have spread overseas in this globalized industry.

In this paper we use a unique dataset of institutional investors trades: the ANcerno database, containing a large sample of asset managers meta-orders on the US markets (Angel et al., 2015, Pagano, 2008, Briere et al., 2019). While most other databases contain the meta-orders of only one asset manager, ANcerno records roughly 10% of total institutional investors activity and 8% of total daily traded volume. Because of this specificity, it is possible to estimate the “imbalance of meta-orders”, i.e. the aggregated net order flow traded by investors, each day on each stock. This variable plays a role of primary importance in the transaction costs (Capponi and Cont, 2019, Bucci et al., 2018). Transaction costs tend to be large when you trade in the same direction as your peers, while you can even have a price improvement (i.e. obtain an average price that is lower than your decision price) if you are almost alone in front of the majority of agents trading that day. Stated differently, you pay to consume liquidity when you are part of the crowd, executing in the same direction as the market, and you are rewarded to provide liquidity to the crowd when you are executing in the opposite direction of the market.

The specificity of this “imbalance” variable is that it cannot be observed by market participants in real-time. Brokers and market makers can have a broad view of the imbalance of their clients’ flows and can provide this information to the rest of market participants with a delay, while asset management dealing desks do only observe their own instructions. Therefore, the imbalance is a “*latent variable*” in the sense of Bayesian modelling. It is linked to some observable explanatory variables and it conditions the transaction costs at the same time. For instance: conditionally to the fact that the investor trades a buy meta-order (rather than a sell one), the imbalance is more likely to be large and positive. This dependence can be inferred using the Bayes’ rule.

In this paper, we show how to use a specific model belonging to the large toolbox provided by machine learning: the Bayesian network, adapted to this kind of conditioning, to predict transaction costs, taking into account market information and trade characteris-

tics. This class of models has been created in the golden age of machine learning (Jordan, 1998); it is also known as graphical models and has been recently used to model analysts predictions (Bew et al., 2018). Such models have two very interesting characteristics. First, they are able to handle missing data. Second, they can infer the distribution of latent variables given the knowledge of other ones. In our case, a model fitted on ANcerno data can be used to forecast transaction costs when the imbalance is no longer observable. In practice, our model could be fitted on data provided ex-post by brokers¹. Afterwards, given other explanatory variables and the observed transaction costs, a Bayesian network can infer the expected distribution of the imbalance on a given day. This is a natural feature of the Bayes' rule: once the joint distribution of a set of variables is known, it is possible to obtain the expected value of any subset of other variables given the observations.

The goal of this paper is to show how Bayesian networks can be used to model the relationship between transaction costs and stock characteristics (bid-ask spread, average turnover, and volatility), meta-order attributes (side and size of the trade), and market pressure (net order flow imbalance). This last variable will be considered as latent because it is only partially observable by investors (typically with a delay, or in real-time but only on the investors' own trades). In practice, a possible way to implement our approach would probably be to implement a learning transfer: first, learn the graphical model on ANcerno or a similar database provided by brokers, then switch to a database in which the imbalance can not be observed.

We find that institutional investors' daily order flow imbalance is a good predictor of transaction costs. Interestingly, because investors' trading tends to be crowded in one direction, and given the fund manager's knowledge of its own meta-order, he can infer the aggregate order flow of the market that day, to better forecast his trading costs. Stated differently, a fund manager could update his beliefs on order flow imbalance distribution of the day, after observing his own trading decision (side and size of his order). We find that his estimation is more accurate when his executed meta-order is large. Besides, we

¹Brokers, exchanges, and custodians are selling the delayed information on the flows they saw the previous day or week. This Bayesian modelling approach is perfectly suited to this kind of partial information.

disclose evidence that a sell order is more informative on imbalance distribution than a buy order, probably because a crowded selling context is more informative about specific market conditions than a crowded buying context. We note that when an asset manager decides to sell a stock with a high participation rate, he could expect a "rushing towards the exit door" behaviour from his peers and assign a high probability for strong negative imbalance. Our finding confirms that the dominating variable for implementation shortfall forecast is indeed the order flow imbalance and not the order size. Moreover, the accuracy of transaction costs and market impact estimates are generally very low (Bacry et al., 2015). Practitioners have long suspected that the difficulty of estimating orders transaction costs is due to the variance of price innovations that is hardly predictable. Thanks to our Bayesian framework, we prove that this is true. The Bayesian network explicitly models the dependencies between the variance of the residuals and the rest of the network nodes. We find that the dominant variable is, indeed, the price volatility with coefficient 0.78, while other nodes' contribution to the variance is insignificant. This allows an investor to assess how confident he could be on each prediction given his meta-order and stock characteristics. Finally, we show that using partially observable order imbalance has value. The Bayesian network provides a better prediction of transaction costs after capturing the conditional dependencies between the nodes and the order flow imbalance than when this information is not used at all (R^2 increase out-of-sample from 0.38% to 0.50%). Besides, the estimates get more accurate when the order size is large (R^2 is 2.39% for the tenth decile of order size compared to -0.17% for the first decile). These results can explain the recent concentration of institutional investors executions on a few dealing desks. By executing the orders of a large and representative set of institutional investors, these dealing desks would have a better grasp of the aggregate order flow imbalance of the day. This information of paramount importance could then be either used for predicting the transaction cost more accurately or to design a better-optimized execution scheme taking the aggregated market pressure into account.

The structure of this paper is as follows: Section 2 reviews the existing literature on transaction costs modelling and Bayesian networks. Section 3 presents the data. Section 4 provides empirical evidence of the influence of investors' trade size and order imbalance

on transaction costs. Section 5 describes the Bayesian network method and its application to transaction costs modelling. Section 6 concludes.

2.2 Related literature

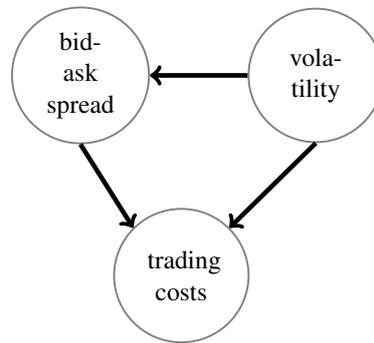
This paper takes place at the crossing of two fields: the transaction costs and market impact literature on the one hand, and Bayesian modelling on the other hand.

Transaction costs and market impact. Market impact attracted the attention of academics following two papers: economists have been initiated to this crucial concept by Kyle's theoretical paper (1985), while researchers in quantitative finance have been largely influenced by Almgren and Chriss, 2001 empirical results. Kyle, 1985 has shown how a market maker should strategically ask informed traders (i.e. asset managers) for a cost to compensate for the difficulty to assess the adverse selection she is exposed to in a noisy environment. This is typically what we observe empirically. Asset managers have to pay for liquidity demand while they can be rewarded for liquidity provision. Other market participants react to the aggregate offer or demand. This aggregate is exactly what we define as *the imbalance of meta-orders for a given day*. Kyle's essential result is that given a linear market-maker pricing rule and within a Gaussian framework, the transaction costs paid by the aggregation of investors are linear in the size of the aggregated meta-order. *Kyle's lambda*, measuring the sensitivity of price impact with respect to volume flow, is a traditional measure of liquidity. This theoretical framework has been sophisticated recently, extending Kyle's game theoretical framework to continuous time, non-Gaussian behaviors, and allowing risk aversion in market makers' strategy (Cetin and Rogers, 2007). It is now understood that the informed trader's optimal strategy is to try to hide its meta-order in the noise, while the market maker has to slowly digest orders flow to try to extract the information it contains and ask for the corresponding price. However, the resulting market impact is not necessarily linear. Empirical studies that followed showed that in practice market impact is more square root than linear in the size of the order (Collins and Fabozzi, 1991, Bouchard et al., 2011 or Robert et al., 2012).

Almgren and Chriss, 2001 seminal paper showed how to split an order optimally to minimize execution cost, making the assumption of concave transient market impact. Bouchard et al., 2011 derived an optimal control scheme to mitigate this cost for large meta-orders. This literature is of primary importance since it answers the regulatory requirements around “proof of best execution” and provides a baseline framework to asset managers and investment banks to improve their best practices and metrics for execution. With the popularity of factor investing, the specific question of the implementation costs of investment strategies following an index or a systematic active strategy has been raised by regulators and market participants. Frazzini et al., 2012, Novy-Marx and Velikov, 2015, or Briere et al., 2019 are attempting to answer the question of the potential maximum capacity of a trading strategy, by modelling transaction costs for large order sizes and estimating the break-even capacity of factor-driven investment strategies.

Bayesian networks. Machine learning is an extension of statistical learning, born with the seminal paper of Vapnik and Chervonenkis, 1971. Following the universal approximation theorem for non-linear Perceptrons (a specific class of neural networks) with at least one hidden layer (Hornik et al., 1989), statisticians and mathematicians started investigating approximation schemes based on the minimization of a possibly non-convex loss function, generally using stochastic gradient descent (Amari, 1993) to reach the global minimum while having good chances to escape from the local minima. Successes in Bayesian statistics focused on coupling a prior and a posterior distribution via the concept of the conjugate (Vila et al., 2000), opened the door to a mix of neural networks and Bayesian statistics, based on maximum likelihood estimations. Bayesian networks were born (see the seminal paper by Pearl, 1986). Bayesian networks are convenient tools for modelling large multivariate probability models and for making inferences. A Bayesian network combines observable explanatory variables with hidden latent variables in an intuitive, graphical representation.

In terms of applications, Bayesian networks have first been used for medical diagnosis, since they have been perceived as a natural extension of *expert systems*. Expert systems emerged with the first wave of artificial intelligence tools: Deterministic decision

Fig. 2.1. A simple graphical model for trading costs modelling

trees. Adding some probabilistic properties to these trees and reshaping them into graphs is another way to see the emergence of Bayesian networks. These models have also been used with success in troubleshooting of computed components, from printers (Skaanning et al., 2000) to computer networks (Lauritzen, 2003). They played an important role in the automation of problem-solving for computers related questions. Recently, they have been applied in finance. Bew et al., 2018 use Bayesian networks to combine analysts' recommendations to improve asset management decisions.

These models can very naturally capture the joint distribution of different variables, specified via a graphical model where nodes represent variables and arrows model the probabilistic dependencies. The very simple example of Figure 2.1 specifies that the stock bid-ask spread and its volatility both influence trading costs, while at the same time, the stock volatility has an influence on the bid-ask spread (Laruelle and Lehalle, 2018). The translation in a probabilistic language of this graph is the following. The *trading costs* TC , follows a law \mathcal{L} which parameters Θ_{TC} are functions of the *bid-ask spread* ψ and of the *volatility* σ : $TC \sim \mathcal{L}(\Theta_{TC}(\psi, \sigma))$. The parameters of the law of the *bid-ask spread* are seen as a random variable, itself a function of the volatility: $\psi \sim \mathcal{L}(\Theta_{\psi}(\sigma))$.

More details on the mechanisms of Bayesian networks are given in Section 5. At this stage, it is enough to say that *latent variables* can be added to the graph. An intermediate variable that is not always observable, but acting as a probabilistic intermediary (i.e. a conditioning variable) between observed variables, is enough to structure a Bayesian model. In the simple example of Figure 2.1, we can observe or not the bid-ask spread. When it is not observed, the Bayesian network will use its law $\mathcal{L}(\Theta_{\psi}(\sigma))$ to infer its

most probable value, conditionally to the observed volatility. To do that, the model uses Bayes' conditional probability chain rule. In our analysis, we always observe the bid-ask spread, but the net order flow imbalance of institutional investors' meta-orders is usually not known. This paper proposes a Bayesian network to model and forecast transaction costs with a graphical model where the imbalance of institutional meta-orders is a latent variable.

To sum up, our paper makes use of Bayesian networks to model the expected transaction costs of institutional investors as a function of the characteristics of the meta-order (essentially its size and direction), the market environment (stock volatility, bid-ask spread, and order flow imbalance). We contribute to the current literature on trading costs estimation by proposing a methodology to account for latent variables, in our case, order flow imbalance. This variable can only be partially observed with a delay or on a subset of all trades, but it is essential to structure the model. Our model has numerous potential applications and could be used to forecast trading costs, estimate the capacity of a strategy, or decide on the optimal trading execution.

2.3 ANcerno database

We obtain institutional trading data for the period from January 1st 2010 to September 30th 2011 from ANcerno Ltd. ANcerno, formerly Abel Noser Corporation is one of the leading consulting companies in providing Transaction Cost Analysis (TCA) in the US. It provides equity trading costs analysis for more than 500 global institutional investors, including pension funds, insurance companies, and asset managers. This database was largely used by academics to investigate institutional investors trading behaviour (see for example Anand et al., 2011, Puckett and Yan, 2011 and Eisele et al., 2017). ANcerno clients send their equity trades in order to monitor their execution quality. ANcerno systematically reports all equity trades it receives. Therefore, costs estimated on ANcerno are representative of what is effectively paid by institutional investors. Besides, previous researches have shown that ANcerno is free from any survivorship or backfill bias (see Puckett and Yan, 2011), constitute approximately 8% of the total CRSP daily dollar

volume (Anand et al., 2013), and 10% of total institutional activity (Puckett and Yan, 2011).

Hence, in our study, we use trade-level data from ANcerno on the historical composition of S&P 500 index. For each execution, ANcerno reports information on the CUSIP and ticker of the stock, the execution time at minute precision, the execution date, execution price, side (i.e., buy or sell), number of shares traded, commissions paid, whether the trade is part of a larger order, and several trade-level benchmarks to evaluate the quality of the execution. In our sample, we have execution data of 285 institutions (i.e., ANcerno clients). They could be either an individual mutual fund, a group of funds, or a fund manager subscribing to Abel Noser's analytical service. Each institution has one or several accounts. In our sample, we successfully track the activity of almost 44 thousand accounts, responsible for 3.9 trillion dollars of transactions, and using the service of 680 different brokerage firms. Compared to the market volume reported in CRSP, ANcerno accounts for an average of 4.5% over the whole period. The traded amount reported in ANcerno is over a trillion dollars every year and is, therefore, large enough to be relevant. We complement ANcerno database with daily bid-ask spread obtained from Reuters Tick History (RTH).

Consistent with machine learning best practices, we split our sample into a training set accounting for 70% of the meta-orders and a testing set accounting for the remaining 30%. The training-set is chosen randomly from meta-orders in our sample such as the number of buy orders and sell orders are equal. This procedure is very important for our study in order to estimate a non-biased net order flow imbalance. In the case of an unbalanced number of buy and sell orders in the training-set, the prior distribution of order flow imbalance will be artificially skewed toward positive values if the number of buy orders is higher or toward negative values otherwise. The training set is used to compute the results of sections 4 and 5, while the testing set is used for the out-of-sample predictions in section 6.

2.4 Transaction cost modelling

We measure trading costs with the traditional measure of implementation shortfall (Perold, 1988). This is the difference between a theoretical or benchmark price and the actual traded price effectively paid for the execution, in percent of the benchmark price. In our study, we define the reference price as the last visible price before the start of the execution (arrival price). The implementation shortfall measures the total amount of slippage a strategy might experience from its theoretical returns. In essence, our cost estimate measures how much of the theoretical returns of a strategy can be achieved in practice.

For a parent-ticket m of size $Q_k(m)$ split into N_{trades} child tickets ² of size $v_{k,m}(i)$ executed at date d in the direction $s_k(m)$, the implementation shortfall is calculated as follows:

$$\text{IS}_k(m, d) = \frac{s_k(m)}{P_k(0)} \left(\sum_{i=1}^{N_{\text{trades}}} \frac{v_{k,m}(i)}{Q_k(m)} \times P_k(i) - P_k^{\text{ref}} \right) \quad (2.1)$$

where $P_k^{\text{ref}} = P_k(0)$ is the reference price (in our case, the arrival price as provided by ANcerno). In this section, we investigate the effect of order size and order flow imbalance on the implementation shortfall of investors transactions.

2.4.1 Order size

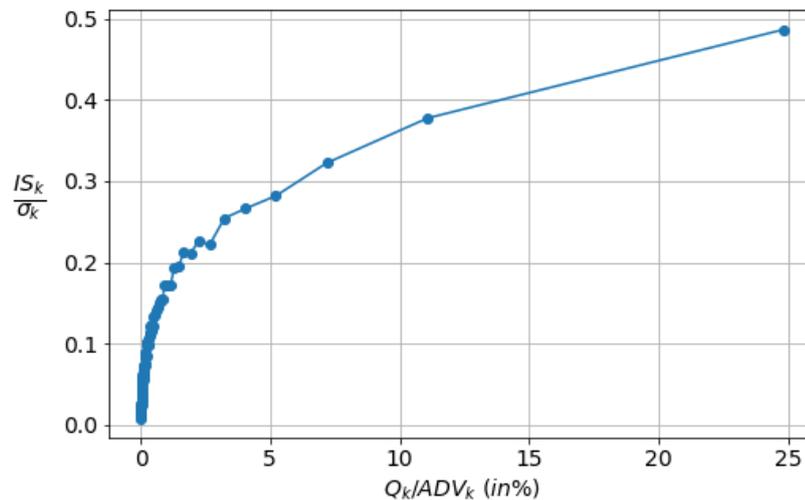
Kyle, 1985 introduced the concept that trades by a market participant may have an impact on the market price. Market impact is a direct consequence of the order size effect. A large meta-order may move the price in an unfavourable direction for the trader, resulting in a higher implementation shortfall. The execution cost is then increasing with order size. A series of empirical studies followed Kyle's theoretical work to confirm the existence of order size effect (Torre and Ferrari, 1999, Moro et al., 2009, Gomes and Waelbroeck, 2015, Bacry et al., 2015, Briere et al., 2019). To illustrate this effect, we regroup ANcerno tickets in 100 bins based on participation rate Q/ADV and plot in Figure 2.2 the average implementation shortfall scaled by the price volatility of the tickets

²Orders in ANcerno (parent tickets) are split within the execution period into smaller orders (child tickets). For each child ticket, ANcerno reports the executed volume, the price and time of execution.

in each bin. The scaling with stock's price volatility makes estimates of implementation shortfall comparable through time and across the universe of stocks. Otherwise, we can not compute the average on each bin as the effect of the participation rate is not the same for large bid-ask spread stocks as those with small bid-ask spreads. ANcerno tickets show a concave relation between the implementation shortfall and order size relative to daily traded volume. We observe a sharp increase in the costs from 0 to 0.2 points of price volatility when order size increases from 0.01% to 2% of the average daily volume. The slope decays afterward. For instance, a ticket with 14% participation rate, costs on average 0.4 points of volatility. A power-law function captures well the dependence of orders trading costs to order size.

Fig. 2.2. Order size effect on trading costs

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1, 2010 to September 30, 2011. We split our sample into 100 bins based on meta-order participation rate $Q_k(m)/ADV_k(d)$ and plots the average implementation shortfall scaled by stock's volatility IS_k/σ_k for each bin (blue dots)



2.4.2 Order flow imbalance

While most trading cost models emphasize on the historical dependence of market impact on stock liquidity and order size, It is only recently that order flow imbalance

has been recognized as a significant factor in explaining the magnitude of orders transaction costs. Using ANcerno database, Capponi and Cont, 2019, compared the explanatory power of order size to the effect of a proxy of market pressure "Order Flow Imbalance" on transaction costs and came to the conclusion that investors should focus on modelling the aggregate dynamics of market pressure during execution period, rather than focusing on optimizing market impact at a trade-by-trade level. Moreover, market pressure is contributed by all market participants at the trading session. But the traders who are responsible for executing institutional investors' orders contribute the most to this pressure and should be specifically taken into account in price movement forecast and transaction costs modelling. These market participants have the same profile as the informed/insider trader introduced by Kyle in 1985. By the end of the trading session, the private information, that was once detained by the insider, spread to the market and get incorporated into the price level. Bucci et al., 2018 argue that price market impact is a function of the aggregate net volume, that for shared indiscriminately between all market participants. Consequently, a small-sized order would cost nearly the same implementation shortfall as a much larger order if executed in the same direction during the same time frame.

We introduce the Net Order Flow Imbalance, to investigate the impact of institutional investors synchronous trading on the implementation shortfall. For a meta-order m executed at date d , the net investors order flow imbalance is defined as the ratio of net volume executed by the other investors at day d over their total traded volume:

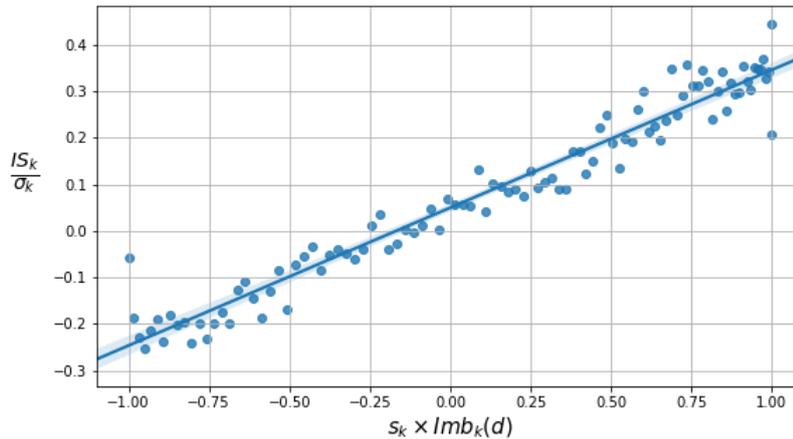
$$\text{Imb}_k(m, d) = \frac{\sum_{m' \neq m} Q_k(m', d) \cdot s_k(m', d)}{\sum_{m' \neq m} Q_k(m', d)} \quad (2.2)$$

Where k designs the stock, $s_k(m', d)$ is the side of the meta-order m' (i.e. 1 for buy orders and -1 for sell orders) and $Q_k(m', d)$ its size.

Figure 2.3 illustrates the dependence of the implementation shortfall to institutional investors trading imbalance. First, we note that the relationship is linear. The stronger the absolute imbalance, the higher the absolute value of price deviation during the execution. But depending on whether the trade is in the same direction as the net order flow imbalance, thus contributes to the existing market pressure, or on the opposite side,

Fig. 2.3. Net order flow imbalance effect on trading costs

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1, 2010 to September 30, 2011. We split our sample into 100 bins based on net order flow imbalance multiplied by the side of the trade and plots the average implementation shortfall scaled by stock's volatility IS_k/σ_k for each bin (blue dots)



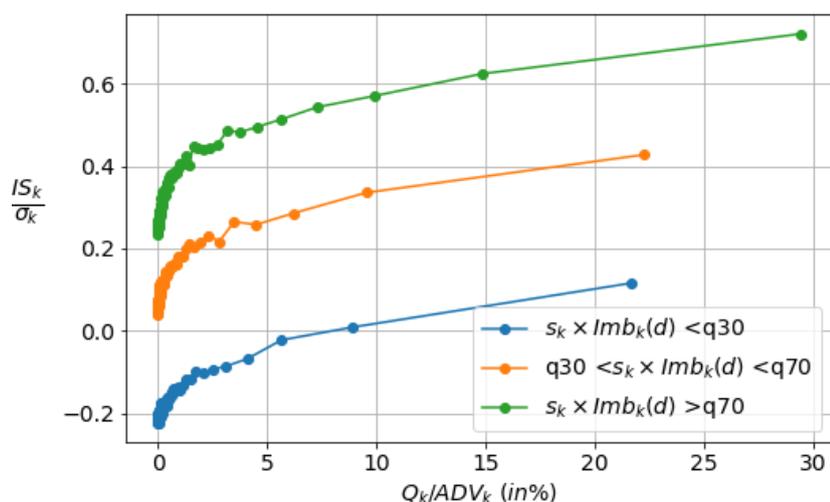
and provides liquidity to the market, one could expect either to pay a significant trading cost up to 0.4 points of price volatility when investors are trading synchronously toward the same directions ($Imb_k(m, d) = 1$) or benefit from a price improvement of 0.3 points of volatility when the trader is almost alone in front of his competitors' aggregate flow ($Imb_k(m, d) = -1$). Also worth noting that the implementation shortfall at zero imbalance is slightly positive. At neutral market pressure, the investors pay a positive transaction cost depending on stock traded and meta-order size.

2.4.3 Joint effect of order size and order flow imbalance

The results in subsection 2.4.1 and 2.4.2 show that the implementation shortfall depends on both the size of the executed order and market pressure during the execution period. Market pressure being approximated by investors net order flow imbalance. To disentangle the two effects, we split our sample on 3 distinct buckets with respect to meta-order signed imbalance ($s_k(m) \cdot Imb_k(m, d)$) 30% and 70% quantiles. Within each bucket, we sort meta-orders into 100 bins based on meta-order participation rate

Fig. 2.4. Joint effect of order size and net order flow imbalance on trading costs

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1, 2010 to September 30, 2011. First, We split our sample on 3 buckets w.r.t meta-order signed imbalance ($s_k(m) \cdot \text{Imb}_k(m, d)$) 30% and 70% quantiles. We sort meta-orders within each bucket into 100 bins based on meta-order participation rate ($Q_k(m)/\text{ADV}_k(d)$) and plots the average implementation shortfall scaled by stock's volatility IS_k/σ_k for each of the bins.



($Q_k(m)/\text{ADV}_k(d)$) and compute the average implementation shortfall scaled by stock's volatility for each of the bins. Figure 2.4 plots the result, where the blue, line shows order size effect for meta-orders executed against high market pressure (signed imbalance is lower than the 30% quantile). The orange line illustrates the effect for meta-orders executed under standard market pressure (signed imbalance between the 30% and 70% quantiles). Whereas the green line shows the result for orders executed in the same direction as the market (signed imbalance larger than the 70% quantile). We observe the impact of meta-order size is persistent in the 3 buckets and the power-law remains valid even after conditioning on net order flow imbalance. The linear effect of the signed imbalance is visible in the difference of transaction cost level between the 3 buckets. This proves that these two explanatory factors do not cancel one another. We also note that most meta-orders executed against investors' net order flow benefit from a price improvement between the moment the execution starts and the moment it ends. During strongly

unbalanced markets, the provider of liquidity is rewarded with a better execution price. However, for larger meta-orders ($Q_k(m)/ADV_k(d) = 23\%$) the market impact of the trade prevail and the trader pays on average a positive transaction cost. The opposite is also true, when traders seek liquidity in the same direction as the remainder of institutional investors, the trading cost gets more expensive than usual.

To further explore the joint effect of order size and net order flow imbalance on the implementation shortfall, we run the following step-wise multivariate regression. First, we perform the regression of order implementation shortfall on stock bid-ask spread and the square root of the order participation rate scale by stock volatility as described in equation (2.3). Then, a regression of the implementation shortfall on the bid-ask spread and the signed imbalance, also scaled by stock volatility (equation 2.4). Finally, we gather the 3 factors on the same regression as in equation (2.5).

$$IS_k(m, d) = \alpha \psi_k(d) + \beta \sigma_k^{GK}(d) \sqrt{\frac{Q_k(m)}{ADV_k(d)}} + \varepsilon_k(m, d) \quad (2.3)$$

$$IS_k(m, d) = \alpha \psi_k(d) + \gamma \sigma_k^{GK}(d) s_k(m) \text{Imb}_k(m, d) + \varepsilon_k(m, d) \quad (2.4)$$

$$IS_k(m, d) = \alpha \psi_k(d) + \beta \sigma_k^{GK}(d) \sqrt{\frac{Q_k(m)}{ADV_k(d)}} + \gamma \sigma_k^{GK}(d) s_k(m) \text{Imb}_k(m, d) + \varepsilon_k(m, d) \quad (2.5)$$

where $IS_k(m, d)$ is the implementation shortfall of meta-order m submitted on stock k at day d . $\psi_k(d)$ is the quoted intraday bid-ask spread of stock k averaged on the month. $\sigma_k^{GK}(d)$ is the Garman and Klass, 1980 intraday volatility of stock k estimated on a 12 month rolling window. $Q_k(m)$ and $s_k(m)$ are respectively size and side (Buy/Sell) of the order. $ADV_k(d)$ is the daily traded volume averaged on a 12 months rolling window, and $Q_k(m)/ADV_k(d)$ is the participation rate. $\text{Imb}_k(m, d)$ is the net investors order flow imbalance estimate for order m at day d . Finally, α , β and γ are model parameters and $\varepsilon_k(m, d)$ is the respective error term.

Table 2.1. Transaction cost model

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1, 2010 to September 30, 2011 on the S&P 500 historical components. $\psi_k(d)$ is the quoted intraday bid-ask spread of stock k averaged on the month, obtained from RTH database. $\sigma_k^{\text{GK}}(d)$ and $\text{ADV}_k(d)$ are respectively the Garman Klass intraday volatility and the average daily volume of stock k estimated on a 12 month rolling window. $Q_k(m)$ and $s_k(m)$ are respectively size and side (Buy/Sell) of the order. $\text{Imb}_k(m, d)$ is the net order flow imbalance for order m at day d .

Model	Dependent variable: $\text{IS}_k(m, d)$		
$\psi_k(d)$	0.399*** (0.032)	0.708*** (0.028)	0.180*** (0.032)
$\sigma_k^{\text{GK}}(d) \sqrt{Q_k(m)/\text{ADV}_k(d)}$	0.951*** (0.021)		0.712*** (0.021)
$\sigma_k^{\text{GK}}(d) s_k(m) \text{Imb}_k(m, d)$		0.234*** (0.002)	0.224*** (0.002)
Observations	7421548	7421548	7421548
R ²	0.005	0.016	0.017
Adjusted R ²	0.005	0.016	0.017
Residual Std. Error	0.017	0.017	0.017
F Statistic	1892.157	5993.682	4391.073
AIC	-3964520	-3972637	-3973801
Note:	*p < 0.1; **p < 0.05; ***p < 0.01		

The results of these three regressions are presented in Table 2.1. In the first regression the coefficient of bid-ask spread and order size term ($\sigma_k^{\text{GK}}(d) \sqrt{Q_k(m)/\text{ADV}_k(d)}$) are respectively 0.4 and 0.95, both statistically significant at the 1% level. Consistently with Briere et al., 2019, we find that for small orders, institutional investors pay only 0.4 times the bid-ask spread. In the second regression, we replace the order size term with the market pressure term. We notice that the coefficient of the bid-ask spread increases (from 0.4 to 0.7) and its confidence interval becomes tighter (lower standard deviation 0.028 vs 0.032). The determination coefficient for the second regression is also much higher (1.6% vs 0.5%). Finally, when we put all explanatory variables together, we find that the coefficient of the order imbalance does not change (0.22-0.23) while both order size term and bid-ask spread have much lower parameters (0.18 for the bid-ask spread and 0.71 for the

order size term) compared to the first two models. Besides, the determination coefficient of the second and third regressions are comparable. The net order flow imbalance seems to be a much better predictor of expected implementation shortfall than the size of the order. Although, all coefficients are statistically significant at the 1% level.

2.5 Bayesian network modelling with net order flow imbalance as latent variable

Institutional investors' net order flow imbalance is a key factor in the estimation of meta-orders transaction costs. However, this variable is only observable with a delay, for example through brokers or custodians' reports, or on a subset of trades only (the investor's own trades). Thus, it can not be used for production purposes. To remedy this issue, we propose a Bayesian network to incorporate all information we could get before the execution of the meta-order, and update our beliefs on the probabilistic distribution of the latent variable. We then use the most probable value of the net order flow imbalance to estimate the meta-order transaction cost. One of the interesting features of Bayesian networks is that they can be explored in both directions, thanks to the Bayes' rule. Therefore, we can give an estimate of the latent variables, by probabilistic inference, before and after the variable of interest is observed. In the context of this study, this means that:

- given the characteristics of the meta-order (side and size of the trade) and stock attributes (bid-ask spread, average daily traded volume and price volatility), we can compute a first estimate of the imbalance and forecast the transaction costs that should be paid by the investor.
- Once we get the effectively paid trading cost, we can recover a more accurate estimate of the distribution of investors' net order flow of the day, and for example, incorporate it in the estimation of order flow imbalance of the following day.

2.5.1 Bayesian inference

The main difference between the frequentist approach and the Bayesian approach is that in the latter, the parameters of the models are no longer unknown constants that need to be estimated, but random variables which parameters have to be estimated. The statistician has the possibility to incorporate his prior belief on the probabilistic distribution of the variable and update his belief step by step as soon as new data becomes available. From one step to the other: the former "posterior distribution" is used as a prior for the next step estimate.

For instance, if y stands for the unknown random variable, x for the observed data, and $\mathbf{P}(y)$ for the prior. The posterior distribution $\mathbf{P}(y|x)$ is obtained as the multiplication of the prior $\mathbf{P}(y)$ with the likelihood $\mathbf{P}(x|y)$ of observing the data, scaled by $\mathbf{P}(x)$. The definition of conditional probabilities applied on this procedure reads:

$$\mathbf{P}(y|x) = \frac{\mathbf{P}(y)\mathbf{P}(x|y)}{\mathbf{P}(x)} \propto \mathbf{P}(y)\mathbf{P}(x|y). \quad (2.6)$$

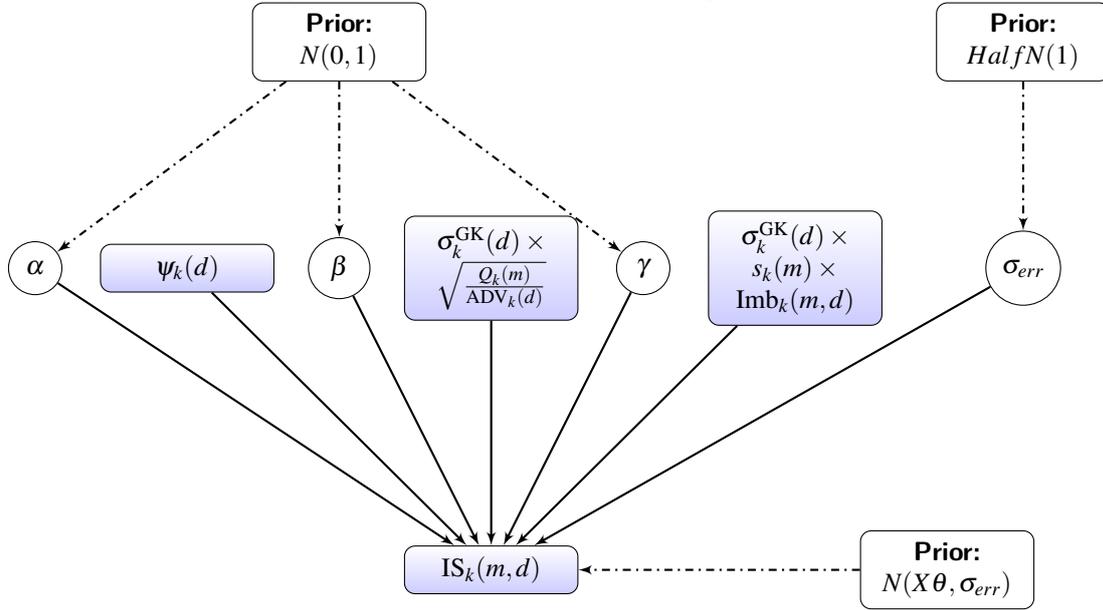
Figure 2.5 shows how to estimate the coefficients of the Bayesian linear regression specified in equation (2.5). First, we start by incorporating our prior beliefs, if any, on the distribution of the parameters $\theta = (\alpha, \beta, \gamma)^T$. Without any belief, a good choice is to take a non-informative prior like the normal distribution $N(0, 1)$. The best initialization for priors is hence a law close to the empirical repartition function of the considered variable. The variable of interest $IS_k(m, d)$ follows a normal distribution centered at the estimated value $\hat{y} = X\theta$ and has variance σ_{err}^2 of the error term $\varepsilon_k(m, d)$. σ_{err}^2 requires a non negative prior distribution, such as the positive part of a Gaussian (i.e. *HalfNormal*) or the positive part of a Cauchy (i.e. *HalfCauchy*). The Bayesian setup gives a direct interpretation of the results: The mean of the posterior distribution is the most probable value of the parameters θ , and the 5% confidence interval is limited by the 2.5% and 97.5% quantiles of the posterior distribution.

MCMC (Markov Chain Monte Carlo, see Hastings, 1970 for one of the first references) methods offers an easy way to sample from the posterior, especially when the posterior does not obey a well-known expression or when we know the expression has

Fig. 2.5. Bayesian inference of a linear regression

Blue rectangle represent observed variables. Circles are the parameters that need to be calibrated. Each have a prior distributions detailed in white rectangle.

$$X\theta = \alpha \psi_k(d) + \beta \sigma_k^{\text{GK}}(d) \sqrt{Q_k(m)/\text{ADV}_k(d)} + \gamma \sigma_k^{\text{GK}}(d) \cdot s_k(m) \cdot \text{Imb}_k(m, d)$$



a multiplicative term. It is very convenient for the Bayesian approach, where the posterior distribution is proportional to the multiplication of the prior and the likelihood. MCMC algorithms make computations tractable for parametric models. The intuition behind MCMC is to define a Markov Chain (x_0, x_1, \dots) on the support of x , such that when the size n of this chain goes to infinity, the new drawn point x_n is distributed accordingly to the law \mathbf{P}_x . The most famous algorithm to generate Markov Chains having this very nice property is the Hasting-Metropolis one, explained in Appendix F, that we use in this study, and the Gibbs sampler³. The marginal distribution of regression coefficients of the calibrated model is shown in the right panel of Table 2.2, while the result of the OLS regression is in the left panel. As expected from Bayesian models when the sample size is large, we end up with the same results. Beside, when the priors are Gaussian, the maximum a posteriori of the parameters is equivalent to a ridge estimate with a quadratic regularization $(\mathbb{E}_{\theta|X,Y} [\theta] = \arg \max_{\theta} \mathbf{P}(\theta|X, Y) = \arg \min_{\theta} \|Y - X\theta\|^2 + \sigma_{err}^2 \|\theta\|^2)$. This formula, similar to the one of Ridge regression (see Hoerl and Kennard, 1970), makes

³We use the PyMC3 python package implementation of Hasting-Metropolis algorithm described in Salvatier, Fongesbeck, et al., 2016 with a large number of iterations $N_{iter} = 10000$

the Bayesian regression more robust to outliers than OLS. It is the case for example for the order size term $\sigma_k^{\text{GK}}(d)\sqrt{Q_k(m)/\text{ADV}_k(d)}$ distribution, which explain the minor difference in coefficient estimate (0.71 vs 0.69). Nevertheless both the OLS and Bayesian regressions give the same economic and statistical conclusions despite having different statistical assumptions.

Table 2.2. OLS regression vs Bayesian regression

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1, 2010 to September 30, 2011 on the S&P 500 historical components. $\psi_k(d)$ is the quoted intraday bid-ask spread of stock k averaged on the month, obtained from RTH database. $\sigma_k^{\text{GK}}(d)$ and $\text{ADV}_k(d)$ are respectively the Garman Klass intraday volatility and the average daily volume of stock k estimated on a 12 month rolling window. $Q_k(m)$ and $s_k(m)$ are respectively size and side (Buy/Sell) of the order. $\text{Imb}_k(m, d)$ is the net order flow imbalance for order m at day d .

	OLS Regression				Bayesian Regression			
	coef	std err	Q 2.5%	Q 97.5%	coef	std err	Q 2.5%	Q 97.5%
$\psi_k(d)$	0.18	0.03	0.12	0.24	0.18	0.03	0.12	0.24
$\sigma_k^{\text{GK}}(d)\sqrt{\frac{Q_k(m)}{\text{ADV}_k(d)}}$	0.71	0.02	0.67	0.75	0.69	0.02	0.65	0.73
$\sigma_k^{\text{GK}}(d)s_k(m)\text{Imb}_k(m, d)$	0.22	0.00	0.22	0.23	0.22	0.00	0.22	0.23
RMSE (%)	1.66				1.66			
R ² (%)	1.77				1.77			

2.5.2 Bayesian network modelling

Most of the OLS assumptions are violated. As shown in Appendix E, the marginal distribution of trading costs has a peaky shape, with fat tails (excess-kurtosis of 23.46). The assumption of homoscedasticity is also violated. The variance of the error term is hardly constant across orders. Forecasting errors are smaller for small orders (implemented in a few minutes) compared to large ones (split over days) that got exposed for a longer period to market volatility. Finally, it is difficult to assume that the observations are independent of one on-other. Meta-orders on the same stock, whatever the execution day, share some common variance related to the stock characteristics. Similarly, orders executed at the same trading session on different stocks face the same market conditions,

and thus cannot be considered independent of one another.

In addition, Bayesian Networks have the advantage of not relying on Normal error distributions (Zuo and Kita, 2012), as do most other machine learning algorithms. Furthermore, Bayesian networks have the advantage of giving a human-readable description of dependencies between considered variables, whereas other more complex models, such as Neural Networks, suffer from being considered as "black box" models.

The structure of our Bayesian network

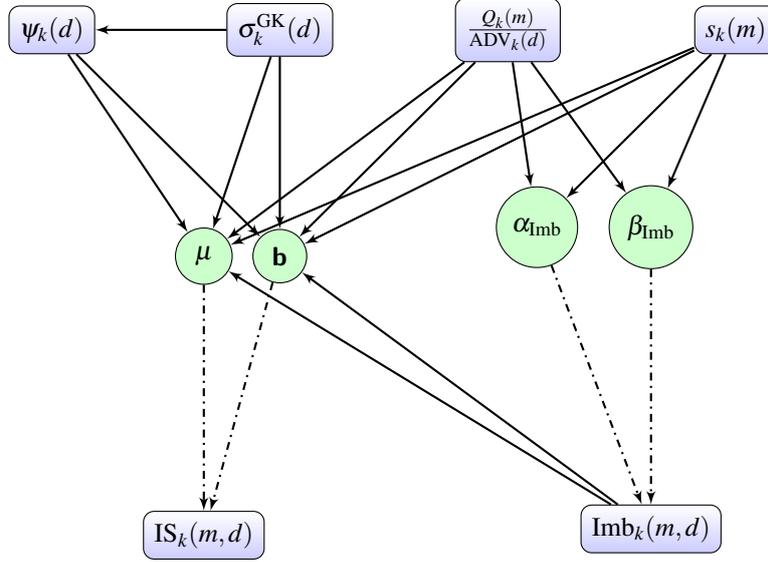
Our goal is to estimate the Implementation Shortfall of an order. We would like the Bayesian network takes into account:

- Attributes of the traded stock, such as average bid-ask spread, price volatility and average turnover;
- Characteristics of the meta-order, mainly order size and side of the trade (Buy/Sell);
- And the level of crowding during the execution: the net order flow of large institutional firms.

Figure 2.6 shows the Bayesian network we engineered. We distinguish 3 key dependencies. First, the bid-ask spread depends on the level of stock volatility. Second, the marginal probability distribution of order flow imbalance is a function of the meta-order size and side. Finally, the implementation shortfall is a function of all network nodes. In the following section, we detail the nature of these dependencies and we set the priors for each group of variable separately.

Bid-ask spread dependencies

The relation between stock volatility and the bid-ask spread is well documented. Theoretically, it is justified by Wyart et al., 2008 that, deriving the P&L of traders submitting market orders and those submitting limit orders, an equilibrium price is only achievable if the bid-ask spread is proportional to price volatility (i.e. $\psi_k(d) \propto \sigma_k^{\text{GK}}(d)$). In the

Fig. 2.6. Bayesian network for transaction costs modelling

same fashion, Dayri and Rosenbaum, 2015 study the optimal tick size, and find that the bid-ask spread that the market would prefer to pay if not constraint by the tick size verifies $\frac{\psi_k(d)}{2} \propto \frac{\sigma_k^{\text{GK}}(d)}{\sqrt{M}}$. The rational is that market makers, setting the best limits of the order book, accept to provide tight bid-ask spreads not only when the volatility (i.e. the risk of a given inventory level) is low, but also when they have more opportunities within the day to unwind their position. The relation between the bid-ask spread and the volatility is confirmed empirically on our data, as illustrated in Figure 2.11 of Appendix B.

Consistent with the literature of stochastic models for volatility, we set the prior of stocks volatility to a log normal distribution $\sigma_k^{\text{GK}}(d) \sim \text{LogNormal}$. Consequently, the bid-ask spread should follow a log normal distribution too, and the conditional probability of bid-ask spread given price volatility is detailed in equation (3.8), where $c^{\psi\sigma}, \rho^{\psi\sigma}, \sigma_{\psi,\sigma}$ are model parameters.

$$\psi_k(d) | \sigma_k^{\text{GK}}(d) \sim \text{LogNormal}(c^{\psi\sigma} + \rho^{\psi\sigma} \log(\sigma_k^{\text{GK}}(d)), \sigma_{\psi,\sigma}) \quad (2.7)$$

Table 2.3. Bayesian inference: Bid-Ask spread, volatility dependencies

The table summaries the posterior distribution of model parameters described in equation (2.7). $\mathbb{E}[X]$, $\text{std}(X)$, Q2.5% and Q97.5% are respectively the mean, the standard deviation, the 2.5% and 97.5% quantile of parameters posterior distribution. The results are obtained from Hasting-Metropolis sampler with $N_{iter} = 10000$ iterations (PyMC3 implementation). The institutional investors trading data are obtained from ANcerno Ltd on the period ranging from January 1st, 2010 to September 30th, 2011.

	$\mathbb{E}[X]$	Std[X]	Q2.5%	Q97.5%
$c_{\psi,\sigma}$	-4.137	0.006	-4.150	-4.126
$\rho_{\psi,\sigma}$	0.777	0.001	0.775	0.780
$\sigma_{\psi,\sigma}$	0.402	0.000	0.401	0.402

Net order flow imbalance dependencies

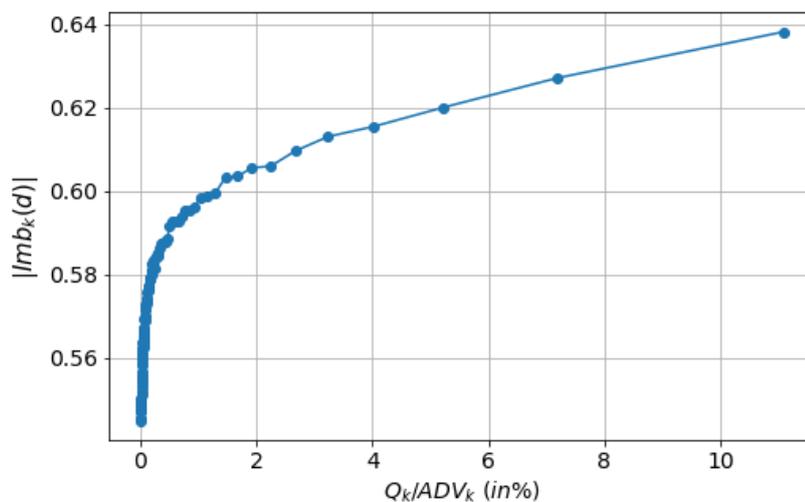
In this section, we quantify the dependence of net order flow imbalance on the remaining variables in the network. Figure 2.7 shows the marginal distribution of the imbalance depending on the sided of the meta-order. The U-shape of the plot confirms that institutional investors have indeed correlated executions, and tend to execute the same stocks in the same directions during the same periods, which intensifies the pressure on price movements. This correlation in trade execution can be explained by various factors. Asset managers compete for the same base of customers and can implement similar strategies (Greenwood and Thesmar, 2011, Koch et al., 2016). Thus, they face similar inflows and outflows, depending on liquidity needs and investment opportunities. Moreover, the asset management industry is subject to a series of regulatory constraints that can push funds to buy or sell simultaneously the same kind of assets. We note also, that the U-shape is decomposed in two skewed distributions depending on the side of the meta-order. So, given the side of the meta-order of an institutional investor, the remainder of large arbitrageurs executions constitute either a positively (for sell) or negatively (for buy) skewed imbalance distribution. Besides, conditional on the level of a meta-order participation rate, Figure 2.8 shows that the intensity of absolute net order flow imbalance of investors meta-orders gets stronger, a confirmation of institutional investors crowding.

Fig. 2.7. Net order flow imbalance distribution given meta-order side

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1, 2010 to September 30, 2011. Given a meta-order m submitted by an institutional investor, the figure plots the distribution of the net order flow imbalance generated by the remainder of investors as defined in equation (2.2) given the side $s_k(m)$

**Fig. 2.8. Net order flow imbalance as a function of participation rate**

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1, 2010 to September 30, 2011. We sort meta-orders into 100 bins based on meta-order participation rate ($Q_k(m)/ADV_k(d)$) and plots the average absolute net order flow imbalance for each of the bins



The Side takes two values Buy and Sell. It is modelled with a Bernoulli distribution $s_k(m) \sim \text{Bernoulli}(p_{side})$ of parameter $p_{side} = \frac{1}{2}$. The data shows that a Beta function is a good approximation of the U-shape for variables defined between $[0,1]$. After applying the linear transformation $x \rightarrow \frac{x+1}{2}$, $\text{Beta}(\alpha, \beta)$ is a plausible distribution of the transformed net order flow imbalance. Moreover, the beta distribution have the particularity to produce different shapes depending on the parameters α and β . It produces a symmetric U-shape when $\alpha = \beta$ and $(\alpha - 1)(\alpha - 2) > 0$, a positive skew when $\alpha < \beta$ and a negative skew when $\alpha > \beta$.

The probability density function of the transformed order flow imbalance PDF_{Beta} is given by:

$$\text{PDF}_{\text{Beta}}(\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\text{B}(\alpha, \beta)} \quad \text{where} \quad \text{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \quad (2.8)$$

$$\mathbf{P} \left(\text{Imb}_k(m, d) \middle| s_k(m), \frac{Q_k(m)}{\text{ADV}_k(d)} \right) = \text{B}(\alpha_{\text{Imb}}, \beta_{\text{Imb}}) \quad (2.9)$$

The dependence on the order side and the order participation rate should be taken into account at the level of the parameters of the Beta function (α, β) .

$$\alpha_{\text{Imb}} = c^\alpha + \rho_s^\alpha \cdot s_k(m) + \rho_p^\alpha \cdot s_k(m) \cdot \frac{Q_k(m)}{\text{ADV}_k(d)} \quad (2.10)$$

$$\beta_{\text{Imb}} = c^\beta + \rho_s^\beta \cdot s_k(m) + \rho_p^\beta \cdot s_k(m) \cdot \frac{Q_k(m)}{\text{ADV}_k(d)} \quad (2.11)$$

The result of the Bayesian inference of the imbalance dependencies is summarized in Table 2.4. When $s_k(m) = 0$ and $Q_k(m)/\text{ADV}_k(d) = 0$, the posterior distribution of net order flow imbalance is given by $\text{B}(0.67, 0.68)$ which produces a U-shape. This means that when the asset manager has no signal or information on price movement, he can only assume the synchronization of institutional activity. Thus, the symmetric distribution with higher probability at the extreme values of the imbalance. But once he detects a signal, since the process leading to generate this signal is independent of the execution process,

he can use his own meta-order as an observation to update his belief on the distribution of the expected market pressure. Note also that ρ_s^β and ρ_p^β are very low compared to ρ_s^α and ρ_p^α . This is not an issue, because what determines the strength of the skew for the Beta function is the difference $(\beta - \alpha)$ (see Appendix C for Beta function properties).

Table 2.4. Bayesian inference: Net order flow imbalance dependencies

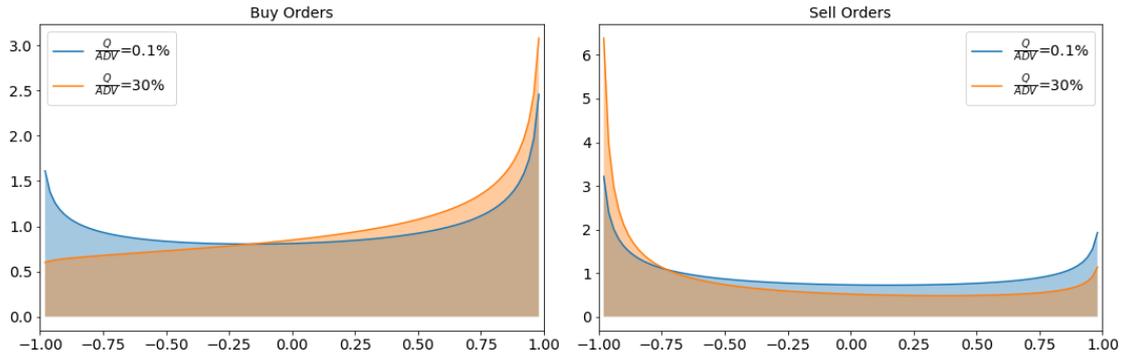
The table summaries the posterior distribution of model parameters described in equations (2.10) and (2.11). $\mathbb{E}[X]$, $\text{std}(X)$, Q2.5% and Q97.5% are respectively the mean, the standard deviation, the 2.5% and 97.5% quantile of parameters posterior distribution. The results are obtained from Hasting-Metropolis sampler with $N_{iter} = 10000$ iterations (PyMC3 implementation). The institutional investors trading data are obtained from ANcerno Ltd on the period ranging from January 1st, 2010 to September 30th, 2011.

	$\mathbb{E}[X]$	Std[X]	Q2.5%	Q97.5%
c^α	0.666	0.001	0.664	0.667
ρ_s^α	0.101	0.001	0.099	0.102
ρ_p^α	0.884	0.021	0.846	0.928
c^β	0.675	0.001	0.673	0.677
ρ_s^β	0.000	0.000	4.2e-08	0.000
ρ_p^β	0.001	0.001	1.4e-07	0.003

To better interpret the results of the table, we plot the posterior distribution of the net order flow imbalance, given two levels of participation rate (0.1% and 30%) for buy and sell trades. As expected, we observe that the side of the trade skews the distribution positively for a buy order and negatively for a sell order. The information of the meta-order participation rate intensifies the skew and increases the probability of having a full synchronization of investors executions $|\text{Imb}| = 1$. However, the shape of the distribution is not symmetrical between buy orders and sell orders. The skew of imbalance distribution is much stronger for sell orders. This means that when an investor is selling massively with large $\frac{Q}{\text{ADV}}$, he could expect a high selling pressure from the market due to investors synchronous inflows and outflow. Because institutional investors are natural buyers, implementing more long-only strategies than short selling ones, a high selling pressure corresponds to a "Rushing toward the exit door" situation. While on the opposite scenario, a buying order with a high participation rate although informative on the market does not give as much evidence on market participants' behaviour.

Fig. 2.9. Inferred net order flow imbalance given the side and the size of the meta-order

The figure shows the posterior distribution of net order flow imbalance given the meta-order characteristics. On the left panel we plots the distribution for buy orders with two levels of participation rate, blue line corresponds to small orders $Q_k(m)/ADV_k(d) = 0.1\%$ and orange line for large orders $Q_k(m)/ADV_k(d) = 30\%$. The right panel shows the result for sell orders with the same levels of participation rate



Implementation shortfall dependencies

Similarly, we model the implementation shortfall as a function of all the other nodes of the network. The data shows the historical distribution of implementation shortfall displays fat tails with pronounced non-Gaussian peaky shape. Thus, a double exponential (Laplace) probability density is a good prior of IS distribution. The probability density function (PDF) of Laplace is given by:

$$IS_k(m, d) \sim Laplace(\mu, b), \quad PDF_{Laplace}(x, \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right) \quad (2.12)$$

The location parameter μ is given by equation (2.13). As in the linear regression, we condition the magnitude of transaction cost to the stock bid-ask spread, the participation rate scaled by the volatility and investors order flow imbalance signed by meta-order side and scaled by stock volatility. Nevertheless, we don't assume the square root function for

meta-order size but a power law highlighted by the exponent γ that we estimate.

$$\begin{aligned} \mu = & a_0 + a_\psi \psi_k(d) + a_{\sigma, \frac{\rho}{\text{ADV}}} \exp \left\{ \log(\sigma_k^{\text{GK}}(d)) + \gamma \log \left(\frac{Q_k(m)}{\text{ADV}_k(d)} \right) \right\} \\ & + a_{s, \text{Imb}} \sigma_k^{\text{GK}}(d) s_k(m) \text{Imb}_k(m, d) \end{aligned} \quad (2.13)$$

Estimation accuracy is function of market condition, speed and duration of the execution algorithm, the aggressiveness in seeking liquidity. This heteroscedasticity of the implementation shortfall is taken into account by making the standard deviation of Laplace distribution b depend on stock attributes (spread and volatility), meta-order characteristic (participation rate) and market condition (absolute imbalance) as follows:

$$\begin{aligned} b_{ln} = & \log(b) \\ = & b_0 + b_\psi \log(\psi_k(d)) + b_\sigma \log(\sigma_k^{\text{GK}}(d)) + b_{\frac{\rho}{\text{ADV}}} \log \left(\frac{Q_k(m)}{\text{ADV}_k(d)} \right) \\ & + b_{\text{Imb}} \log(|\text{Imb}_k(m, d)|) \end{aligned} \quad (2.14)$$

Table 2.5. Bayesian inference: Implementation shortfall dependencies

The table summaries the posterior distribution of model parameters described in equations (2.13) and (2.14). $\mathbb{E}[X]$, $\text{std}(X)$, Q2.5% and Q97.5% are respectively the mean, the standard deviation, the 2.5% and 97.5% quantile of parameters posterior distribution. The results are obtained from Hasting-Metropolis sampler with $N_{iter} = 10000$ iterations (PyMC3 implementation). The institutional investors trading data are obtained from ANcerno Ltd on the period ranging from January 1st, 2010 to September 30th, 2011.

	$\mathbb{E}[X]$	Std[X]	Q2.5%	Q97.5%
a_0	0.00	0.00	-0.00	0.00
a_ψ	-0.60	0.37	-1.36	0.12
$a_{\sigma, \frac{\rho}{\text{ADV}}}$	0.67	0.24	0.26	1.15
γ	0.41	0.11	0.20	0.62
$a_{s, \text{Imb}}$	0.20	0.02	0.17	0.24
b_0	0.06	0.17	-0.27	0.40
b_ψ	0.09	0.03	0.05	0.14
b_σ	0.78	0.03	0.71	0.85
$b_{\frac{\rho}{\text{ADV}}}$	0.05	0.01	0.04	0.06
b_{Imb}	0.01	0.01	-0.01	0.03

Table 2.5 summarizes the first two moments and the 2.5% and 97.5% quantiles of the posterior distribution of the parameters. First, we note that the exponent of the order size

term is a bit lower than the square root $\hat{\gamma} = 0.41$, consistent with the previous finding of Bacry et al., 2015 that used a proprietary database of a broker execution in Europe. The literature usually document a power law with exponent between 0.4 and 0.5 (Gomes and Waelbroeck, 2015 and Briere et al., 2019). The parameters relative to the location term are consistent with the ones estimated with OLS regression. As expected, the intercept parameter is null, the coefficient of the order size and order flow imbalance terms are similar to those estimated by the OLS regression. Only the coefficient of bid-ask spread differs significantly. The parameters of the scale of Laplace distribution are small except the stock volatility coefficient. It proves that the main contributor to the heteroscedasticity is not the order size but stock volatility, consistent with the findings of Capponi and Cont, 2019 suggesting that conditionally to the level of stock volatility and execution duration, the order size have a small impact on transaction costs.

2.5.3 Forecasting implementation shortfall

We gather the different blocks of variable dependencies to constitute the Bayesian network of Figure 2.6. The parameters $(\mu, b, \alpha_{imb}, \beta_{imb})$ are estimated via Bayesian inference using Hasting-Metropolis algorithm. Once the network is calibrated on 70% of the meta-orders, we use it to infer the latent variable of net order flow imbalance given meta-order and stock characteristics and estimate orders implementation shortfall both in-sample (on the training set) and out-of-sample (on the testing set, the remaining 30% of the meta-orders not yet seen by the algorithm). Table 2.6 displays the results for both the linear regression and the Bayesian network predictions in- and out-of-sample. For the linear regression, we compare a model without order flow imbalance (equation (2.3), column 1) and one with order flow imbalance (equation (2.5), column 2). In this last model, the realized imbalance is fully observed in real-time, which is never achievable in practice but can serve as a benchmark case. We then show the results of three Bayesian networks: The first network (column 3) has never seen the information of the imbalance neither during the training phase nor the prediction phase. In that sense, it is comparable to the first linear regression (OLS when the imbalance is not available) in the first column

of Table 2.6. The second network (column 4) was trained with the information of order flow imbalance. Once this information is captured by the conditional probabilities of network edges, the network is exploited without the use of the imbalance. In this regard, order flow imbalance is partially observed. The last network (column 5) has full information on the imbalance, both at the training and testing phases, and is similar to the second OLS regression displayed in column 2. Adding information on imbalance improves the forecasting accuracy, both for the OLS regression and the Bayesian network. In-sample, it increases the R^2 from 0.52% to 1.77% for the two models, and reduces the forecasting error (RMSE from 1.67% to 1.66% and MAPE from 98.74% to 98.48%). For the same set of information (imbalance observable or not), the Bayesian network has the same accuracy as the OLS on the training set. However, the absolute average error is much smaller for the Bayesian network (-0.43 bps vs -0.86 when the imbalance is available, -1.41 bps vs -1.43 bps when it is not). In-sample, and when all explanatory variables are observable, the Bayesian network has only a limited advantage over simple linear regressions in terms of prediction accuracy. Out-of-sample, when the imbalance is not available (Panel B, columns 1 and 4), the Bayesian network is also similar to the linear regression (lower average error = 0.08 bps vs 0.16 bps, but similar RMSE= 1.39 % and R^2 =0.38%, and slightly higher MAPE= 99.3% vs 99.0%. But when imbalance is available (Panel B, columns 2 and 5), the Bayesian network has higher forecasting accuracy than the linear regression on all criteria (R^2 = 1.20% vs 1.10%, average error= -0.43 bps vs -1.08 bps, RMSE= 1.388 % vs 1.389%, and MAPE= 99.41 % vs 99.57%).

The Bayesian Network is particularly valuable when a subset of variables are only partially observable. In this case, the network captures the conditional dependencies between the nodes and fills the missing information with the most probable values of the latent variables. In our case, the realized imbalance is not used for the prediction, but the Bayesian network is trained on imbalance to infer its distribution given meta-orders characteristics. This gives a better forecast for the realized transaction cost, both in-sample and out-of-sample (for example higher R^2 =0.56% vs 0.52% in-sample, 0.50% vs 0.38% out-of-sample) than OLS or Bayesian networks that could not rely on this information.

Table 2.7 provides similar results to those in Table 2.6, but for ten deciles of orders

Table 2.6. Performance of the Bayesian network compared to the standard OLS regression

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1, 2010 to September 30, 2011. In-sample predictions are computed on 70% of the data such as the number of buy orders is equal to the number of sell orders. The remaining 30% serves for the out-of-sample prediction. RMSE and MAPE are respectively the Root Mean Squared Error and the Mean Absolute Percentage Error of the estimates

Imbalance Available	OLS Regression		Bayesian Network		
	No	Yes	No	Partial	Yes
Panel A: In-sample Estimation					
$\mathbb{E}[\text{IS}]$ (bps)	9.020	9.020	9.020	9.020	9.020
$\mathbb{E}[\hat{\text{IS}}]$ (bps)	7.590	8.161	7.606	8.617	8.588
$\mathbb{E}[\hat{\text{IS}} - \text{IS}]$ (bps)	-1.430	-0.859	-1.414	-0.403	-0.431
RMSE (%)	1.669	1.659	1.669	1.669	1.659
MAPE (%)	98.743	98.476	98.739	98.686	98.476
R^2 (%)	0.517	1.773	0.517	0.558	1.771
Panel B: Out-of-sample Estimation					
$\mathbb{E}[\text{IS}]$ (bps)	6.394	6.394	6.394	6.394	6.394
$\mathbb{E}[\hat{\text{IS}}]$ (bps)	6.557	5.317	6.482	8.030	5.960
$\mathbb{E}[\hat{\text{IS}} - \text{IS}]$ (bps)	0.163	-1.076	0.088	1.637	-0.434
RMSE (%)	1.394	1.389	1.394	1.393	1.388
MAPE (%)	99.022	99.570	99.301	99.340	99.410
R^2 (%)	0.377	1.104	0.378	0.502	1.204

size, and for Bayesian networks using partial or full information on the imbalance. We split the training and testing sets into 10 bins with respect to the training set order size. The first bin contains small orders, lower than 0.01% of daily volume, while the last one contains very large orders, higher than 4.34% of daily volume. We assess the accuracy of the Bayesian network within the three configurations of information availability (order flow imbalance fully, partially or not available). Consistent with intuition, we find that the inferred order flow imbalance distribution is more accurate when the investor holds a larger order. The posterior distribution of order flow imbalance given a small order is a symmetric U-shape function (Figure 2.9). At best it is slightly skewed, either positively or negatively, depending on the direction of the order. Thus, the larger the investors' trade, the more informative it is on the estimation of order flow imbalance, and as a consequence, the more accurate is the forecast of resulting implementation shortfall. We observe that

the R^2 increases steadily, whatever the configuration of information availability (partial or full), in-sample and out-of-sample, starting at the seventh bin. For example, the in-sample estimation of transaction costs when the imbalance is partially available goes from an R^2 of 0.18% for the seventh decile to 2.13% for the tenth decile, while smaller deciles of order size have relatively small R^2 (from -0.03% to 0.09% for the first 4 bins). Actually, for small order sizes, the market impact is very limited and disappears in market noise. Even if the Bayesian network is trained using the information on order flow imbalance, it has no advantage when the investor uses its trades attributes, if he trades only small order sizes. Said differently, it is hard to make good prior predictions of the order flow and thus the transaction cost when executing small orders. But we see how information on the investors' own orders becomes more informative on the aggregate net order flow as the investors' own order size gets larger. This is in line with the recent concentration of institutional investors executions on few dealing desks. Because the large dealing desk has a more accurate picture of investors' order flow imbalance of the day, it can assess the expected transaction cost more accurately and potentially design a better optimized executing scheme using this information. Note also that the RMSE does not drop, because higher-order size bins have few orders with large implementation shortfall that increases the average transaction cost for the bin. This is visible in the difference between the mean and the median realized trading cost (30.64 bps vs 24.79 bps in-sample for the tenth bin and 1.89 bps vs 1.67 bps for the first bin). The MAPE on the other hand, not suffering from this bias, gets smaller as the order size increase.

2.5.4 Inference of investors order flow imbalance given post-trade cost and market conditions

Investors' net order imbalance is a latent variable, thus not observable by the asset manager before executing his trade. His best prediction of market pressure is the inferred imbalance, after observing his own trading decision. However, his decision although usually in line with investors' trading because of the crowd effect, can depart from what is actually traded by his peers. One of the interests of our Bayesian network model is that

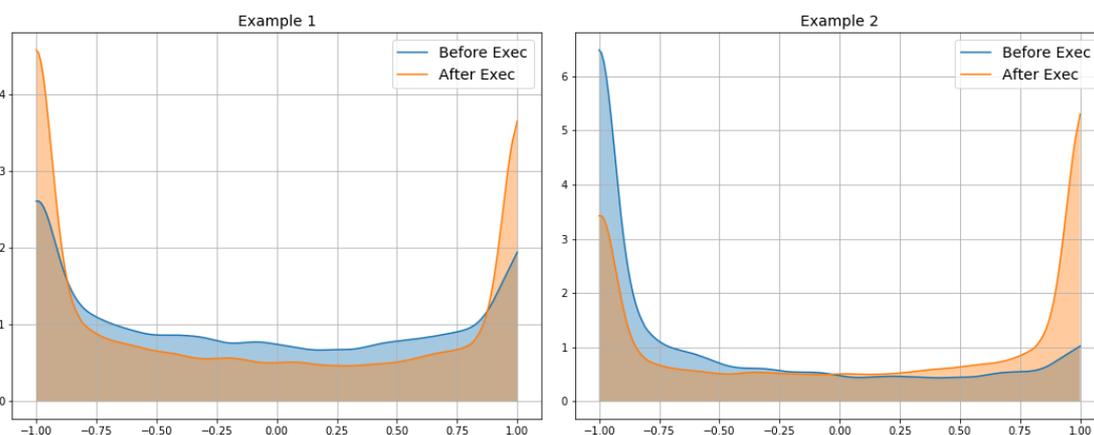
Table 2.7. Performance of the Bayesian network given order size

Institutional trading data are obtained from ANcerno Ltd on the period ranging from January 1, 2010 to September 30, 2011. In sample predictions are computed on 70% of the data such as the number of buy orders is equal to the number of sell orders. The remaining 30% serves for the out-of sample prediction. The sample are split in 10 bins w.r.t training set orders size. Q50 is the 50% quantile of implementation shortfall realized distribution. RMSE and MAPE are respectively the Root Mean Squared Error and the Mean Absolute Percentage Error of the estimates

Bins	1	2	3	4	5	6	7	8	9	10
$\mathbb{E}[\frac{Q}{ADV}]$ (%)	0.01	0.02	0.03	0.04	0.06	0.09	0.15	0.28	0.64	4.34
Panel A: In-sample Bayesian Estimation										
Effective Trading Costs										
$\mathbb{E}[IS]$ (bps)	1.89	2.87	2.61	4.85	6.88	7.82	7.01	9.25	16.50	30.64
Q50 (bps)	1.67	2.67	2.34	2.89	5.09	6.11	5.99	8.06	11.54	24.79
Imbalance Partially Available										
$\mathbb{E}[\hat{IS}]$ (bps)	4.26	4.52	4.82	5.25	5.78	6.51	7.58	9.32	12.59	25.62
RMSE (%)	1.41	1.44	1.51	1.52	1.56	1.61	1.68	1.74	1.92	2.17
MAPE (%)	100.19	99.78	100.11	99.69	98.92	98.59	98.46	97.86	96.91	93.60
R ² (%)	-	0.04	-	0.09	0.17	0.24	0.18	0.27	0.77	2.13
	0.03		0.00							
Imbalance Available										
$\mathbb{E}[\hat{IS}]$ (bps)	3.72	4.09	4.36	4.86	5.54	6.36	7.63	9.54	13.07	26.80
RMSE (%)	1.39	1.43	1.50	1.51	1.55	1.60	1.67	1.73	1.90	2.17
MAPE (%)	98.76	98.72	98.76	99.02	98.66	98.70	98.59	98.54	98.06	96.89
R ² (%)	1.50	1.69	1.43	1.37	1.55	1.35	1.33	1.35	2.09	2.89
Panel B: Out-of-sample Bayesian Estimation										
Effective Trading Costs										
$\mathbb{E}[IS]$ (bps)	-	-	0.75	2.70	3.80	4.83	8.09	8.81	12.71	28.09
	1.44	0.14								
Q50 (bps)	0.00	0.00	0.00	1.65	1.28	3.62	5.49	6.73	8.01	19.09
Imbalance Partially Available										
$\mathbb{E}[\hat{IS}]$ (bps)	4.53	4.77	5.05	5.39	5.87	6.50	7.46	8.98	11.85	22.73
RMSE (%)	1.22	1.24	1.27	1.37	1.32	1.35	1.40	1.45	1.61	1.71
MAPE (%)	101.28	101.25	101.23	100.99	101.33	98.96	98.23	97.59	97.22	93.65
R ² (%)	-	-	-	0.06	0.13	0.17	0.39	0.41	0.72	2.39
	0.17	0.09	0.05							
Imbalance Available										
$\mathbb{E}[\hat{IS}]$ (bps)	2.19	2.28	2.69	2.98	3.67	4.33	5.57	6.97	9.91	22.13
RMSE (%)	1.21	1.23	1.26	1.36	1.32	1.35	1.39	1.45	1.60	1.72
MAPE (%)	99.47	99.69	99.51	99.78	99.73	99.37	99.62	99.38	99.23	97.95
R ² (%)	1.18	1.25	1.15	0.85	0.92	1.18	0.81	0.96	1.34	2.08

Fig. 2.10. Bayesian Inference of Net Order Flow Imbalance

The figure plots in two panels the posterior distribution of net order flow imbalance given two example of market conditions and order characteristics. Each time, the blue curve plots the inferred distribution when only meta-orders attributes are considered and the orange line is the updated distribution once the resulting transaction cost is observed.



it can be used to recover the aggregate order flow imbalance prevailing during the investor's order execution, knowing his transaction costs. After receiving his *Transaction Cost Analysis* report, the investor could update his belief on investors imbalance during his execution using the calibrated Bayesian Network. We explore two cases as an example. First case: the investor sells a stock $s_k(m) = -1$; with a small participation rate $Q_k(m)/ADV_k(d) = 0.01\%$. His order is not very informative on market pressure since his trade is small, so his best estimate using the Bayesian network is a U-shape slightly skewed towards negative values of mean -0.10 (blue distribution of the left panel of Figure 2.10). Unexpectedly the resulting trading cost is huge $IS_k(m, d) = 3.02\%$ because the imbalance is very large and negative $Imb_k(m, d) = -0.94$. The investor could update his belief on the true distribution prevailing during his execution. The posterior distribution after incorporating the realized trading cost gives a higher probability to values at -1 (Orange line of the left panel). The average posterior imbalance distribution is -0.17 (Table 2.8). Second case: the investor takes the decision to sell massively a stock, $Q_k(m)/ADV_k(d) = 31.8\%$. This is usually happening during market panic where other investors sell massively as well. Therefore, his prior distribution is highly skewed to the left ($\mathbb{E}[Imb] = -0.40$, blue distribution of the right panel, Figure 2.10). While the in-

vestor expects a high transaction cost, he got lucky to be against the aggregate order flow ($\text{Imb}_k(m, d) = 0.94$) and benefited from a price improvement of 2.6%. The posterior imbalance distribution after incorporating this information is displayed on the right panel of Figure 2.10 (orange line) with a 0.05 average.

Table 2.8. Bayesian inference of net order flow imbalance

The table summarizes the first 2 moment and the 2.5% and 97.5% quantiles of net order flow imbalance inferred distribution for two scenarios before and after order implementation shortfall become available

	$\mathbb{E}[\hat{\text{Imb}}]$	Std $[\hat{\text{Imb}}]$	Q2.5%	Q97.5%
Inferred Imbalance: Example 1				
Before exec	-0.097	0.333	-1.0	0.953
After exec	-0.175	0.381	-1.0	0.999
Inferred Imbalance: Example 2				
Before exec	-0.400	0.328	-1.0	0.909
After exec	0.050	0.386	-1.0	1.000

2.6 Conclusion

In this paper, we use a Bayesian network to model transaction costs on US equity markets using ANcerno data, a large database of asset managers' instructions. Our main motivation is to make use of a variable of paramount importance for transaction costs, the *Net Order Flow Imbalance*. This variable is not observed by all market participants. Brokers and market makers have access to the imbalance of their clients' flows while dealing desks of asset managers do only observe their own instructions. Nevertheless, brokers, custodians, and even exchanges started recently reselling aggregate information on their clients' flows with a delay. Bayesian networks open new perspectives to model transaction costs using *latent variables*, i.e. variables that are not always known when the model has to be used but can be partially observed during the learning process. They enable to design a model linking observed and latent variables, based on conditional probabilities. The partially observable data can then be used to train the model.

Bayesian networks are able to estimate not only expected values but the whole probability distribution of a given variable. They are thus able to estimate the variance of the

residuals of their estimation. Because of the heteroskedasticity of the error term, market impact models and transaction costs estimates have traditionally a very small R^2 . A common belief among practitioners is that the effect of small mechanical price pressure is disappearing in the "market noise" (i.e. innovation on prices). We confirm this intuition in our model, by allowing the accuracy of trading costs forecasts to depend on market conditions and the investors' order characteristics. We find that the main variable explaining the variance of the residuals is the stock volatility, with a coefficient of 0.78.

Last but not least, we show several advantages of Bayesian networks for transaction costs forecasting. First, even when the latent variables (in our case, the imbalance of institutional orders at the start of the day) cannot be observed, the estimation relies on its pre-captured relationships with other observable variables (like the size and side of the investor's order to be executed). This allows the model to provide a better prediction than standard (for example OLS) models. Second, we show that the estimates get more accurate with the size of the meta-order the investor has to execute, because the larger the meta-order, the better the estimation of the order flow imbalance. This gives an informational advantage for large dealing desks in charge of executing the orders of numerous or large investors as they have a better picture of the aggregate imbalance. This finding is consistent with the current evolution of market practice. Small asset managers increasingly use the services of large dealing desks to benefit from this information, leading to the recent concentration of institutional investors' orders on a few dealing desks. Finally, these models can use Bayesian inference to deduce the expected distribution of the latent variable. We show how it is feasible to ask the Bayesian network the expected distribution of large orders of other investors, either at the start or at the end of the day, once the resulting trading costs are observed.

Bayesian networks are very promising models to account for partial information. They could prove particularly valuable for "alternative datasets", like airlines activity, web traffic, or financial flows, that often provide very detailed information on a small subset of transactions. They are difficult to use in standard models, that do not accept missing values. Bayesian networks structurally model the relationship between missing and known variables. They could naturally fill this gap.

References

- Almgren, R., & Chriss, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk*, 3, 5–40.
- Amari, S.-i. (1993). Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4-5), 185–196.
- Anand, A., Irvine, P., Puckett, A., & Venkataraman, K. (2011). Performance of institutional trading desks: An analysis of persistence in trading costs. *The Review of Financial Studies*, 25(2), 557–598.
- Anand, A., Irvine, P., Puckett, A., & Venkataraman, K. (2013). Institutional trading and stock resiliency: Evidence from the 2007–2009 financial crisis. *Journal of Financial Economics*, 108(3), 773–797.
- Angel, J. J., Harris, L. E., & Spatt, C. S. (2015). Equity trading in the 21st century: An update. *The Quarterly Journal of Finance*, 5(01), 1550002.
- Bacry, E., Iuga, A., Lasnier, M., & Lehalle, C.-A. (2015). Market impacts and the life cycle of investors orders. *Market Microstructure and Liquidity*, 1(02), 1550009.
- Bew, D., Harvey, C. R., Ledford, A., Radnor, S., & Sinclair, A. (2018). Modelling analysts' recommendations via bayesian machine learning. *Available at SSRN*.
- Bouchard, B., Dang, N.-M., & Lehalle, C.-A. (2011). Optimal control of trading algorithms: A general impulse control approach. *SIAM Journal on financial mathematics*, 2(1), 404–438.
- Briere, M., Lehalle, C.-A., Nefedova, T., & Raboun, A. (2019). Stock market liquidity and the trading costs of asset pricing anomalies. *Available at SSRN 3380239*.
- Bucci, F., Mastromatteo, I., Eisler, Z., Lillo, F., Bouchaud, J.-P., & Lehalle, C.-A. (2018). Co-impact: Crowding effects in institutional trading activity. *arXiv preprint arXiv:1804.09565*.
- Capponi, F., & Cont, R. (2019). Trade duration, volatility and market impact. *Volatility and Market Impact (March 14, 2019)*.
- Cetin, U., & Rogers, L. (2007). Modeling liquidity effects in discrete time. *Mathematical Finance*, 17(1), 15–29.

- Collins, B. M., & Fabozzi, F. J. (1991). A methodology for measuring transaction costs. *Financial Analysts Journal*, 47(2), 27–36.
- Dayri, K., & Rosenbaum, M. (2015). Large tick assets: Implicit spread and optimal tick size. *Market Microstructure and Liquidity*, 1(01), 1550003.
- Eisele, A., Nefedova, T., Parise, G., & Peijnenburg, K. (2017). Trading out of sight: An analysis of cross-trading in mutual fund families.
- Frazzini, A., Israel, R., & Moskowitz, T. (2012). Trading costs of asset pricing anomalies.
- Garman, M. B., & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, 67–78.
- Gomes, C., & Waelbroeck, H. (2015). Is market impact a measure of the information value of trades? market response to liquidity vs. informed metaorders. *Quantitative Finance*, 15(5), 773–793.
- Greenwood, R., & Thesmar, D. (2011). Stock price fragility. *Journal of Financial Economics*, 102(3), 471–490.
- Haldane, A. Et al. (2014). The age of asset management. *speech at the London Business School*, 4(4).
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Applications to nonorthogonal problems. *Technometrics*, 12(1), 69–82.
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5), 359–366.
- Jordan, M. I. (1998). *Learning in graphical models* (Vol. 89). Springer Science & Business Media.
- Koch, A., Ruenzi, S., & Starks, L. (2016). Commonality in liquidity: A demand-side explanation. *The Review of Financial Studies*, 29(8), 1943–1974.
- Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315–1335.
- Laruelle, S., & Lehalle, C.-a. (2018). *Market microstructure in practice*. World Scientific.

- Lauritzen, S. L. (2003). Some modern applications of graphical models. *Oxford Statistical Science Series*, 13–32.
- Moro, E., Vicente, J., Moyano, L. G., Gerig, A., Farmer, J. D., Vaglica, G., Lillo, F., & Mantegna, R. N. (2009). Market impact and trading profile of large trading orders in stock markets.
- Novy-Marx, R., & Velikov, M. (2015). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(1), 104–147.
- Pagano, M. S. (2008). Which factors influence trading costs in global equity markets? *The Journal of Trading*, 4(1), 7–15.
- Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial intelligence*, 29(3), 241–288.
- Perold, A. F. (1988). The implementation shortfall: Paper versus reality. *Journal of Portfolio Management*, 14(3), 4.
- Puckett, A., & Yan, X. (2011). The interim trading skills of institutional investors. *The Journal of Finance*, 66(2), 601–633.
- Robert, E., Robert, F., & Jeffrey, R. (2012). Measuring and modeling execution cost and risk. *The Journal of Portfolio Management*, 38(2), 14–28.
- Salvatier, J., Fonnesbeck, C. Et al. (2016). Pymc3: Python probabilistic programming framework. *Astrophysics Source Code Library*.
- Skaanning, C., Jensen, F. V., & Kjærulff, U. (2000). Printer troubleshooting using bayesian networks, In *International conference on industrial, engineering and other applications of applied intelligent systems*. Springer.
- Torre, N., & Ferrari, M. (1999). The market impact model tm. barra research insights. 1999 barra. Inc.
- Vapnik, V. N., & Chervonenkis, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities, In *Measures of complexity*. Springer.
- Vila, J.-P., Wagner, V., & Neveu, P. (2000). Bayesian nonlinear model selection and neural networks: A conjugate prior approach. *IEEE Transactions on neural networks*, 11(2), 265–278.

Wyart, M., Bouchaud, J.-P., Kockelkoren, J., Potters, M., & Vettorazzo, M. (2008). Relation between bid–ask spread, impact and volatility in order-driven markets. *Quantitative Finance*, 8(1), 41–57.

Zuo, Y., & Kita, E. (2012). Stock price forecast using bayesian network. *Expert Systems with Applications*, 39(8), 6729–6737.

Appendices

2.A Garman Klass volatility definition

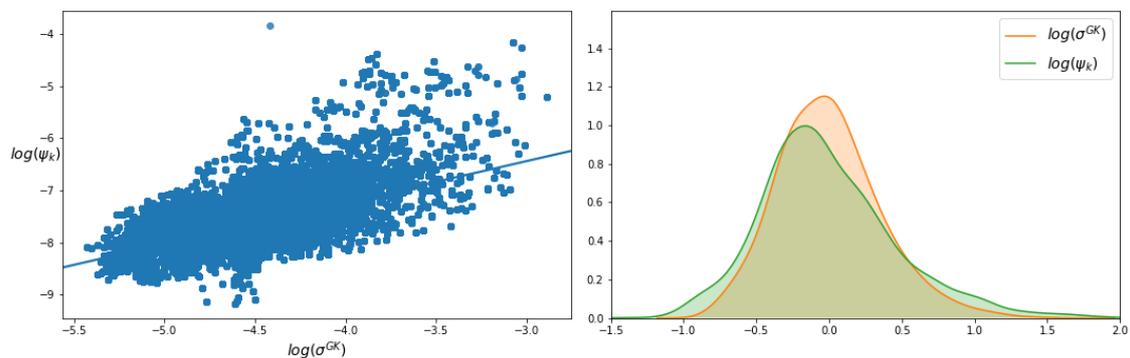
Garman-Klass estimate of the volatility uses the open, high, low and close prices of the day. This estimate is robust and very close in practice to more sophisticated ones. The formula is given by:

$$\sigma_k^{\text{GK}}(d) = \sqrt{\frac{1}{N} \sum_{t=1}^N \frac{1}{2} \log \left(\frac{H_{d-t}^k}{L_{d-t}^k} \right)^2 - (2 \log(2) - 1) \log \left(\frac{C_{d-t}^k}{O_{d-t}^k} \right)^2} \quad (2.15)$$

where the indexation k refers to the stock. d to the calculation day. N is the length of the rolling window in day. In our case 252 trading days. $O_t^k, H_t^k, L_t^k, C_t^k$ are respectively the open, high, low, close prices of stock k at day t

2.B Bid-Ask spread and volatility distribution dependencies

Fig. 2.11. Bid-Ask spread and volatility distribution dependencies



The left panel of figure 2.11 shows the scatter plot of the log bid-ask spread and log volatility of S&P 500 components of 2010 and 2011. It proves that the variables are

related. The right panel displays the centered distributions $(X - \bar{X})$ of the log bid-ask spread and the log volatility.

2.C Beta distribution properties

The probability density function of the Beta distribution PDF_{Beta} is given by:

$$\forall x \in [0, 1] \quad PDF_{Beta}(\alpha, \beta, x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \quad \text{where} \quad B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \quad (2.16)$$

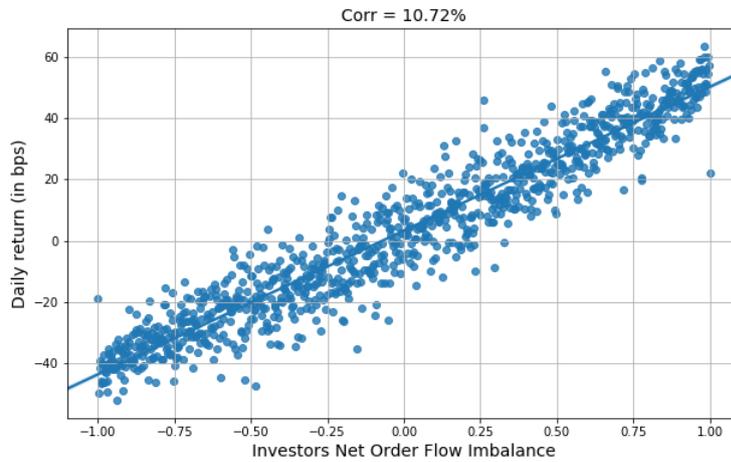
The first 3 moments of the distribution are as following:

$$\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta} \quad \text{Var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \quad \text{Skew}[X] = \frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}}$$

Note that the skew of the distribution is proportional to $(\beta - \alpha)$. So when $\alpha \ll \beta$ the probability density function is significantly skewed toward values at 0 and in the opposite case $\alpha \ll \beta$ the probability density function is skewed toward values at 1. The particular case where $\alpha = \beta$ the distribution is symmetric around the mean $\mathbb{E}[X] = \frac{1}{2}$. ($PDF_{Beta}(\alpha, \beta, \frac{1}{2} + x) = PDF_{Beta}(\alpha, \beta, \frac{1}{2} - x)$) and the skew is null. if in addition the condition $(\alpha - 1)(\alpha - 2) > 0$ is fulfilled the distribution has a U-shape. Otherwise the Beta distribution produces a concave function.

2.D Net order flow imbalance properties

Net order flow imbalance has a strong predictive power of daily returns. The cross sectional average correlation for S&P 500 index components on our period of study is significantly positive up to 10.72% (Figure 2.12). Furthermore, investors trading imbalance prevail through time. Table 2.9 shows that the daily imbalance auto-correlation decays slowly from 12.03% for the first lag to 7.44% after 5 days. Since order flow imbalance is only available with a delay, the long memory of the imbalance is appreciated.

Fig. 2.12. Net order flow imbalance, daily returns correlation**Table 2.9. Net Order Flow Imbalance auto-correlation**

	Imb_{t-1}	Imb_{t-2}	Imb_{t-3}	Imb_{t-4}	Imb_{t-5}
Imb_t	12.03	9.11	8.37	7.69	7.44

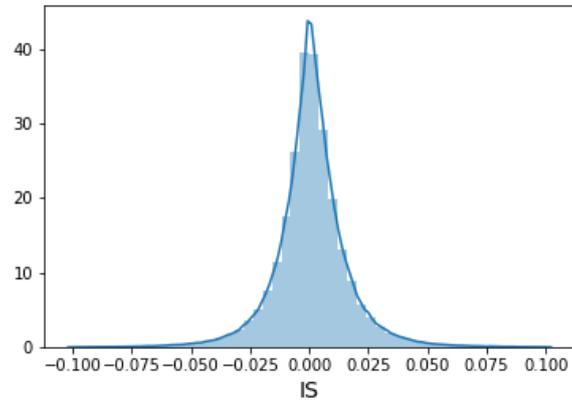
2.E Implementation shortfall distribution

The implementation shortfall estimated on ANcerno meta-orders on S&P 500 components of 2010 and 2011, displays a non-normal distribution centered at 0, with standard deviation equal to 0.64, a positive skew of 0.34, and highly significant excess kurtosis of 23.46. These moments are more comparable to a double exponential distribution.

2.F Hasting-Metropolis algorithm

Hasting-Metropolis is one of the pioneer Markov Chain Monte Carlo algorithms developed in the early 90s to sample from an unknown distribution. Given a function f proportional to the desired probability distribution $P(x)$ (a.k.a the target distribution) and a proposal distribution $q(\cdot) = q(\cdot|x)$ easy to simulate, the algorithm constructs a series of variables (x_1, x_2, \dots, x_n) such as given x_n

1. Generate $y_n \sim q(y|x_n)$,

Fig. 2.13. Implementation shortfall marginal distribution

2. Generate $u \sim \mathcal{U}[0, 1]$ a uniform distribution
3. Compute the acceptance rate $\alpha = \min \left\{ \frac{f(y_n)q(x_n|y_n)}{f(x_n)q(y_n|x_n)}, 1 \right\}$
4. Accept the new candidate y_n with probability α if $u \leq \alpha$ Otherwise reject.

$$X_n = \begin{cases} y_n, & \text{if } u \leq \alpha \\ x_n, & \text{otherwise} \end{cases} \quad (2.17)$$

Chapter 3

Liquidity Provision and Market-Making in Different Uncertainty Regimes: Evidence from the Covid-19 Market Crash

About this chapter

I want to thank Euronext Paris Analytics team for their help and support to understand and query the order-book database efficiently. A special thank you to Gaelle Le Fol, Robert Kosowski, and Frederic Abergel for the fruitful discussion during the Pre-Defense that helped to improve the content of this chapter

Abstract

Kyle, 1985 builds a pioneering and influential model, in which an insider observing private information submits an optimal order given the market-maker's pricing rule, which is assumed a linear function of the aggregated order flow. We propose an extension to Kyle's model where different types of uncertainty regimes exist and where the market maker estimates market uncertainty and uses it to set her price. The model implies that the elasticity of prices to liquidity demand will increase in high uncertainty regimes. We test the outcome of the model empirically by studying the price formation process during the COVID-19 pandemic crash. A period of agitation with important announcements having a major impact on financial markets, such as the state lockdown and the Fed's fiscal response. We find that indeed the elasticity of prices to liquidity demand during the COVID-19 period increased threefold

3.1 Introduction

The micro-structure of financial markets has increasingly attracted the attention of academics, regulators, and market participants as it allows speculators to design efficient trading algorithms, market-makers to manage their inventory risk, asset-managers to reduce their transaction costs, and regulators to design efficient policies to improve stocks' liquidity and prices informativeness. However, modeling order-book dynamics and the price formation process is not straightforward. In theory, prices are formed in the markets in real-time, as traders confront their views about the asset's future value. Traders are assumed to acquire pieces of information that they believe could influence the asset price and trade upon it. In the trading process, prices are formed as market participants update their beliefs and adjust their orders when interacting with other traders. Koiijen et al., 2020 show that different investors, with different investment horizons and mandates, allow to incorporate different firms characteristics into prices. Kyle, 1985 is one of the reference models in this field, close to a game-theoretic approach, where an informed trader ob-

serves the fundamental value of a risky asset and chooses her traded amount strategically in order to maximize her profit, and a market-maker sets the clearing price by inferring the value of the asset from the noisy aggregated volume submitted by all traders, informed and uninformed indiscriminately. This model, although simple, highlights the complex interactions between liquidity providers and liquidity consumers and the resulting price formation process.

A few academic papers showed evidence that besides the reaction to market flows, market-makers also react to exogenous information, and change their trading behavior whenever new information is disclosed. For instance, Megarbane et al., 2017 analyzed the impact of the 2015 European Central Bank announcements on monetary policy and the 2016 Brexit vote on financial markets and reported a regime change of HFTs trading during those two events. The authors emphasize that HFTs are the main market-makers on the limit order book contributing to more than 80% of market depth in normal market conditions. However, during these two events, HFTs reduced their role of liquidity provision, which was taken over by the rest of the market participants. The same behavior was documented by Tung, 2000 on the impact of earning announcement on financial markets. The author argues that what makes the bid-ask spreads widen during earning announcements is the level of information asymmetry between informed traders and market-makers. Lyle et al., 2019 study price reaction to earnings announcement when the information is disclosed during the trading session or just before the market opens compared to an announcement after the close of the previous day. They find a significantly lower price reaction in the post-closing announcement because it gives market participants sufficient time to digest the newly disclosed information and confront their views about the firm's value at the opening auction of the next day.

In this paper, we explore whether market-makers react solely to market flows, or also take into account exogenous information in their trading scheme. To do so, we propose a theoretical extension to Kyle, 1985 model with two trading regimes. A regime of high uncertainty, where market-makers are exposed to high information asymmetry and characterized by a high variance of expected informed prices, and a regime of low uncertainty, with probably lower information asymmetry between market-makers and in-

formed traders. Our model assumes that the variability of the risky asset value changes with the regime of uncertainty, that the informed trader observes the realization of the fundamental value of the risky asset and chooses her traded amount strategically to maximize her informational rent, and that the market-maker reacts to the noisy market flow according to his estimate of the prevailing regime of uncertainty. This allows us to model the market-maker's excess price reaction to exogenous information arrival which is not explained by market flows. This setup is optimal from an information cost perspective. As Grossman and Stiglitz, 1980 stated, the decision of acquiring information is theoretically motivated by the cost-benefit trade-off. The informed trader needs to acquire complete information to submit the optimal volume that will maximize her expected profit. Based on the partial information related to the level of uncertainty in the market, the market-maker could decide whether or not to provide liquidity. He then increases the cost for liquidity provision during periods of high uncertainty and quotes more aggressively during stable periods where trades contain less to no information. All the remaining assumptions are kept identical as the Kyle, 1985 model.

There have been other attempts to extend Kyle, 1985 strategic trading model. Holden and Subrahmanyam, 1992 considers the competition among multiple insiders each endowed with perfect private information. Foster and Viswanathan, 1996 study the competition with heterogeneous private signals. S. Huddart et al., 2001 examines the case where an insider must announce her trading volume after the submission while S. J. Huddart et al., 2004 study the case with pre-announcement of insider trade. More recently, Caldentey and Stacchetti, 2010 study the extended Kyle model with insider observing a signal that tracks the evolution of the asset's fundamental value and with a random public announcement time revealing the current value of the asset. A common feature of these different models is that they all focus on the informed trader behavior and assume the same pricing rule for the market-maker. On the contrary, we postulate that market makers change their pricing rule based on their estimate of the level of uncertainty in the market. Hence, our extension of the Kyle model by considering a different trading behavior for market-makers might have potential applications on various models based on Kyle, 1985 framework.

In the empirical part of this paper, we test model implication on the 40 stocks composing the CAC40 index in the first quarter of 2020 using the Euronext tick-by-tick database. This period was marked by the Covid-19 market crash. Baker et al., 2020 argues that no previous infectious disease outbreak, including the Spanish Flu, has ever impacted the stock market prices as powerfully as the COVID-19 pandemic. As it could be seen in the unprecedented high levels of the VIX index, markets have been taken by surprise and were unable to anticipate a shock of this nature. The crisis thus provides a unique opportunity to test theories on the price formation process and the trading behavior of market participants. We propose a methodology to estimate empirically Kyle's lambda on central limit order book data and apply it to our universe of securities. It consists of deriving the price set by market-makers from their activity in the order-book, insertion, and cancellation of volume in the best bid and best ask limits and estimate the elasticity of market-makers quotes to the aggregate liquidity demand of the market as specified by Kyle model with a linear form¹. We then study the incorporation of the information of the health crisis exogenous shock on market-makers pricing rule. We find that indeed market-makers' quotations are well explained by the signed liquidity demand in the market. The results are significant to 1% or more and robust to all controls on firm characteristics and day fixed effects. The elasticity of market-makers' prices to market liquidity demand increased by 39bps in the high uncertainty context of the health crisis, suggesting that market makers take into account the exogenous information in their pricing rule. Moreover, during periods of stress, natural risk holder increase their trading activity and contribute more to the amount of liquidity provision. Nevertheless, their liquidity provision has no explanatory power of daily cross-sectional price returns. We conclude that the price response to liquidity demand is a characteristic of market-participants insuring the intermediation role in the market, as claimed by Kyle, 1985.

The rest of the paper is organized as follows: Section 2 explains the model, its as-

¹Although, empirical evidence suggests a square root model, with marginal price impact diminishing as the number of shares traded increases (Briere et al., 2019, Almgren et al., 2005, ...), we keep a linear function as in Kyle, 1985, Back and Baruch, 2007 and Chordia and Subrahmanyam, 2004 because it results from the assumption of exponential utility, risk neutrality of traders' preferences and Gaussian payoff. Models with non-linear market impact are usually analytically intractable and suited for high volume orders, while we are interested in modeling market-makers response to the volume at the micro-structure level.

assumptions, and implications. In section 3 we present Euronext order-book data and give some descriptive statistics on market conditions and market agent's trading behavior during the COVID-19 pandemic crash. In Section 4, we propose an empirical estimation of Kyle's lambda using order-book data and study market-makers pricing rule during two regimes of uncertainty. Finally, Section 5 explains the implications on asset pricing at the daily level.

3.2 Model and assumptions

In this section, we present the model assumptions and implications. Similarly to Kyle, 1985, we assume the existence of a risky asset with an uncertain liquidation value θ . We introduce a variable $\xi \sim \text{Bernoulli}(P_\xi)$ that indicates the type of the regime of uncertainty of day t . $\xi = \mathbf{1}_{d \in RI_1}$ equals 1 when the regime of uncertainty high RI_1 , 0 otherwise, and we assume θ to be normally distributed with a mean θ^* and a time dependent standard deviation $\sigma_\theta(t)$ $\theta \sim N(\theta^*, \sigma_\theta(t))$ that depends on the regime of uncertainty of the day t . Formally,

$$\sigma_\theta(t) = \sigma_{\theta,0} + (\sigma_{\theta,1} - \sigma_{\theta,0})\xi \quad \text{where } \sigma_{\theta,0} < \sigma_{\theta,1} \quad (3.1)$$

The variability of the asset liquidation value is higher during uncertain or highly informative periods (regime RI_1). We assume the presence of three types of market players: an informed trader, a noise trader, and a market-maker

- **Informed trader:** a trader who can access alternative sources of data and assess the efficient price faster than other market participants, creating an asymmetric information situation in the market. We can think of as sophisticated financial institutions that analyze different data sources, like text data, customer transactions, and satellite images to extract valuable information on the efficient price. Therefore, we assume in the model, the informed trader knows the realization of the asset value θ in the future and sets the size of her trade strategically to maximize her profit. The informed trader buys, (or sells) until her expected profit from holding (liquidating)

another share is exactly offset by the price impact of her trade. The informed trader solves the following optimization problem.

$$\operatorname{argmax}_Q \mathbb{E}[Q(\theta - P^M)|\theta] \quad (3.2)$$

- Noise trader: also referred to as a liquidity trader. A set of traders who submits orders to the market in a zero-intelligence fashion either while seeking liquidity or for hedging purposes or simply their signals are inaccurate and cancel out. Noise traders' total demand u is endogenously-generated, normally distributed with mean 0 and a standard deviation σ_u . $u \sim N(0, \sigma_u)$.
- Market-Maker: provides liquidity to market participants. We assume the market-maker holds partial information ξ^M about the type of regime of the day t . ξ^M is a Bernoulli variable, such that $\mathbf{P}[\xi^M = 0|\xi = 0] = 1$ and $\mathbf{P}[\xi^M = 1|\xi = 1] = P_{\xi^M}$. Meaning that, when there is no uncertainty about the level of the stock market, and there is no macro-economic news or firm's related publication, $\xi = 0$, the market-maker considers the day belonging to a standard informational regime, $\xi^M = 0$. However, whenever there is information or uncertainty about the firm's fundamentals $\xi = 1$, the market maker identifies this information only with probability P_{ξ^M} and mistake a day with high uncertainty for a standard day with probability $(1 - P_{\xi^M})$. The market-maker sets the clearing price upon observing the aggregate order flow $w = Q + u$ and estimating the information regime to be as close as possible to the expected value of the asset.

$$P^M = \mathbb{E}[\theta|Q + u, \xi^M] \quad (3.3)$$

Consistent with Kyle, 1985, we define the Bayesian Nash Equilibrium of this economy as on one side, the informed trader's strategy maximizing her expected profit, given the market-maker's pricing rule and signal and on the other side, the market-maker setting

²It is straight forward to relax this assumption by considering a time-dependent standard deviation of noise traders aggregate demand varying with the type of informational regime. The derivation of the results is straight forward

the price to be equal to the expected value of the asset given aggregated order flow. This is quite a game-theoretic concept as the market-maker does not maximize an objective function but rather sets the price to be equal to the expected value of the payoff.

We assume the market-maker holds a linear pricing rule, with λ_K the elasticity of the price to the aggregate demand, also known as Kyle's lambda.

$$P^M = \tilde{P} + \lambda_K(Q_\theta + u) \quad (3.4)$$

The unique equilibrium in linear strategies of this economy is characterized by the informed traders demand strategy Q_θ

$$Q_\theta = \frac{\theta - \theta^*}{2\lambda_K} \quad (3.5)$$

and the market-maker's pricing rule:

$$\begin{cases} \tilde{P} & = \theta^* \\ \lambda_K & = \frac{\text{cov}(\theta, Q_\theta + u | \xi^M)}{\text{Var}(Q_\theta + u | \xi^M)} = \frac{\sqrt{\text{Var}[\theta | \xi^M]}}{2\sigma_u} \end{cases} \quad (3.6)$$

where,

$$\begin{aligned} \text{Var}[\theta | \xi^M = 1] &= \sigma_{\theta,1}^2 \\ \text{Var}[\theta | \xi^M = 0] &= \frac{(1 - P_\xi)}{(1 - P_{\xi^M}) \cdot P_\xi + (1 - P_\xi)} \sigma_{\theta,0}^2 + \frac{(1 - P_{\xi^M}) \cdot P_\xi}{(1 - P_{\xi^M}) \cdot P_\xi + (1 - P_\xi)} \sigma_{\theta,1}^2 \end{aligned}$$

We have the following result proved in Appendix B.

Meaning that the market-maker's reaction to the aggregate flow depends on her estimation of the regime of uncertainty ξ^M and the variance of noise traders' aggregate volume. Market-makers' pricing rule (equation 3.4) could be rewritten:

$$P^M = \theta^* + \sqrt{\frac{\text{Var}[\theta | \xi^M = 0]}{4\sigma_u^2}} (Q_\theta + u) \mathbf{1}_{\xi^M=0} + \frac{\sigma_{\theta,1}}{2\sigma_u} (Q_\theta + u) \mathbf{1}_{\xi^M=1} \quad (3.7)$$

Using the Bayes rule, we can easily prove that whenever the market-maker estimates that the day belongs to a regime of high uncertainty $d \in RI_1$, it is actually the case — $\mathbf{P}[\xi = 1 | \xi^M = 1] = 1$. This is because the market-maker will not change her trading behavior unless there is a good reason to. For instance an external shock on the economy or scheduled news. In such cases, the market-maker adjusts her pricing rule, by asking for the highest risk premium, $\lambda_K = \frac{\sigma_{\theta,1}}{2\sigma_u}$ to compensate for the risk on price uncertainty. However, when the market-maker does not have any reason to consider a day t as belonging to a regime of high uncertainty, $\xi^M = 0$, and knowing she misses pieces of information about the firm's fundamental value, she sets the price taking into account the probability she might be mistaking the level of uncertainty regime ($\mathbf{P}[\xi = 1 | \xi^M = 0]$). Therefore the $\mathbf{Var}[\theta | \xi^M = 0]$ is a probability-weighted average of $\sigma_{\theta,0}$ and $\sigma_{\theta,1}$. As a result, a market maker who estimates properly the type of regime detains a high probability of $P_{\xi^M} \approx 1$ and thus reacts the least to the aggregate flow. $\lim_{P_{\xi^M} \rightarrow 1} \lambda_K = \frac{\sigma_{\theta,0}}{2\sigma_u}$. In a competitive market-making environment, this will make a significant difference during the standard market conditions, because the market-maker with the highest P_{ξ^M} asks for the lowest risk premium λ_k for liquidity provision and wins more trades. This is because she is better informed and less exposed to the risk of adverse selection. On the other side of the equation, we note that the traded amount chosen strategically by the informed trader is inversely proportional to the market-maker's risk aversion λ_K , as specified by Kyle, 1985 model. Thus during major macro-economic events, stress periods, or recently disclosed information about the company, the informed trader would expect the market-maker to incorporate this information in his quoted prices and thus reduces the size of her trade proportionally. In a word, the more information moves prices (i.e. large $\sigma_{\theta,1}$) the higher the price impact and the more non-informative flow, the more difficult for the market-maker to identify information, hence the less he moves the price. Although this model is not an accurate description of how real-world traders interact with market makers, however, it is a simple way of capturing key facts: informed traders trade more aggressively when their information is private and surrounded with noise trading and trade more cautiously when the information is leaked to the market and market-makers are vigilant to trading signal in the market.

3.3 Data

The following analysis is based on Euronext tick-by-tick data of the CAC40 Index components from January the 1st 2020 to April the 1st 2020. Euronext Paris is the main trading venue of the studied securities. According to the CBOE website tracking the market share of European equities trading venues, Euronext accounts for 72.68% of the traded amounts of Paris listed companies in March 2020, followed by CBOE Europe 16.37%, then Aquis 4.98%. Turquoise and Equiduct come at 4th and 5th position with respectively 3.7% and 1.91% market share. This makes Euronext order book the place where the offer meets the demand and where the prices are primarily formed. The database we use tracks the entire history of the orders, from their initial submission to full execution including order submissions, partial fills, potential modifications, and cancellations. We use Euronext order-book data to study market conditions during the COVID-19 market crash and test empirically model implications. Table 3.1 gives summary statistics of market variables: price returns of CAC40 index components, daily traded volume, the 5min Garman-Klass volatility³, and the time-weighted average bid-ask spread in basis points, during the first quarter of 2020, and Figure 3.1 shows their evolution in view of the three days with major events during this period of market turmoil. February 24: the first trading day after the lockdown, in Lombardy, Northern Italy, not far from the country's main economic center of Milan. March 11th the day the President of the United States announced a travel ban on EU countries. It is also the day the World Health Organisation (WHO) declared the Coronavirus a pandemic case. Finally, the 18th of March 2020, when the Federal Reserve began making purchases under the Commercial Paper Funding Facility to alleviate the strain in short-term credit markets. It is also the date President Trump signed the second Coronavirus Emergency Aid Package (CEAP) (the Families First Corona Response Act). First, we note that the Italy lockdown decision marked the beginning of the market crash. This can be clearly seen in the drop in securities prices, the sudden rise in traded volume, and the jump in the intraday volatility. Market sanity gradually deteriorated as the virus spread, however the recognition of the virus as a pandemic case by WHO accelerated

³Garman Klass volatility at 5min scale uses the open high low close prices of every 5min bin of the continuous session to estimate the intraday volatility. The calculation is detailed in the appendix

the trend. Intraday volatility exploded to reach 100% on average between the 11th and the 18th of March and above 140% for some securities. Regarding the bid-ask spread, market-makers succeeded to keep their quotations tight in the period between Italy lockdown and the pandemic declaration by the WHO, even if the market was already abnormally volatile reaching 60% volatility, but it soon increased significantly from 3bps on average before the crisis to 15bps, and this is for the 40 most liquid companies in the French market. Mid and small-capitalization experienced even more severe liquidity shock. Finally, the Fed's announcement was the first stabilizing factor of the financial markets. Volatility like the bid-ask spread starts converging back to their pre-crisis average and prices reverted back progressively. Note also that the price drop during the first phase of the pandemic was much severe during the continuous session compared to the uncrossing phase, Panel(a). This price movement is particularly interesting in the scope of this analysis as it is formed by the interaction between market participants including market-makers and the various potential informed trader, unlike the overnight jump resulting from an equilibrium price established at the open auction. In the rest of the paper, we identify the COVID-19 pandemic shock as the period after Italy's lockdown until the end of our sample. The same period has been considered in papers studying the COVID-19 market crash, such as Ramelli and Wagner, 2020, and Albuquerque et al., 2020.

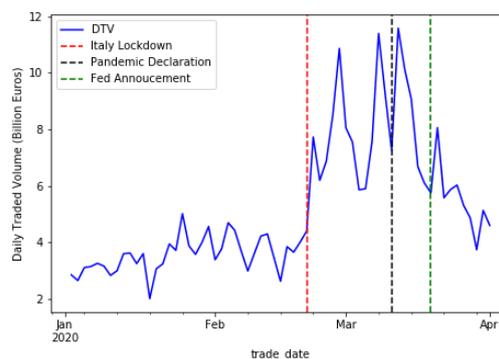
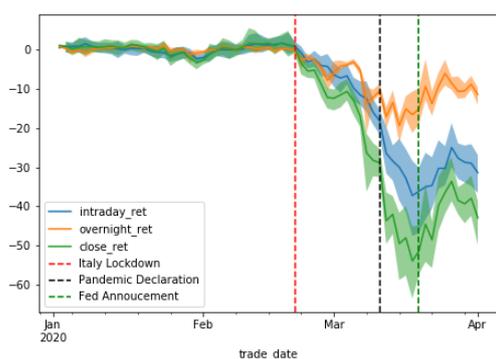
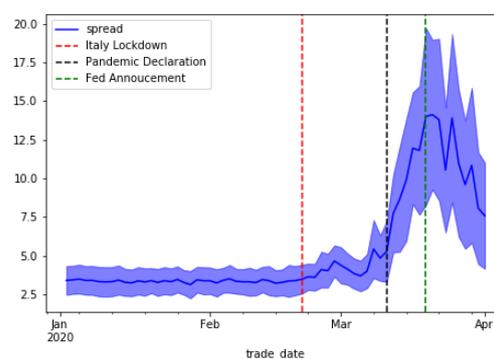
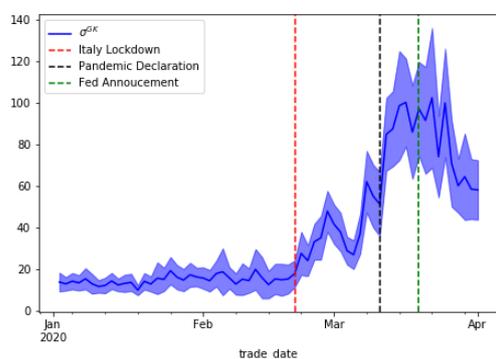
Table 3.1. Summary statistics on Market conditions

This table reports the summary statistics (number of observations, mean, standard deviation [SD], and 25th, 50th [median], and 75th percentiles) of market variables describing trading conditions on the CAC 40 Index components, from 1st January to 1st of April 2020. The intraday price change (in log %), the overnight price jumps (in log %), the daily price return (in log %), the traded volume (in millions of Euros), the average 5min Garman-Klass volatility (in %), and the average bid-ask spread (in bps).

	Mean	SD	25%	Median	75%	Obs.
Intraday log return (%)	-0.48	3.04	-1.46	-0.22	0.74	2600
Overnight log return (%)	-0.18	2.61	-0.80	0.03	0.69	2600
Daily log return (%)	-0.66	4.17	-1.75	-0.21	0.92	2600
Daily traded volume (MEur)	128.16	115.83	52.36	92.16	156.69	2600
Garman Klass 5min volatility (%)	35.13	31.45	12.58	19.74	50.03	2600
Bid-Ask spread (bps)	5.27	3.98	3.10	3.90	5.26	2600

Fig. 3.1. French stock market reaction to COVID-19 Pandemic

This figure plots the average intraday price movement (in log %), the average overnight price jumps (in log %), the sum of traded volume in (billions of Euros), the average 5min Garman-Klass volatility (in %), and the average bid-ask spread (in bps) of CAC 40 Index components from 1st January to 1st April 2020. The red, black and green dotted line represent respectively the 24th February, the 11th March, and the 18th March. The shaded area around the mean is the 5% confidence interval in all panels.

(a) Intraday vs Overnight Cumulative Returns (%)
(b) Sum of Daily Traded Volume (B€)

(c) Garman Klass 5min volatility
(d) Average Bid-Ask Spread (bps)


Euronext tick-by-tick database indicates the type of order's origin in a specific field called account type. It is a flag retrieved from the order message that indicates the nature of the flow. The main account type categories on Euronext central limit order book are the Client, the House, the Liquidity Provider "LP", and the Retail Member Organization "RMO"⁴. When the order is flagged "Client", it means that the member is providing ex-

⁴The order could also be flagged "RLP" —Retail Liquidity Providers. It is a type of order dedicated to retail investors only, where liquidity providers could offer price improvements compared to the central limit order book, in a commercial package called "best of the book". RLP members are not allowed to interact with trading members other than retail. Therefore, we exclude their flow in this analysis

ecution service for a third party institutional investor (asset managers, hedge funds, pension funds, etc.). This category is in contrast with the “House” account type indicating that the member is bearing the risk of trading for her own account. Typically, arbitragers and proprietary trading desks of banks submit ”House” flagged orders. The third major account type is “LP” for Liquidity Provision. For the CAC 40 Index components, only members registered in the market-making program on blue-chip securities, called Supplemental Liquidity Provider (SLP), could submit such orders. These members have to meet market-making requirements, including order book presence time and competitive quotes at the bid and ask sides on a regular basis. Finally, the RMO account type corresponds to brokers submitting orders on behalf of their retail clients. Note that the same trading member could submit all kinds of orders if it is involved in different activities (i.e Brokerage, Arbitrage, Market-Making...). Orders related to each activity will be channeled to the exchange with the corresponding account type flag to distinguish between the different flows. Hence, this categorization allows us to uncover the role of different trading schemes in the market even if they are implemented by the same entity.

In order to evaluate the trading behavior of the different categories of market players, we study the evolution of account-types traded volume in percentage daily turnover, and account-types contribution to market overall liquidity provision. We define the following metrics:

- **Share in amount traded:** is the amount of volume traded by members of the account type as a percentage of daily traded volume

$$MS_A(s,t) = \frac{Q_A(s,t)}{\sum_A Q_A(s,t)}$$

Where $Q_A(s,t)$ is the amount traded by members of account type A on the stock s at day t

- **Share in liquidity provision:** is the percentage amount of daily traded volume where members of the account type were providing liquidity to other market participants.

$$MSLP_A(s,t) = \frac{Q_A^{Pa}(s,t)}{\sum_A Q_A^{Pa}(s,t)}$$

where $Q_A^{Pa}(s,t)$ is the amount traded passively by members of category A during day t on the stock s . A passive order is an order that adds liquidity to the order-book by entering bids and offers, which are not immediately executed but will be consumed later by other traders' market orders.

Table 3.2 reports the account-types trading activity on CAC40 index constituents during the first quarter of 2020. We see that the LP category is by far the most active in terms of traded volume. It accounts on average for 48.88% of daily traded volume, the Client and the House flows come second with comparable traded volume market share respectively 24.37% and 24.08% while the RMO flow comes at the last position with a 2.67% average percentage of the trading volume. This distribution of traded volume market share illustrates well the role of intermediation played by the LP category between Client and House flows. Glosten and Milgrom, 1985 claim that mature markets should see market-makers participate in a significant proportion of transactions in electronic order books, converging asymptotically to half of the daily traded amount because they take part in every transaction where the natural buyers and sellers do not perfectly meet in the market at the same time with the same desired quantity and opposite sides. Therefore, market-makers wait for the buyers after providing liquidity to the sellers and vice-versa. In terms of liquidity provision measured by the percentage amount of passive trades contributed by each account type, the LP category comes first representing 41.20% of passive traded volume in the market. The Client flow composed of institutional investors executions has the second-highest share of liquidity provision after market-makers with 31.44% of overall market liquidity provision. Since the introduction of the concept of "best execution" by the first Markets in Financial Instruments European directive (MiFID I), brokers are required to prove through transaction costs analysis reports that order execution price is comparable to the market VWAP (Volume Weighted Price). Thus, the broker's execution algorithms, nowadays, are very liquidity seeking. The proprietary flow "House" has the lowest liquidity provision share among the three most active categories 25.34%. Finally,

we note that retail investors contribute to 2.02% of market liquidity provision⁵

Table 3.2. Summary statistics on Account type trading

This table describes account types trading activity for CAC 40 Index components trading activity, from 1st January to 1st of April 2020. Share in amount traded is the percentage amount of volume traded by the category. Share in liquidity provision is the percentage amount of volume traded passively by members of the account type over the daily traded volume.

		Mean	SD	25%	Median	75%	Obs.
Panel A: Share in amount traded							
All Sample	Client	24.37	5.81	20.31	23.97	27.93	2600
	House	24.08	4.79	20.64	23.75	27.08	2600
	LP	48.88	7.23	43.59	49.00	53.88	2600
	RMO	2.67	2.86	1.04	1.78	3.02	2600
Without COVID-19 effect	Client	24.05	5.59	20.02	23.57	27.70	1440
	House	25.23	4.76	21.86	25.09	28.25	1440
	LP	48.52	6.42	43.61	48.58	53.04	1440
	RMO	2.20	2.12	0.99	1.60	2.54	1440
With COVID-19 effect	Client	24.77	6.05	20.61	24.52	28.38	1160
	House	22.65	4.44	19.58	22.38	25.43	1160
	LP	49.33	8.10	43.43	49.61	54.84	1160
	RMO	3.25	3.49	1.12	2.05	3.79	1160
Panel B: Share in liquidity provision							
All Sample	Client	31.44	9.70	24.34	30.58	37.40	2600
	House	25.34	6.44	20.77	24.78	29.21	2600
	LP	41.20	11.36	32.95	41.52	49.71	2600
	RMO	2.02	2.87	0.52	1.08	2.17	2600
Without COVID-19 Effect	Client	29.36	8.72	23.00	28.12	34.59	1440
	House	25.47	6.29	21.10	25.10	29.25	1440
	LP	43.49	10.01	35.81	43.80	51.02	1440
	RMO	1.68	2.42	0.47	0.98	1.71	1440
With COVID-19 effec	Client	34.03	10.22	26.94	33.55	40.47	1160
	House	25.16	6.62	20.32	24.38	29.15	1160
	LP	38.35	12.26	29.75	37.79	47.05	1160
	RMO	2.46	3.30	0.62	1.24	2.71	1160

⁵Retail investors profit from the "best of the book" offer, where dedicated market-maker offer them better prices than those available in the lit central limit order book.

Table 3.2 makes also the parallel between the trading behavior of the account types in terms of share in traded amount and share in market liquidity provision before the Coronavirus pandemic outbreak and the period of agitation after the Italy lockdown. First, we point out that the Client category increased their contribution to market liquidity provision during the market crash going from 29.36 % before the crash to 34.03% after. This while keeping a stable percentage daily volume in a period where the number of transactions increased threefold. Some institutional investors, having a longer-term investment horizon, have seen the market drop a good entry point. In a selling market pressure, they were able to execute their buying position mostly with limit orders. We also observe a similar behavior from retail investors, increasing their percentage traded volume from 2.2% to 3.25% and their share in liquidity provision from 1.68% to 2.46%. The AMF report on retail behavior in the COVID-19 confirms this substantial activity of retail clients. Bonnet, 2020 highlight that some retail investors had already been active in the months preceding the Covid-19 crisis and 150,000 new investors bought SBF 120 shares for the first time since at least January 2018. On the contrary, we note a slight decrease in the proprietary flow percentage of trading volume from 25.23% to 22.65% trading activity with relatively the same contribution to market liquidity provision and a decrease in market-makers share of liquidity provision from 43.49% to 38.35%. Market-makers prices are composed of at least three components. The order processing cost. It represents a fee charged by market makers for standing ready to match buy and sell orders (Tinic, 1972). This component includes compensation for the market makers to insure their work of intermediation, this component remains identical in the two periods. The inventory holding cost, modeled in Stoll, 1978 and Ho and Stoll, 1981. It compensates dealers for holding less than fully diversified portfolios. During periods of agitation, the risk associated with price changes between the moment market-makers provide liquidity and the moment they liquidate their inventory is higher. Finally, the adverse selection component (Copeland and Galai, 1983 and Glosten and Milgrom, 1985) represent the risk premium that market-makers charge to accept dealing with traders who may have superior information. Krinsky and Lee, 1996 find that the adverse selection costs component is the most prominent before and following public announcements and increase significantly due to increased information

asymmetry. The market-maker is the market agent who takes into account all these cost components in setting his price. Which explains the decrease in liquidity provision.

3.4 Kyle model estimation

In this section, we estimate empirically the pricing rule of market-makers as a function of the liquidity demand. However, contrary to the Kyle, 1985 model setup, where the market-maker sets the clearing price of the market after all traders have submitted their desired quantity and side, in electronic order-book the trading is continuous and market-makers do not have the advantage of playing last. Instead, they submit limit orders to position themselves optimally in the bid and ask queues in a way to maximize their profit while keeping their inventory as balanced as possible. For this reason, we define a price contributed by market-makers that takes into account their activity in the order book. Several papers have studied the market-making optimal pricing strategy in a limit order-book, such as Abergel et al., 2020, Avellaneda and Stoikov, 2008, Guéant et al., 2013, and Lehalle and Mounjid, 2017. All these models agree that the distribution of orders arrival in the order-book is a function of the bid-ask imbalance. For example, if the volume at the ask limit is notably higher than the volume at the bid limit ($Q_a(t) \gg Q_b(t)$), then there are higher chances that the next transaction will occur at the best bid price and not at the best ask price. Degryse et al., 2005 shows that aggressive orders take place when order depth in the opposing side is relatively low, because either the next seller would choose to submit a market order instead of adding herself to an already filled queue or one of the traders already placed in the ask queue would decide to cancel her order to cross the bid-ask spread if she deems the opportunity cost exceeds the transaction cost. Inversely, the next buyer would have less incentive to pay the bid-ask spread in this configuration. Therefore, market-makers provide liquidity in line with their estimation of the most probable future price, skewing their quotes if needed to reduce the amount of future liquidity provision in the side where they have already accumulated inventory, or submitting market orders if the inventory risk is unbearable. We consider the instantaneous expected price (i.e. the most probable next traded price conditionally to the state of the order book

— $\mathbb{E}[P(t)]$), and derive the market-maker’s contributed price as the expected price change due to market-makers activity in the bid and ask best limits. Harris, 2013 refers to this price as “the true value of the asset” and explains why, theoretically, it corresponds to the instantaneous equilibrium price in a linear utility framework for supply and demand. (see Appendix B for the details). Stoikov, 2018 refers to this metric as “the micro-price” and shows empirically its predictive power to the short-term price movement.

Formally, if $Q_b(t)$ and $Q_a(t)$ are respectively the bid and ask volumes available at the first limits at time t , and $P_b(t)$ and $P_a(t)$ are their corresponding prices. Then, the micro-price is defined as:

$$\mathbb{E}[P(t)] = \frac{Q_b(t) P_a(t) + Q_a(t) P_b(t)}{Q_b(t) + Q_a(t)}$$

Financial markets today allow market participants to submit a variety of orders (Modify, Iceberg, Market-to-Limit, etc), but all these orders could be decomposed into 3 elementary orders: Fill or Market order, Insert order, and Cancel order⁶. These three orders allow traders to manifest their impatience for liquidity. As stated in Foucault et al., 2005, when a trader submits a Market/Fill order, she chooses to pay the cost of immediacy (the bid-ask spread) to get executed. However, when a trader chooses to insert an order, she accepts to delay her execution by waiting for the next trader to cross the bid-ask spread and consume the liquidity she posted. Finally, the trader could cancel an already inserted order hoping to get a better price in the future. For example, if the price has high chances to drop (the imbalance is highly negative), the buyer waiting at the best bid limit has high chances to be executed, but she may choose to wait longer, by canceling her order and insert it farther in the order-book at a better price. From the market-makers’ perspective, they insert orders to provide liquidity to the rest of the market and cancel them to withdraw liquidity when they sense a higher risk of adverse selection, or their estimation of the fair price has moved, or they can no longer bear the risk of the accumulated inventory. As

⁶For example, a Modify order is a Cancel order followed by an Insert of a new order at a different price. An Iceberg is an Insert order triggered by a transaction that consumes all the volume in the best limit. A Market-to-limit order is a Market/Fill order with a size capped at the volume available in the opposing best limit, followed by an Insert order with the remaining volume, if any, at the price of the consumed limit

a consequence, we compute the contribution of market-makers to the price as the sum of the micro-price movements resulting from market-makers actions that change the quantities available at the best bid and best ask limits. If $\tau_0^{MM}, \dots, \tau_k^{MM}, \dots, \tau_N^{MM}$ are the stopping times of such events in the order-book during the time frame $[d + (t - 1)r, d + tr]$ (the t^{th} period of time of length r of the trading day d , $r = 5\text{min}$ in our case), then the expected price change due to market makers' activity during this time frame is given by

$$\Delta P^{MM}(s, d, k) = \sum_{t=1}^N \mathbb{E}[P(\tau_t^{MM})] - \mathbb{E}[P(\tau_{t-1}^{MM})]$$

Note that the state of the order book could change between two consecutive stopping times τ_{t-1}^{MM} and τ_t^{MM} , if other market participants contribute to the order-book depth, insert or cancel orders or when liquidity consumers take part or all of the volume available in the best limits. These other actions taken by the rest of the market participants are one of the reasons why a market-maker may choose to adjust his quotation, to either place himself in a better position to maximize his chances to be the one to provide liquidity to the next trader or inversely to lessen this probability by canceling his well-positioned orders.

The liquidity demand ($Q + u$ in the model) within the time frame $[d + (k - 1)r, d + kr]$ is measured by the net traded volume executed by all market participants and signed by the sign of the liquidity consumer order (+ for buy and - for sell). To ensure comparability between stocks with different liquidity characteristics in our sample and take into account the U shape of the volume curve during the day (Laruelle and Lehalle, 2018), we rebase this measure by dividing the net volume by traded volume during the considered time frame. This adjustment is fairly common in empirical estimations in market micro-structure literature (see Capponi and Cont, 2019, Said et al., 2017 or Laruelle and Lehalle, 2018)

$$LD(s, d, k) = \frac{Q_{Buy}(s, d, k) - Q_{Sell}(s, d, k)}{Q(s, d, k)}$$

where $Q_{Buy}(s, d, k)$ and $Q_{Sell}(s, d, k)$ are the traded amounts respectively initiated by the buyers and the sellers on stock s during the time frame $[d + (k - 1)r, d + kr]$, and $Q(s, d, k) =$

$Q_{Buy}(s, d, k) + Q_{Sell}(s, d, k)$ is the traded volume during the same period. As we are interested in the stationary behavior of market-participants in the statistical validation, we consider a linear model with a single representative period, where the price movement contributed by market-makers orders reacts to the aggregate liquidity demand of all market participants. Although empirical evidence suggests a square root model with marginal price impact diminishing with the number of shares traded (Briere et al., 2019, Almgren et al., 2005, ...), the concave market impact function is only suited for large sized orders. On the contrary, orders studied at the micro-structure level are smaller and are executed immediately. Thus, their market impact is linearly subject to the limit order book depth.

$$\Delta P^{MM}(s, d, k) = \alpha + \lambda LD(s, d, k) + \beta_s \text{ Firm Controls}_s + \beta_d \text{ Day FE}_d + \varepsilon(s, d, k) \quad (3.8)$$

We use this specification to quantify the average market-makers' price reactions to the liquidity demand in the first quarter of 2020. We identify market-makers orders as the ones submitted by members of the "LP" account type insuring the market-making rule in the Euronext regulated market. The unit of observation (s, d, k) refers respectively to the stock, the trading day, and the 5min time period. We check the robustness of the result by controlling for several firm characteristics. Namely, the size of the firm calculated as the log free-float market capitalization of the firm on the 31st of December 2019, the annualized volatility of daily logarithmic returns of stocks prices in 2019, the average bid-ask spread in 2019, and the financial leverage in 2019 calculated as the ratio of the book value of debt over book assets. Then, we estimate market-makers' pricing rule in response to the uncertainty regime as defined in equation 3.4. We distinguish the elasticity of market-makers micro-price movement to the net volume in a regime of low uncertainty ($\xi^M = 0$) taken as the period before Italy lockdown, and the elasticity of their prices when they are aware of the exogenous shock of Covid-19 ($\xi^M = 1$). We run the following cross-sectional regression:

$$\Delta P^{MM}(s, d, k) = \alpha \mathbf{1}_{d \in \text{COVID-19}} + \lambda_1 LD(s, d, k) + \lambda_2 LD(s, d, k) \mathbf{1}_{d \in \text{COVID-19}} + \beta_s \text{ Firm Controls}_s + \beta_d \text{ Day FE}_d + \varepsilon(s, d, k) \quad (3.9)$$

Table 3.3 presents the results of the two regressions. In the first four columns, we give the result of the market-makers' contribution to the micro-price movement as a func-

Table 3.3. Cross-sectional regression of market-makers contribution to the micro-price on the aggregate liquidity demand

$\Delta P^{MM}(s, d, k)$ is the log market-makers contribution to the micro-price, presented in basis points. $LD(s, d, k)$ is the net liquidity demand of stock s during the t^{th} 5min bin of day d (in %). $\mathbf{1}_{d \in \text{COVID-19}}$ is a dummy variable that equals 1 during the period after the Italy lockdown in February 24. Historical spread and historical volatility are the average bid-ask spread and the average daily price return volatility calculated over 2019. Market capitalization is the log free float market capitalization at the 31st of December 2019. The financial leverage is the book value of debt over book assets in 2019. The numbers in parentheses are t-statistics. * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$.

Dependent variables	Without Covid effect			With Covid effect		
	$\Delta P^{MM}(s, d, k)$			$\Delta P^{MM}(s, d, k)$		
$LD(s, d, k)$	26.23*** (78.56)	26.37*** (78.84)	26.45*** (78.98)	14.45*** (36.32)	14.55*** (36.56)	14.65*** (36.77)
$LD(s, d, k)$				38.86*** (53.78)	39.15*** (54.04)	39.15*** (53.99)
$\mathbf{1}_{d \in \text{COVID-19}}$				0.32 (1.45)		
Leverage		-0.1*** (-2.9)			-0.1*** (-3.9)	
B-A spread		-0.5*** (-2.9)			-0.6*** (-3.5)	
Volatility		-0.03* (-1.7)			-0.05*** (-2.6)	
Market Cap (ln)		1.18*** (3.29)			1.14*** (3.18)	
Firm FE	No	No	Yes	No	No	Yes
Day FE	No	Yes	Yes	No	Yes	Yes
R2 (%)	2.29	2.67	2.74	3.35	3.73	3.80

tion of the aggregate liquidity demand of the market during the full period (equation 3.8). In the first column, we use the liquidity demand as the only independent variable to explain the price movement. In the second column, we add firm control variables. In the third column, we add day fixed effects to make sure the results are not driven by market conditions of some specific days where an exogenous shock happened, and in the

fourth column, we control for both day and firm fixed effects. Standard errors are robust to heteroscedasticity. The effect of the net volume on market-makers contribution to the micro-price movement is positive and significant at the 1% level or better, and stable after controlling for all variables. The magnitude of the coefficient estimate suggests that every 1% increase in the polarized net liquidity demand results in a change of market-makers quotation that drives the micro-price by 14.45 basis points, on average in the direction of market pressure. The economic magnitude of this coefficient encompasses the response of market-makers quotations both before and after the pandemic outbreak. After distinguishing between the period of low uncertainty and high uncertainty (Columns from 5 to 8 presenting the results of equation 3.9), we find that market-makers reaction to liquidity demand is 4 times larger after the COVID-19 exogenous shock. Immediately after the lockdown in Italy, every 1% increase in liquidity demand, resulted in an increase of 39bps of market-makers contributed micro-price in the direction of market pressure. This difference in quotation adjustment to market pressure illustrates well how market-makers are risk-averse during periods of high uncertainty. Finally, firms with high liquidity attributes (low bid-ask spread, low volatility, and high market capitalization) and firms with less financial leverage have larger price adjustment from market-makers. Since the offer and demand of liquid shares are usually balanced outside any context, a large net flow in absolute value is more informative for more liquid assets resulting in larger price adjustment from market-makers.

3.5 Implications on the Price Formation Process

One consequence of information exogenous shocks triggered by public announcements or information disclosure is the increased natural risk holders trading activity who either need to trade immediately to manage their risk or see in the event a good opportunity to position themselves in the market and bet on the informational rent of announcement. However, this liquidity flow is different from the liquidity provided by market-makers. On one hand, the designated intermediaries ensuring the market-making rule in the market are bound by the inventory risk and the adverse selection risk. Hence, as specified by

Kyle, 1985, they must set the price as close as possible to the assets' true value by inferring this information from the noisy aggregated volume. On the other hand, the rest of market-participants may provide liquidity in different circumstances. For instance, while implementing their trading strategies or managing their risk constraints. Therefore, we expect that their trading scheme is not liquidity-driven and their prices move less in reaction to market flows. In this section, we study the joint effect of the liquidity provided by market-makers and by non-market-makers on the daily price returns, in view of the COVID-19 exogenous shock.

Similarly to the previous section, we define the daily liquidity demand $LD(s, d)$ as the ratio of the net traded volume signed by the liquidity consumer order side and the continuous traded session volume. This definition is equivalent to the volume-weighted average of the 5min liquidity demand considered above and correct for the U shape of the daily traded volume curve.

$$\begin{aligned} LD(s, d) &= \sum_k^N \frac{Q(s, d, k)}{Q(s, d)} LD(s, d, k) \\ &= \frac{Q_{Buy}(s, d) - Q_{Sell}(s, d)}{Q(s, d)} \end{aligned} \quad (3.10)$$

where $Q_{Buy}(s, d)$ and $Q_{Sell}(s, d)$ are the traded volume initiated respectively by the buyers and the sellers. We distinguish between the traded volume provided by market-makers, $(Q_{Buy,MM}(s, d), Q_{Sell,MM}(s, d))$ defined as the volume initiated by buyers and seller in the market and where the counterparty was a member of the LP category, and the volume provided by non-market makers $(Q_{Buy,O}(s, d)$ and $Q_{Sell,O}(s, d))$. The liquidity demand answered by each category of market player, market-makers and non-market-makers follows naturally as the ratio of the amount of flow imbalance reduced by each category and the traded volume

- The daily liquidity demand addressed to market-makers

$$LD_{MM}(s, d) = \frac{Q_{Buy,MM}(s, d) - Q_{Sell,MM}(s, d)}{Q(s, d)}$$

- the daily liquidity demand addressed the non market-makers

$$LD_O(s, d) = \frac{Q_{Buy,O}(s, d) - Q_{Sell,O}(s, d)}{Q(s, d)}$$

We run the cross-sectional regressions of stocks' intraday price returns $\Delta P(s, d)$ on market liquidity demand $LD(s, d)$, and one the liquidity demand addressed by the two categories of market-participants $LD_{MM}(s, d)$ and $LD_O(s, d)$ to study their relative effect of daily price return. Then, we isolate the crisis effect on price elasticity to volume, by making all variables of interest interact with the Covid-19 dummy.

$$\begin{cases} \Delta P(s, d) = \alpha + \lambda LD(s, d) + \beta_s \text{ Firm Controls}_s + \varepsilon(s, d) \\ \Delta P(s, d) = \alpha + \lambda_{MM} LD_{MM}(s, d) + \lambda_O LD(s, d, O) + \beta_s \text{ Firm Controls}_s + \varepsilon(s, d) \end{cases} \quad (3.11)$$

Table 3.4 shows the result of the different specifications of intraday price returns regression on the liquidity demand. In the first column, we regress the intraday price movement on the aggregate liquidity demand in the market submitted to all market participants. Not surprisingly, the daily price movement is increasing with the liquidity demand, as it represents the signed market pressure. More buyers than sellers during a trading session increases the price and vice-versa. However, we find that the daily price elasticity to the signed volume has increased substantially during the period of the health crisis outbreak, going from 1.61 before Italy's lockdown to 3.03 after (column 3). The second and the fourth columns, compare the effect of the liquidity demand addressed by market-makers and by non-market-makers on the daily price movement. This shows the marginal contribution of each type of liquidity provision in the price formation process. We find that when market-makers reduce market liquidity demand by 1% they move daily prices by 7.28 %, while other market participants move the daily price by 2.79 % only to reduce market flow imbalance by the same amount. This difference in price elasticity to volume is even larger during the period of high uncertainty, where the liquidity demand addressed to market-makers has the most impact on daily price movement 11.96 % vs 2.13 % for non-market-makers liquidity. This increase in the price reaction to liquidity demand cor-

responds to an increase in the risk premium that market-makers ask to provide liquidity in an uncertain environment with several unknowns. The result shows that market makers' reaction to flows follows Kyle's model far more than other participants, even when the latter partly provide liquidity

Table 3.4. Cross-sectional regression of intraday price changes on the daily liquidity demand

This table reports the results of the cross-sectional regressions of the price intraday change (in ln) on the liquidity demand $LD(s, d)$, the liquidity demand addressed to the designated market-maker $LD_{MM}(s, d)$ and the liquidity demand addressed to other trading members $LD_O(s, d)$, on CAC 40 Index constituents, from January 1st to April 1st 2020. The numbers in parentheses are t-statistics. * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$.

Dependent variables $\Delta P(s, d)$	Without COVID-19 effect		With COVID-19 effect	
	(1)	(2)	(3)	(4)
$LD(s, d)$	3.38*** (4.78)		1.61* (1.78)	
$LD(s, d) \cdot \mathbf{1}_{d \in \text{COVID-19}}$			3.03** (2.17)	
$LD_{MM}(s, d)$		7.28*** (4.02)		3.12 (1.44)
$LD_{MM}(s, d) \cdot \mathbf{1}_{d \in \text{COVID-19}}$				11.96*** (3.15)
$LD_O(s, d)$		2.79*** (3.73)		1.34 (1.38)
$LD_O(s, d) \cdot \mathbf{1}_{d \in \text{COVID-19}}$				2.13 (1.45)
$\mathbf{1}_{d \in \text{COVID-19}}$			-1.08*** (-9.06)	-1.1*** (-9.27)
Leverage	-0.03*** (-3.06)	-0.03*** (-3.0)	-0.03*** (-3.0)	-0.03*** (-2.95)
Bid-Ask spread	0.05 (0.52)	0.05 (0.5)	-0.08 (-1.19)	-0.07 (-1.14)
Market Cap (ln)	0.56*** (2.87)	0.54*** (2.77)	0.48** (2.57)	0.47** (2.5)
R2 (%)	1.88	2.09	7.59	8.07

3.6 Conclusion

The COVID-19 pandemic and the subsequent lockdown brought about an exogenous and unparalleled shock on the stock market. The crisis thus provides a unique opportunity to test theories on the price formation process and the change in the trading behavior of market participants. In this paper, we highlight the role of exogenous information on market-makers' pricing rules. We propose an extension to Kyle, 1985 model where the market-maker estimate the level of market uncertainty and incorporate it into his pricing rule. The model implies that the market-maker will react differently to market flows in different market regimes. Thus, the elasticity of prices to liquidity demand will be higher during a crisis and high uncertain regimes. When the market-maker holds no information that would allow him to assume high uncertainty, he defines the price relative to the probability of mistaking the nature of the day's pattern. This assumption is more in line with the empirical observations of market-makers changing their trading behavior during periods of information disclosure. We test model implication empirically on the Coronavirus outbreak in the first quarter of 2020. The magnitude of the market reaction to the pandemic was commensurate to the severity of the economic shock, but also to the fact that market participants have been taken by surprise and were unable to anticipate a shock of this nature. We propose an empirical estimation of Kyle's lambda which exploits the richness of the order book to derive market-makers prices from their activity in the best bid and ask limits. We find that indeed market-makers' quotations are well explained by the signed liquidity demand in the market as specified by the Kyle model. The results are robust to all controls and significant at 1% or better. Besides, the exogenous information of the Covid-19 pandemic has been incorporated in market-makers pricing rule. We note a lower price resistance to liquidity demand from market-makers post-Covid of 39bps compared to the more stable period before Italy's lockdown. Moreover, during periods of high uncertainty, the rest of the market-participants increase their trading activity and share in liquidity provision. Nevertheless, we show that when other market-participants provide liquidity, their prices take into account the polarization of the flows to a certain extent but have no significant effect on daily price returns. Thus, price reaction to flows

is a market-maker characteristic, as specified by Kyle's model.

References

- Abergel, F., Huré, C., & Pham, H. (2020). Algorithmic trading in a microstructural limit order book model. *Quantitative Finance*, 20(8), 1263–1283.
- Albuquerque, R., Koskinen, Y., Yang, S., & Zhang, C. (2020). Resiliency of environmental and social stocks: An analysis of the exogenous covid-19 market crash. *The Review of Corporate Finance Studies*.
- Almgren, R., Thum, C., Hauptmann, E., & Li, H. (2005). Direct estimation of equity market impact. *Risk*, 18(7), 58–62.
- Avellaneda, M., & Stoikov, S. (2008). High-frequency trading in a limit order book. *Quantitative Finance*, 8(3), 217–224.
- Back, K., & Baruch, S. (2007). Working orders in limit order markets and floor exchanges. *The Journal of Finance*, 62(4), 1589–1621.
- Baker, S., Bloom, N., Davis, S. J., Kost, K., Sammon, M., & Viratyosin, T. (2020). The unprecedented stock market reaction to covid-19. *Covid Economics: Vetted and Real-Time Papers*, 1(3).
- Bonnet, S. F. C. (2020). Retail investor behaviour during the covid-19 crisis. *AMF Report/Study*.
- Briere, M., Lehalle, C.-A., Nefedova, T., & Raboun, A. (2019). Stock market liquidity and the trading costs of asset pricing anomalies. *Available at SSRN 3380239*.
- Caldentey, R., & Stacchetti, E. (2010). Insider trading with a random deadline. *Econometrica*, 78(1), 245–283.
- Capponi, F., & Cont, R. (2019). Trade duration, volatility and market impact. *Volatility and Market Impact (March 14, 2019)*.
- Chordia, T., & Subrahmanyam, A. (2004). Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Economics*, 72(3), 485–518.
- Copeland, T. E., & Galai, D. (1983). Information effects on the bid-ask spread. *the Journal of Finance*, 38(5), 1457–1469.

- Degryse, H., Jong, F. D., Ravenswaaij, M. V., & Wuyts, G. (2005). Aggressive orders and the resiliency of a limit order market. *Review of Finance*, 9(2), 201–242.
- Foster, F. D., & Viswanathan, S. (1996). Strategic trading when agents forecast the forecasts of others. *The Journal of Finance*, 51(4), 1437–1478.
- Foucault, T., Kadan, O., & Kandel, E. (2005). Limit order book as a market for liquidity. *The review of financial studies*, 18(4), 1171–1217.
- Garman, M. B., & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, 67–78.
- Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics*, 14(1), 71–100.
- Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, 70(3), 393–408.
- Guéant, O., Lehalle, C.-A., & Fernandez-Tapia, J. (2013). Dealing with the inventory risk: A solution to the market making problem. *Mathematics and financial economics*, 7(4), 477–507.
- Harris, L. (2013). Maker-taker pricing effects on market quotations. *USC Marshall School of Business Working Paper*. Available at <http://bschool.huji.ac.il/upload/hujibusiness/Maker-taker.pdf>.
- Ho, T., & Stoll, H. R. (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial economics*, 9(1), 47–73.
- Holden, C. W., & Subrahmanyam, A. (1992). Long-lived private information and imperfect competition. *The Journal of Finance*, 47(1), 247–270.
- Huddart, S. J., Hughes, J. S., & Williams, M. (2004). Pre-announcement of insiders' trades. Available at SSRN 216168.
- Huddart, S., Hughes, J. S., & Levine, C. B. (2001). Public disclosure and dissimulation of insider trades. *Econometrica*, 69(3), 665–681.
- Koijen, R. S., Richmond, R. J., & Yogo, M. (2020). *Which investors matter for equity valuations and expected returns?* (Tech. rep.). National Bureau of Economic Research.

- Krinsky, I., & Lee, J. (1996). Earnings announcements and the components of the bid-ask spread. *The Journal of Finance*, 51(4), 1523–1535.
- Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315–1335.
- Laruelle, S., & Lehalle, C.-A. (2018). *Market microstructure in practice*. World Scientific.
- Lehalle, C.-A., & Mounjid, O. (2017). Limit order strategic placement with adverse selection risk and the role of latency. *Market Microstructure and Liquidity*, 3(01), 1750009.
- Lyle, M. R., Stephan, A., & Yohn, T. L. (2019). The speed of the market reaction to pre-open versus post-close earnings announcements. *Kelley School of Business Research Paper*, (17-75).
- Megarbane, N., Saliba, P., Lehalle, C.-A., & Rosenbaum, M. (2017). The behavior of high-frequency traders under different market stress scenarios. *Market Microstructure and Liquidity*, 3(03n04), 1850005.
- Ramelli, S., & Wagner, A. F. (2020). Feverish stock price reactions to covid-19.
- Said, E., Ayed, A. B. H., Husson, A., & Abergel, F. (2017). Market impact: A systematic study of limit orders. *Market Microstructure and Liquidity*, 3(03n04), 1850008.
- Stoikov, S. (2018). The micro-price: A high-frequency estimator of future prices. *Quantitative Finance*, 18(12), 1959–1966.
- Stoll, H. R. (1978). The supply of dealer services in securities markets. *The Journal of Finance*, 33(4), 1133–1151.
- Tinic, S. M. (1972). The economics of liquidity services. *The Quarterly Journal of Economics*, 79–93.
- Tung, S. (2000). The effect of information asymmetry on bid-ask spreads around earnings announcements by nasdaq firms. *Review of Pacific Basin Financial Markets and Policies*, 3(03), 331–346.

Appendices

3.A Model Proof

The informed trader maximizes her expected profit $\operatorname{argmax}_Q \mathbb{E}[Q(\theta - P^M | \theta)]$. To solve the model, first, we plug the functional form of the market maker's pricing rule into the informed trader's optimization problem:

$$\begin{aligned} \frac{\partial \mathbb{E}[Q(\theta - P^M) | \theta]}{\partial Q} = 0 &\Rightarrow \mathbb{E}[\theta - \tilde{P} - \lambda_K u | \theta, \xi^M] - 2\lambda_K Q = 0 \\ &\Rightarrow Q_\theta = \frac{\theta - \tilde{P}}{2\lambda_K} \end{aligned} \quad (3.12)$$

The market maker sets the price equal to his conditional expectation of the asset's value given the aggregate demand and his signal. Thus, the market-maker have to choose (\tilde{P}, λ_K) so that

$$P^M = \mathbb{E}[\theta | Q + u, \xi^M] = \tilde{P} + \lambda_K(Q_\theta + u)$$

$$\left. \begin{aligned} \mathbb{E}[P^M] &= \mathbb{E}[\mathbb{E}[\theta | Q + u, S^M]] = \theta^* \\ \mathbb{E}[P^M] &= \frac{1}{2}\tilde{P} + \frac{1}{2}\theta^* \end{aligned} \right\} \Rightarrow \tilde{P} = \theta^* \quad \text{and} \quad Q_\theta = \frac{(\theta - \theta^*)}{2\lambda_K} \quad (3.13)$$

The equilibrium coefficient λ_K is the regression coefficient with the following func-

tional form:

$$\left. \begin{aligned}
 \lambda_K &= \frac{\text{cov}(\theta, Q+u|\xi^M)}{\text{Var}[Q+u|\xi^M]} \\
 \text{Var}[Q+u|\xi^M] &= \text{Var}\left[\frac{\theta-\theta^*}{2\lambda_K} + u|\xi^M\right] \\
 &= \frac{1}{4\lambda_K^2} \text{Var}[\theta|\xi^M] + \sigma_u^2 \\
 \text{cov}(\theta, Q+u|\xi^M) &= \mathbb{E}[(\theta - \theta^*)(Q+u)|\xi^M] \\
 &= \mathbb{E}\left[\frac{(\theta-\theta^*)^2}{2\lambda_K} + u(\theta - \theta^*)|\xi^M\right] \\
 &= \frac{1}{2\lambda_K} \text{Var}[\theta|\xi^M]
 \end{aligned} \right\} \Rightarrow \lambda_K = \frac{\sqrt{\text{Var}[\theta|\xi^M]}}{2\sigma_u} \tag{3.14}$$

The variance of the asset liquidation value depends on the informational regime. Therefore, the expected variance of θ given the market-maker estimation of the informational regime is given by:

$$\mathbf{Var}[\theta|\xi^M] = \sigma_{\theta,0}^2 + (\sigma_{\theta,1}^2 - \sigma_{\theta,0}^2) \mathbb{E}[\xi|\xi^M] \tag{3.15}$$

Since ξ is a Bernoulli variable, $\mathbb{E}[\xi|\xi^M] \in [0, 1]$. Therefore, the expected variance of the asset liquidation value given the market-maker estimation of the regime of information is within $[\sigma_{\theta,0}, \sigma_{\theta,1}]$

If the market-maker identifies a day as belonging to a highly informative regime RI_1 , then using the Bayes rule we prove that it is actually the case $\xi = 1$. Therefore the

elasticity of market-maker's prices to the liquidity demand is maximal and equals to $\frac{\sigma_{\theta,1}}{2\sigma_u}$.

$$\begin{aligned}\mathbb{E}[\xi|\xi^M = 1] &= 1 \cdot \mathbf{P}[\xi = 1|\xi^M = 1] \\ &= \mathbf{P}[\xi^M = 1|\xi = 1] \frac{\mathbf{P}[\xi = 1]}{\mathbf{P}[\xi^M = 1]} \\ &= \frac{P_{\xi^M} \cdot P_{\xi}}{\mathbf{P}[\xi^M = 1]}\end{aligned}$$

$$\begin{aligned}\mathbf{P}[\xi^M = 1] &= \mathbf{P}[\xi^M = 1|\xi = 1]\mathbf{P}[\xi = 1] + \mathbf{P}[\xi^M = 1|\xi = 0]\mathbf{P}[\xi = 0] \\ &= \mathbf{P}[\xi^M = 1|\xi = 1]\mathbf{P}[\xi = 1] \\ &= P_{\xi^M} \cdot P_{\xi}\end{aligned}$$

$$\begin{aligned}\Rightarrow \mathbb{E}[\xi|\xi^M = 1] &= 1 \quad \text{and} \quad \mathbf{Var}[\theta|\xi^M = 1] = \sigma_{\theta,1}^2 \\ \Rightarrow \lambda_K &= \frac{\sigma_{\theta,1}}{2\sigma_u}\end{aligned}$$

$$\begin{aligned}\mathbb{E}[\xi|\xi^M = 0] &= 1 \cdot \mathbf{P}[\xi = 1|\xi^M = 0] \\ &= \mathbf{P}[\xi^M = 0|\xi = 1] \frac{\mathbf{P}[\xi = 1]}{\mathbf{P}[\xi^M = 0]} \\ &= \frac{(1 - P_{\xi^M}) \cdot P_{\xi}}{\mathbf{P}[\xi^M = 0]}\end{aligned}$$

$$\begin{aligned}\mathbf{P}[\xi^M = 0] &= \mathbf{P}[\xi^M = 0|\xi = 1]\mathbf{P}[\xi = 1] + \mathbf{P}[\xi^M = 0|\xi = 0]\mathbf{P}[\xi = 0] \\ &= (1 - P_{\xi^M}) \cdot P_{\xi} + (1 - P_{\xi})\end{aligned}$$

$$\begin{aligned}\Rightarrow \mathbb{E}[\xi|\xi^M = 0] &= \frac{(1 - P_{\xi^M}) \cdot P_{\xi}}{(1 - P_{\xi^M}) \cdot P_{\xi} + (1 - P_{\xi})} < 1 \\ \Rightarrow \mathbf{Var}[\theta|\xi^M = 0] &= \frac{(1 - P_{\xi})}{(1 - P_{\xi^M}) \cdot P_{\xi} + (1 - P_{\xi})} \sigma_{\theta,0}^2 + \frac{(1 - P_{\xi^M}) \cdot P_{\xi}}{(1 - P_{\xi^M}) \cdot P_{\xi} + (1 - P_{\xi})} \sigma_{\theta,1}^2\end{aligned}$$

If the market-maker does not recognize any shock or information that could influence

the variability of the asset value, $\xi^M = 0$, then the market-makers sets a price while taking into account the probability of a wrongly estimated regime. The variance of θ given ξ^M is then a probability-weighted average of $\sigma_{\theta,0}$ and $\sigma_{\theta,1}$ depending on the average number of occurrences of the informative days P_ξ and the accuracy of his estimate P_{ξ^M} .

3.B Micro price

Harris, 2013 details the theoretical framework in which the defined micro-price is indeed the expected price, resulting from the maximization of the linear demand utility function. If we assume that

- Demand and supply schedules are linear in the difference between potential trade prices and the unobserved true value
- The absolute values of the slopes of these schedules are equal
- Supply and demand are both equal to zero when the price is equal to the unobserved true value

then we can estimate the unobserved true value from the quoted prices and sizes by simply expressing the slopes of both schedules as a function of the market quote and the true value, V

$$\frac{Q_a(t)}{P_a(t) - V} = \frac{Q_b(t)}{V - P_b(t)}$$

The resulting estimate is the size-weighted average of the bid and ask prices where the bid is weighted by the ask size and the ask is weighted by the bid size:

$$\hat{V} = \frac{Q_b(t) P_a(t) + Q_a(t) P_b(t)}{Q_b(t) + Q_a(t)}$$

The linear supply and demand schedules that motivate the derivation of this estimate are easily derived from the maximization of an exponential utility function, which generally can serve as a local approximation to any utility function.

3.C Liquidity metrics

3.C.1 Volatility:

We measure the daily price volatility of a stock returns using the Garman and Klass, 1980 formula applied to 5min intervals and averaged on the day. Garman-Klass estimate of the volatility uses the open, high, low and close prices of the time window. This estimate is robust to micro-structure noise and very close in practice to more sophisticated ones. It allow us to study the timely shocks of volatility within a day without considering a rolling window that could smooth and delay the shock. The formula is given by:

$$\sigma_k^{\text{GK}}(d) = \sqrt{\frac{1}{N_d} \sum_{t=1}^{N_d} \frac{1}{2} \log \left(\frac{H_t^{d,k}}{L_t^{d,k}} \right)^2 - (2 \log(2) - 1) \log \left(\frac{C_t^{d,k}}{O_t^{d,k}} \right)^2} \quad (3.16)$$

where the indexation k refers to the stock. d to the calculation day. N_d to the number of 5 min intervals in the continuous trading session of day d . It equals to 102 for days where the open is not delayed. $O_t^{d,k}$, $H_t^{d,k}$, $L_t^{d,k}$, $C_t^{d,k}$ are respectively the open, high, low, close prices of the t^{th} 5min interval at day d of stock k

3.C.2 Bid-Ask spread:

We take the volume weighted average of the bid-ask spread just before each transaction weighted by the volume of the trade. if $\zeta_{k,d}$ is the set of stopping times where transactions occur, so the equation

$$\Psi_{\tau}^k = \frac{1}{V_{k,d}} \sum_{\tau \in \zeta_{k,d}} V_{\tau} \cdot \Psi_{\tau}^k \quad \forall \tau \in \zeta_{k,d} \quad \Psi_{\tau}^k = \frac{Ask_{\tau-}^k - Bid_{\tau-}^k}{\frac{1}{2} (Ask_{\tau-}^k + Bid_{\tau-}^k)} \quad (3.17)$$

Where $V_{k,d}$ is the continuous trading phase volume. $Bid_{\tau-}^k$ and $Ask_{\tau-}^k$ are respectively the bid and ask price just before the transaction at τ . This measure could be interpreted as the transaction cost of the first € and give insight on the cost of liquidity during market stress.

General Conclusion

The chapters of this thesis provide original contributions to various aspect of the literature on the stock market liquidity. It aims to better understand market frictions, particularly the ones supported by institutional investors, the price formation process, and the behavior of market participants. Although, we've studied the research questions extensively from different angles, we also thought of possible extension, enhancement, generalizations or further analysis that could complement our work.

In the first chapter, we estimate the profitability of one of the most implemented strategies nowadays in the asset management industry, namely factors or asset pricing anomalies. These strategies are expected to generate above average risk-premia but also involve more rebalancing and generate higher transaction costs compared to the traditional buy and hold market cap based strategies. Using ANcerno database containing the execution of a representative sample of institutional investors in the US, we aim to estimate the costs effectively paid by asset managers to replicate these strategies. However, we note nowadays that funds provide fine-tuned portfolios with particular risk-return profiles exposed to several factors simultaneously (Broeders and Jansen (2019)). Thus, estimating the implementation cost of the factors separately, would make less sense. Therefore, a possible extension of this paper is to estimate the cost of one unit of risk exposure to each factor. To do so, we propose to cross-reference mutual funds executions in ANcerno with a database containing funds holdings, in order to identify their transactions. Compute the funds' exposure to each risk factors, then with a second stage regression, estimate the cost of executing one unit of risk exposure for each factor.

In the second chapter, We provide a methodology to estimate transaction costs in a

crowded environment. Since investors' synchronous trading is difficult to measure before the start of the trading session. We train a Bayesian network to capture the dependencies between the market order flow imbalance and investors' historical trading decisions and use it to better predict orders' transaction cost. This modeling has demonstrated its value by recovering the missing value of order flow imbalance and enhancing the accuracy of the estimates. However, the same model could be improved further by adding more layers to the network. One of the intuitive generalizations of the model is to incorporate a temporal aspect. In this paper, observations are assumed independent and identically distributed. We can make the model dynamic by adding layers to the network to account for previous values of order flow imbalance in the estimation of the daily imbalance and consequently the implementation shortfall. This should improve the estimates because the order flow imbalance is positively auto-correlated.

Finally, in the third chapter, we model market-participants interaction in different regimes of uncertainty. We extend the Kyle model set-up by assuming that the variability of the risky asset is time dependent and changes with the level of uncertainty. As a result, the market-makers response to the aggregate liquidity demand is also function of the regime of uncertainty. The model provided in this paper, takes the simple case of 2 levels of uncertainty: low or standard where no major event that could impact financial market occurred and a high uncertain regime. However, market uncertainty can rise for micro- and macro-economic reasons. Macro-uncertainty, impacts the level of systematic risk in the market. It could be triggered for instance, by the adoption of new monetary or fiscal policies, the publication of government statistics, an exogenous shocks on the economy such as natural disasters or health crises, that leads to increased macro-volatility. Micro-events on the other hand, leads to changes in firms' idiosyncratic variables. It results from periodic financial reporting, earnings calls announcements, analyst forecasts, or news on the future prospects and latest achievements of the company. The model presented in this chapter can be easily generalized to a model with two price reactions depending on the type of uncertainty (micro or macro). In the empirical part of this paper, It will be interesting to compare market-maker's response to the Covid-19 outbreak, to firm specific events for example earning calls.

RÉSUMÉ

Cette thèse propose trois contributions originales, sous la forme d'articles autonomes, à la littérature sur les coûts de transaction, le processus de formation des prix et l'encombrement. Le premier chapitre de la thèse étudie de manière approfondie la rentabilité de l'une des stratégies les plus mises en œuvre aujourd'hui dans le secteur de la gestion d'actifs, à savoir les facteurs nommés ou les anomalies de prix des actifs. Ces stratégies sont censées générer une prime de risque supérieure à la moyenne, mais elles impliquent également davantage de transactions. Nous évaluons dans quelle mesure les frictions du marché constituent une limite à l'arbitrage pour ces stratégies, et quelles seraient leurs capacités à atteindre le seuil de rentabilité compte tenu des habitudes de négociation des investisseurs institutionnels. Le deuxième chapitre souligne le rôle important de la négociation synchrone des investisseurs institutionnels dans la prévision des coûts de transaction des ordres. Cependant, cette variable est difficile à connaître avant le début de la séance de négociation. Nous proposons une méthodologie pour estimer les coûts de transaction dans un environnement encombré, en utilisant un réseau bayésien qui saisit les dépendances du déséquilibre du flux d'ordres sur le marché et des décisions de négociation historiques des investisseurs, afin de mieux prédire le coût de transaction des ordres. Enfin, le troisième chapitre modélise la réponse des teneurs de marché à l'offre et à la demande globale de liquidité du marché dans différents régimes d'incertitude. Nous remettons ensuite en question les implications du modèle pendant la période très incertaine de l'épidémie de COVID-19.

MOTS CLÉS

Coûts de Transaction, Impact sur le marché, Processus de Formation des Prix, Investissement Factoriel

ABSTRACT

This paper-based thesis is composed of three autonomous chapters and contributes to the trading costs and price formation process literature. The first chapter of the thesis extensively studies the profitability of one of the most implemented strategies nowadays in the asset management industry, named factors or asset pricing anomalies. These strategies are expected to generate above average risk-premia but also involve more trading. We assess to what extent market friction constitutes a limit to arbitrage for these strategies, and what would be their break-even capacities taking the institutional investors' trading patterns. The second chapter highlights the important role of institutional investors' synchronous trading in predicting order transaction costs. However, this variable is difficult to know before the start of the trading session. We provide a methodology to estimate transaction costs in a crowded environment, using a Bayesian network that captures the dependencies of the market order flow imbalance and investors' historical trading decisions, to better predict order transaction cost. Finally, the third chapter models market-makers' response to the aggregate liquidity supply and demand of the market in different regimes of uncertainty. We then challenge the model implications during the highly uncertain period of the COVID-19 outbreak.

KEYWORDS

Trading Costs, Market Impact, Price Formation Process, Asset Pricing Anomalies, Factor-based Investments