



**HAL**  
open science

## Essays in Behavioral Finance

Hedi Benamar

► **To cite this version:**

Hedi Benamar. Essays in Behavioral Finance. Business administration. HEC, 2014. English. NNT : 2014EHEC0004 . tel-01082424

**HAL Id: tel-01082424**

**<https://pastel.hal.science/tel-01082424>**

Submitted on 13 Nov 2014

**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.

**ECOLE DES HAUTES ETUDES COMMERCIALES DE PARIS**  
**Ecole Doctorale « Sciences du Management/GODI » - ED 533**  
**Gestion Organisation Décision Information**

**« Essays in Behavioral Finance »**

THESE  
présentée et soutenue publiquement le 4 juillet 2014  
en vue de l'obtention du  
DOCTORAT EN SCIENCES DE GESTION  
Par

**Hedi BENAMAR**

**JURY**

<b>Président du Jury :</b>	<b>Monsieur Denis GROMB</b> Professeur INSEAD, Fontainebleau – France
<b>Directeur de Recherche :</b>	<b>Monsieur Thierry FOUCAULT</b> Professeur, HDR HEC Paris – France
<b>Rapporteurs :</b>	<b>Monsieur Maxime MERLI</b> Professeur des Universités EM Strasbourg – France  <b>Monsieur Sébastien POUGET</b> Professeur IAE, Université de Toulouse Capitole 1– France
<b>Suffragant :</b>	<b>Monsieur Laurent CALVET</b> Professeur HEC Paris – France

**Ecole des Hautes Etudes Commerciales**

**Le Groupe HEC Paris n'entend donner aucune approbation ni improbation aux  
opinions émises dans les thèses ; ces opinions doivent être considérées  
comme propres à leurs auteurs.**

# Remerciements

Je tiens tout d'abord à remercier mon directeur de thèse Thierry Foucault pour son soutien absolument fondamental durant ces années de thèse à HEC. En le cotoyant, j'ai eu la chance de bénéficier de sa grande curiosité intellectuelle, de la finesse de son raisonnement économique et de son expertise du monde de la recherche académique en Finance. Sans son aide, cette thèse n'aurait sans doute pas vu le jour.

J'adresse aussi mes plus sincères remerciements aux membres de mon jury de thèse. Merci infiniment, donc, à Maxime Merli et Sébastien Pouget d'avoir accepté d'être rapporteurs, et à Laurent Calvet et Denis Gromb d'avoir accepté d'être suffragants. J'ai souvent trouvé dans leurs travaux une source d'inspiration essentielle pour améliorer mon travail de recherche.

Je souhaiterais aussi remercier tous les membres du département de Finance d'HEC Paris, professeurs et doctorants, pour nos nombreuses discussions et pour l'ambiance particulièrement collégiale qui régnait au département. Les moments passés avec vous tous resteront des temps forts de mon passage à HEC. En particulier, j'ai une dette morale envers Ulrich Hege, Johan Hombert, Stefano Lovo, Ioanid Rosu et David Thesmar pour leurs commentaires essentiels sur mon travail, et leur soutien.

Je souhaiterais de plus exprimer ma gratitude envers la fondation HEC, l'Institut Europlace de Finance et le LABEX pour leur aide financière, et envers le courtier français anonyme qui m'a permis de développer mon travail de thèse.

Enfin, sur un plan plus personnel, je dois beaucoup à femme Elise, mon petit garçon Ulysse, ma famille et à mes amis. Ils ont plus que contribué à l'achèvement de cette merveilleuse aventure doctorale. Qu'ils trouvent en ces quelques lignes l'expression de ma gratitude la plus sincère.

# Contents

<b>Remerciements</b>	<b>3</b>
<b>Introduction</b>	<b>10</b>
<b>Résumé en français</b>	<b>16</b>
Chapitre premier . . . . .	16
Chapitre second . . . . .	19
Chapitre troisième . . . . .	23
<b>1 Chapter one</b>	<b>26</b>
1.1 Introduction . . . . .	28
1.2 Testable Hypothesis . . . . .	34
1.3 Data . . . . .	36
1.3.1 The Brokerage Dataset . . . . .	36
1.3.2 Variable Definition . . . . .	38
1.4 Methodology and Identification Strategy . . . . .	40
1.4.1 Trader+ . . . . .	40
1.4.2 Identification Strategy . . . . .	42
1.4.3 The Control Group . . . . .	44
1.5 Results . . . . .	47
1.5.1 Trading performance . . . . .	47
1.5.2 Order Choice . . . . .	49

1.5.3	Trading Volume . . . . .	50
1.6	Robustness Analyses . . . . .	53
1.6.1	Standard difference-in-differences estimation . . . . .	53
1.6.2	Placebo Treatment . . . . .	54
1.7	Alternative Explanations . . . . .	55
1.7.1	Dynamic Self-Selection . . . . .	55
1.7.2	Trading Speed . . . . .	57
1.7.3	Overconfidence . . . . .	58
1.8	Conclusion . . . . .	59
1.9	Figures . . . . .	64
1.10	Tables . . . . .	76
<b>2</b>	<b>Chapter Two</b>	<b>86</b>
2.1	Introduction . . . . .	88
2.2	Methodology . . . . .	92
2.2.1	The brokerage dataset . . . . .	92
2.2.2	Investors' round-trips . . . . .	94
2.3	Results . . . . .	97
2.3.1	Ranking investors by trading performance . . . . .	97
2.3.2	Differences in trading behavior across investors . . . . .	99
2.3.3	Is there persistence in trading performance? . . . . .	102
2.4	Limits-to-arbitrage and liquidity provision . . . . .	104
2.5	Conclusion . . . . .	107
2.6	Figures . . . . .	110
2.7	Tables . . . . .	114
<b>3</b>	<b>Chapter Three</b>	<b>124</b>
3.1	Introduction . . . . .	126

3.2	Data . . . . .	130
3.2.1	The brokerage dataset . . . . .	130
3.2.2	Earnings Announcements . . . . .	131
3.3	Methodology . . . . .	132
3.3.1	Event Study . . . . .	132
3.3.2	Investors' round-trips . . . . .	134
3.4	Results . . . . .	138
3.4.1	Abnormal trading volume and participation around earnings announcements . . . . .	138
3.4.2	Individual trading performance and trading patterns around earnings announcements . . . . .	139
3.5	Conclusion . . . . .	143
3.6	Figures . . . . .	146
3.7	Tables . . . . .	150
	<b>Executive Summary</b>	<b>158</b>

# List of Tables

1.1	Summary statistics of the brokerage dataset . . . . .	77
1.2	Estimation of the probability of using Trader+ at least once after June 2003 (using pre-treatment data only) . . . . .	78
1.3	Comparisons between treated investors and control investors obtained via propen- sity matching . . . . .	79
1.4	Summary statistics on investors' trading channels and investors' order choice . .	80
1.5	Does the display format of market data matters for retail investors? . . . . .	81
1.6	Robustness: standard difference-in-difference estimation . . . . .	82
1.7	Robustness: placebo test . . . . .	83
1.8	Heterogeneous treatment effects . . . . .	84
1.9	Dependent variables . . . . .	85
2.1	Summary statistics of the brokerage dataset . . . . .	115
2.2	Identification of Trading Strategies . . . . .	116
2.3	Sharpe ratios summary statistics . . . . .	116
2.4	Trading performance and liquidity provision . . . . .	117
2.5	Order choice . . . . .	118
2.6	Contrarian behavior . . . . .	119
2.7	Round-trips durations . . . . .	120
2.8	Performance Persistence . . . . .	121
2.9	Trading performance and limits-to-arbitrage I . . . . .	122

2.10 Trading performance and limits-to-arbitrage II . . . . .	123
3.1 Summary statistics of the brokerage dataset . . . . .	151
3.2 Summary Statistics - Sample . . . . .	152
3.3 Summary Statistics - Announcement Time . . . . .	153
3.4 Identification of Trading Strategies . . . . .	154
3.5 Round-Trips . . . . .	155
3.6 Abnormal Trading Activating around Earnings Announcements . . . . .	156
3.7 Trading performance around Earnings Announcements . . . . .	157

# List of Figures

1.1	Investors' decision tree . . . . .	65
1.2	Main trading channels used by the broker customers before Trader+ . . . . .	66
1.3	Screenshot of Trader+ . . . . .	67
1.4	Orders submitted by treated and matched control investors, 1999-2010 . . . . .	68
1.5	Monitoring activity and performance . . . . .	69
1.6	Monitoring activity and performance: market orders vs. limit orders . . . . .	70
1.7	Monitoring activity and order choice . . . . .	71
1.8	Monitoring activity and trading volume . . . . .	72
1.9	Monitoring activity and trading horizon . . . . .	73
1.10	Dynamics of entry into treatment . . . . .	74
1.11	Causes before consequences, or consequences before causes? . . . . .	75
2.1	Aggregate Net Imbalance: Top Decile . . . . .	111
2.2	A tiny market maker . . . . .	112
2.3	Abnormal returns dynamics over the last decade . . . . .	113
3.1	Abnormal trading activity and abnormal market participation around earnings announcements . . . . .	147
3.2	Profit-Taking patterns on and after the announcement date . . . . .	148
3.3	Net Individual Imbalance around earnings announcements . . . . .	149

# Introduction

In the traditional finance paradigm, the homo economicus is meant to be rational: he correctly updates and processes new information according to Bayes' rule (rational beliefs), and he makes decisions that are consistent with the Subjective Expected Utility model defined by Savage more than 50 years ago (rational preferences) ([Barberis and Thaler \[2003\]](#)).

However, over the last decades the growing behavioral finance literature has consistently pointed out that the fundamental axioms of the Expected Utility (EU) paradigm generate empirical predictions that do not appear to be consistent with the data.

[Barberis and Thaler \[2003\]](#) recall that, to address these difficulties, the Behavioral Finance literature uses two “building blocks”. The first one is the relaxation of some problematic assumptions of the rational paradigm. For example, if processing information is costly for individuals, then their financial decisions may be more consistent with bounded-rationality than full rationality ([Simon \[1955\]](#), [Barber and Odean \[2007\]](#)). In other words, psychology can explain the observed deviations from the normative rational models. In my thesis, I consider two possible deviations from the rational setup in the first chapter (bounded rationality) and third chapter (disposition effect).

The second building block is limits to arbitrage (see [Shleifer and Vishny \[1997\]](#), [Gromb and Vayanos \[2002\]](#) and [Gromb and Vayanos \[2010\]](#)). Limits-to-arbitrage expresses the fact that prices deviate from their fundamental “true” value because of the actions of irrational agents (“noise traders”) in the market. In this context, arbitrage opportunities may take a lot of time

and may be very costly to be exploited by rational arbitrageurs. I consider some important implications of the limits-to-arbitrage paradigm for retail investors in the second chapter of my thesis.

Broadly speaking, understanding the financial decisions of individuals is fundamental, for at least two reasons according to [Barberis and Thaler \[2003\]](#). First, today more and more individuals are investing in equities because of lowering costs of entry into the market and second, the recent reforms in the pension sector imply that individuals are more and more responsible for their financial well being at retirement time. Therefore, perhaps not surprisingly, to understand how individuals use financial instruments and make financial decisions, Household Finance has quickly emerged as a new and very dynamic field ([Campbell \[2006\]](#)).

Often, the literature in this field has focused on general trading activity and portfolio choices made by individuals. In particular researchers have already established several links between particular psychological traits of investors and their trading behavior. For instance, overconfidence reduces and self-monitoring increases trading performance ([Biais et al. \[2005\]](#)), retail investors are prone to the disposition effect ([Barberis and Xiong \[2009\]](#), [Boolell-Gunesh, Broihanne, and Merli \[2012\]](#)) and confirmation bias among investors can generate bubbles, crashes, and reversals ([Pouget and Villeneuve \[2012\]](#)). In the same vein, individual variables such as IQ ([Grinblatt, Keloharju, and Linnainmaa \[2011\]](#)), smartness ([Korniotis and Kumar \[2013\]](#)) and financial sophistication ([Boolell-Gunesh, Broihanne, and Merli \[2009\]](#), [Calvet and Sodini \[2014\]](#)) are strongly associated with more efficient and better financial decisions at the individual level.

Other papers have more focused on the welfare implications of those “investment mistakes” ([Barber, Lee, and Odean \[2006\]](#)). The usual stylized facts are that individuals are on average insufficiently diversified ([Jacobs, Muller, and Weber \[2013\]](#)), trade too much and do not participate enough in the financial market. However, this average behavior hides in reality a very large herogeneity in trading skills among the retail population. For instance, using Swedish government records covering all Swedish households between 1999 and 2002, [Calvet, Camp-](#)

[bell, and Sodini \[2007\]](#) argue that while a few households are indeed poorly diversified, most of them are through international diversification. Similarly, while [Barber and Odean \[2000\]](#) argue that active retail investors trade too much to their detriment, [Linnainmaa \[2011\]](#) suggest that this intense trading activity hides some important learning mechanisms.

In this context, another interesting vein of research has recently appeared. In their “questions for further research” chapter, [Parlour and Seppi \[2008\]](#) argue that understanding how individuals choose between using market or limit orders when trading is a “useful window through which to observe investor heterogeneity” such as private trading motives or urgency for trading. This new approach is particularly interesting because nowadays most of equity and derivatives exchanges around the world (such as NYSE-Euronext, the Nasdaq or Xetra) are electronic limit order markets, which allow the submission of market orders and limit orders.

Limit orders specify a number of shares and a price. Because of their pre-specified price, they must be matched with opposite market orders in order to be executed. In the meantime they are stored with other unexecuted limit orders in the limit order book, which they will exit if they are hit by subsequent orders or cancelled by the order submitter. Market orders specify a quantity, but no particular price. They are matched with the best opposite limit orders, in terms of price, available at the time of their submission and with respect to time and price priority rules.

In this new literature, [Linnainmaa \[2010\]](#) is a pioneer work. He reaches the important conclusions that the use of limit orders by individual can alter the inferences on individual’s skill that researchers derive from the data. Similarly, [Kelley and Tetlock \[2013\]](#) study the role of retail investors in stock pricing by separately examining the performance of retail executed market orders vs. retail executed limit orders, and obtain different implications about the information content of each order type.

The scope of my thesis was to understand more deeply which important factors can explain the individual financial decisions. Using a brokerage dataset which contains more than 15

million investor trades at the daily level from 1999 to 2010, I have had the opportunity to contribute a bit to this fascinating literature.

# References

- Barber, B., Lee Y. and Liu, Y., and Odean, T. (Nov. 2006). “Just How Much Do Individual Investors Lose by Trading?” *Review of Financial Studies* 22.2.
- Barber, B. and Odean, T. (Dec. 2007). “All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors”. *Review of Financial Studies* 21.2.
- Barber, B. and Odean, T. (2000). “Trading is hazardous to your wealth: The common stock investment performance of individual investors”. *The Journal of Finance* 55.2.
- Barberis, N. and Thaler, R. (2003). “A survey of behavioral finance”. *Handbook of the Economics of Finance*. Ed. by G. M. Constantinides, M. Harris, and R. Stulz.
- Barberis, N. and Xiong, W. (2009). “What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-Based Explanation”. *The Journal of Finance* June.
- Biais, B., Hilton, D., Mazurier, K., and Pouget, S. (2005). “Judgemental Overconfidence, Self-Monitoring, and Performance in an Experimental Financial Market”. *Review of Economic Studies* 72.2.
- Booell-Gunesh, S., Broihanne, M., and Merli, M. (2009). “Disposition effect, investor sophistication and taxes: Some French Specificities”. *Finance* 30.
- (2012). “Sophistication of Individual Investors and Disposition Effect Dynamics”. *Finance* 33.
- Calvet, L. E., Campbell, J. Y., and Sodini, P. (Oct. 2007). “Down or Out: Assessing the Welfare Costs of Household Investment Mistakes”. *Journal of Political Economy* 115.5.

- Calvet, L. E. and Sodini, P. (2014). "Twin Picks : Disentangling the Determinants of Risk-Taking in Household Portfolios". *The Journal of Finance*.
- Campbell, J. (2006). "Household Finance". *Journal of Finance* 61.4.
- Grinblatt, M., Keloharju, M., and Linnainmaa, J. (2011). "IQ, trading behavior, and performance". *Journal of Financial Economics*.
- Gromb, D. and Vayanos, D. (2002). "Equilibrium and welfare in markets with financially constrained arbitrageurs". *Journal of financial Economics* 66.
- (Dec. 2010). "Limits of Arbitrage". *Annual Review of Financial Economics* 2.1.
- Jacobs, H., Muller, S., and Weber, M. (Aug. 2013). "How should individual investors diversify? An empirical evaluation of alternative asset allocation policies". *Journal of Financial Markets*.
- Kelley, K. and Tetlock, P. (2013). "How wise are crowds? Insights from retail orders and stock returns". *The Journal of Finance* January.
- Korniotis, G. M. and Kumar, A. (Jan. 2013). "Do Portfolio Distortions Reflect Superior Information or Psychological Biases?" *Journal of Financial and Quantitative Analysis* 48.01.
- Linnainmaa, J. (2010). "Do limit orders alter inferences about investor performance and behavior?" *The Journal of Finance* LXV.4.
- (2011). "Why Do (Some) Households Trade So Much?" *Review of Financial Studies* 24.5.
- Parlour, C. A. and Seppi, D. J. (2008). "Limit Order Markets : A Survey". *Handbook of Financial Intermediation and Banking* 5.33. Ed. by A. V. Thakor and A. W. A. Boot.
- Pouget, S. and Villeneuve, S. (Mar. 2012). "A Mind is a Terrible Thing to Change: Confirmation Bias in Financial Markets". 720.
- Shleifer, A. and Vishny, R. (1997). "The limits of arbitrage". *The Journal of Finance* LII.1.
- Simon, H. (1955). "A behavioral model of rational choice". *The Quarterly Journal of Economics* 69.1.

# Résumé en français

## Chapitre premier

### Motivation

Les régulateurs sont de plus en plus préoccupés non seulement par la quantité d'information financière accessible aux investisseurs individuels, mais aussi par *la façon* avec laquelle cette information est affichée, car l'on sait désormais que le format d'affichage peut influencer les décisions financières individuelles. <sup>1</sup>

Dans le premier chapitre, je teste si le *format d'affichage* des données de marché utilisées par les investisseurs individuels affecte leur performance boursière. Cette question me semble importante car il est fondamental d'identifier quels sont les déterminants des décisions financières des ménages (Campbell [2006]; Calvet, Campbell, and Sodini [2007]). Et, en particulier, la grande hétérogénéité observée dans les données en termes de performance boursière des ménages reste encore à expliquer.

Pour comprendre pourquoi le format d'affichage des données de marché peut avoir son importance pour les investisseurs individuels, rappelons que lorsque les investisseurs ont choisi une allocation de portefeuille, ils doivent mettre en œuvre les stratégies de trading nécessaires

---

<sup>1</sup>Voir Benartzi and Thaler [1999], Saez [2009], Looney and Hardin [2009] et Kaufmann, Weber, and Haisley [2012].

pour rééquilibrer leurs inventaires. Ces stratégies de trading impliquent nécessairement un choix entre les ordres de marché et les ordres à cours limité.

Si les investisseurs utilisent des ordres à cours limité, ils doivent alors surveiller le marché, après la soumission de leur ordre, afin de réviser ou annuler leur ordre dans le but d'atténuer les problèmes de sélection adverse et les risques de non-exécution qui découlent de l'utilisation d'ordres de ce type.

En d'autres termes, les investisseurs doivent surveiller le marché et leurs ordres à cours limité pour améliorer leur performance boursière sur ces ordres. L'hypothèse de travail de mon premier chapitre est donc la suivante: si un format d'affichage des données de marché rend cette *de surveillance du marché plus efficace*, toute chose égale par ailleurs, il devrait également améliorer la performance boursière des ordres à cours limité des individus.

## Méthodologie

Pour tester cette hypothèse, j'exploite un contexte expérimental assez unique en son genre dans lequel la quantité d'informations de marché accessible aux investisseurs est restée inchangée, alors que son *format d'affichage* a subi un changement exogène à un moment donné. Plus précisément, j'utilise l'introduction en Juin 2003 par une grande maison de courtage française d'un logiciel de trading (ci-après de Trader +). Le logiciel permettait d'afficher les données de marché d'une manière plus efficace car il affichait simultanément tous les items d'information pertinents (données du marché et les ordres en cours des investisseurs) sur un écran personnalisé par l'utilisateur. La présentation simultanée de ces items permet aux investisseurs de mieux comprendre comment les mouvements des prix des actions sont liés les uns aux autres et devrait aider les investisseurs à mieux assimiler cette information lors de la surveillance de leurs ordres à cours limité.

Vu que le logiciel affiche ces données de marché plus efficacement qu'auparavant, je m'attends donc à ce que les utilisateurs de Trader + "monitorent" mieux leurs ordres à cours limité, ce

qui devrait avoir pour effet d'augmenter leur performance boursière sur ces ordres. Afin de tester ces implications, j'utilise un estimateur par différence-de-différences où un groupe de traitement, contenant les individus qui utilisent le logiciel au moins une fois, sont appariés (avec un algorithme par score de propension) avec des investisseurs contrôle similaires qui, eux, n'utilisent jamais le logiciel.

## Résultats

Je montre que le format d'affichage des données de marché a un impact fort sur le comportement des investisseurs individuels. Comme dans [Grinblatt, Keloharju, and Linnainmaa \[2011\]](#) et [Linnainmaa \[2010\]](#), pour chaque ordre d'achat et de vente, je calcule une mesure de performance qui a souvent été utilisée dans la littérature comme un proxy pour le risque de sélection adverse encouru par les investisseurs qui soumettent des ordres à cours limité. Mes résultats suggèrent un effet positif de Trader+ de l'ordre de 10 points de base sur les returns intraday des ordres à cours limités.

L'ampleur économique de ce résultat est plutôt importante à la lumière des études antérieures sur la performance boursière des investisseurs. Par exemple [Grinblatt, Keloharju, and Linnainmaa \[2011\]](#) montrent que les returns intradays des investisseurs finlandais ayant un QI élevé (dans le top 5 % de la distribution) surperforment les personnes à faible QI d'environ 11 points de base, ce qui suggère que le format d'affichage des données de marché a un effet sur la performance boursière similaire aux capacités cognitives individuelles.

Dans le premier chapitre, j'explique cette augmentation de la performance boursière en proposant que l'amélioration du format d'affichage dûe à Trader + permet aux individus de surveiller le marché plus efficacement et, par conséquent, de placer/réviser/annuler leurs ordres à cours limités de façon plus optimale. Cette augmentation de la performance a deux conséquences, si l'on se réfère aux les modèles classiques de la microstructure des marchés.

Premièrement, les investisseurs "traités" devraient utiliser les ordres à cours limités plus souvent qu'auparavant, car il est plus facile de gérer ces ordres avec Trader+. C'est ce que je trouve dans les données. Deuxièmement, je m'attends à ce que les investisseurs traités soient en mesure de repérer des opportunités de trading à court terme qui étaient probablement trop difficiles à saisir avant. Ce changement de comportement devrait conduire à une diminution de l'horizon de trading boursier individuel, défini comme le temps nécessaire pour déboucler une ligne de trading ouverte. Cette implication est aussi confirmée dans les données.

En résumé, les changements de comportement boursier observés dans le groupe traité, après l'introduction du logiciel, sont compatibles avec une explication basée sur les coûts cognitifs des l'activité de surveillance du marché.

## Chapitre second

### Motivation

Dans le deuxième chapitre, j'explique tout d'abord que des études très récentes suggèrent que les investisseurs individuels utilisent les ordres à cours limité afin d'être apporteurs de liquidités pour les investisseurs institutionnels qui ont besoin de déboucler leurs positions rapidement ([Kaniel, Saar, and Titman \[2008\]](#), [Kelley and Tetlock \[2013\]](#)). Dans ce contexte, on s'attend donc à ce que les investisseurs institutionnels fassent des concessions sur le prix d'échange boursier afin d'inciter les investisseurs individuels à remplir leur rôle de contrepartie ([Grossman and Miller \[1988\]](#), [Campbell, Grossman, and Wang \[1993\]](#)).

Cependant, il n'est pas clair dans la littérature actuelle si les individus bénéficient réellement, *après coûts de transaction*, de l'utilisation de ces stratégies de trading. Par exemple, [Kelley and Tetlock \[2013\]](#) reconnaissent que “ *[leurs] données ne contiennent pas les portefeuilles des individus, ou même les coûts de transaction encourus, ce qui exclut toute analyse directe de la performance brute ou nette des investisseurs particuliers dans [leur] échantillon*”. Cette question

est néanmoins importante car il est difficile d'expliquer la forte hétérogénéité en termes de performance boursière des individus qui a été identifiée dans la littérature ([Grinblatt, Keloharju, and Linnainmaa \[2011\]](#), [Barber et al. \[2013\]](#)). En outre, du point de vue du bien-être des investisseurs, il est fondamental de savoir si les investisseurs particuliers s'engagent ou non dans des stratégies de trading qui les conduisent à des lourdes pertes financières.

L'objectif de mon deuxième chapitre est donc de tester directement *si* ces stratégies apportées de liquidités peuvent être une source de profits éventuelle pour les investisseurs particuliers actifs qui les implémentent. Pour ce faire, j'exploite deux caractéristiques uniques de mes données.

Tout d'abord, les données contiennent la majeure partie des frais de transaction payés par les investisseurs particuliers. Deuxièmement, les investisseurs actifs dans ma base semblent se comporter comme des "teneurs de marché" au sens traditionnel du terme: ils utilisent beaucoup d'ordres à cours limité et ont des périodes de détention des actifs financiers plus courtes que les autres investisseurs. Par conséquent, en me concentrant sur les investisseurs les plus actifs de mon échantillon, je considère en quelque sorte la population idéale pour comprendre si être apporteur de liquidités, pour un individu, peut être profitable ou pas.

## **Méthodologie**

Pour aborder cette question de recherche, la difficulté principale est de bien mesurer la rentabilité des stratégies de trading individuelles, ce qui est loin d'être évident. Par exemple, étant données les grandes disparités en termes d'horizons de trading, de motivations pour trader et en termes de stratégies employées, les chercheurs sont souvent forcés de choisir des règles ad-hoc arbitraires comme évaluer la performance d'un trade sur un horizon fixé par le chercheur (jour, semaine, mois). Ces choix peuvent potentiellement masquer de nombreuses informations importantes sur la performance réelle des individus.

Afin de tenter de surmonter ces difficultés, je tire profit de la structure en panel de mon jeu de données en adaptant une méthodologie décrite initialement par [Schlarbaum, Lewellen, and Lease \[1978\]](#). La méthode, simple et puissante, repose sur l'identification des aller-retours effectivement implémentés et exécutés par les investisseurs. C'est une approche "hard-data" qui ne fait que très peu d'hypothèses: les données parlent pour les individus.

## Résultats

Les individus peuvent-ils générer des profits en utilisant des stratégies apportées de liquidités? Je classe les investisseurs dans mon échantillon selon les ratios de Sharpe de leurs stratégies de trading et je trouve que seuls les investisseurs classés dans les trois déciles supérieurs en termes de performance sont en mesure d'obtenir des rendements positifs anormaux *-déduction faite des frais-* en fournissant de la liquidités au marché. Un fait intéressant ressort de l'étude: ces top-traders actifs semblent être beaucoup plus contrariants que les autres groupes de traders actifs. Ce comportement suggère que ces retails se comportent bien comme des teneurs de marché en achetant des actions quand il y a une pression à la vente et en vendant quand il y a une pression à l'achat.

Ces résultats confirment certaines études précédentes, comme celle de [Grinblatt, Keloharju, and Linnainmaa \[2011\]](#) qui documente que les individus ayant un QI élevé en Finlande sont plus contrariants que les investisseurs ayant un QI faible, sans toutefois pouvoir trancher (faute de pouvoir utiliser les coûts de transactions), s'il s'agit d'un comportement rationnel ou pas. Ils affirment qu' en vendant des actions près des maxima mensuels et en achetant près des minima mensuels, les investisseurs au QI élevé dans leur échantillon suivent probablement une stratégie de trading rationnelle apportée de liquidités.

Afin de dissocier la chance des compétences réelles de trading, je teste ensuite si les performances individuelles sont persistantes. J'obtiens un niveau significatif de persistance au cours du temps, en ligne avec les résultats présentés dans d'autres études récentes. Il semble donc

probable que certains investisseurs dans mon échantillon ont des compétences des trading authentiques.

Ensuite, j'étudie si certaines frictions particulières des marchés financiers permettent à ces top traders de générer des profits après coûts de transaction. En effet, il peut sembler surprenant qu'un sous-ensemble d'investisseurs individuels, aussi doués soient-ils, est en mesure de saisir des opportunités d'arbitrage *avant* d'autres arbitrageurs plus sophistiqués, tels que les fonds spéculatifs ou des investisseurs institutionnels. Une possibilité est que les investisseurs les plus performants dans mon sample fournissent de la liquidité lorsque les arbitragistes institutionnels sont plus contraints et donc pas en mesure d'exploiter les opportunités d'arbitrage disponibles.

Je fournis dans mon deuxième chapitre des tests empiriques en faveur de cette explication liée au concept de "limites-à-l'arbitrage" ([Shleifer and Vishny \[1997\]](#)) de deux façons différentes. En particulier, pour chaque aller-retour dans mon échantillon je récupère le niveau de volatilité intraday des actions négociées le jour où cet aller-retour donné a été *initié*.

En effet, en conditionnant par la date de départ d'un aller-retour il est probable que j'identifie les moments exacts où les individus cherchent activement à exploiter une opportunité de trading. J'utilise la volatilité intraday car elle est généralement utilisée comme un indicateur du risque d'arbitrage: plus la volatilité est élevée, plus le risque d'arbitrage est fort ([Pontiff \[2006\]](#) et [Lam and Wei \[2011\]](#)) et plus la liquidité est faible ([Hameed, Kang, and Viswanathan \[2010\]](#) et [Nagel \[2012\]](#)).

Je montre que les retails les plus performants initient leurs aller-retours les plus rentables lorsque le risque d'arbitrage est particulièrement élevé. De façon intéressante, cette relation positive entre la volatilité intraday et la rentabilité future ne tient pas pour les retails les moins performants du sample, ce qui suggère que seuls les traders les plus doués sont en mesure d'exploiter les moments où le marché manque de liquidités.

# Chapitre troisième

## Motivation

Le point de départ du troisième chapitre est qu'il existe toute une littérature qui montre que le flux d'ordres agrégé des investisseurs individuels est à même de prédire les rendements futurs des actions autour des annonces de résultats (Kaniel, Liu, et al. [2012], Kelley and Tetlock [2013]).

Hélas, ces études exploitent des données de courtage qui ne permettent pas de suivre un même investisseur au fil du temps. Autrement dit, les auteurs des deux études ci-dessus ne peuvent tout simplement pas identifier les stratégies qui sont effectivement implémentées par des investisseurs individuels avant, pendant et après l'annonce.

Cependant, l'identification des stratégies de trading mises en œuvre par les individus autour de l'annonce est importante, pour au moins deux raisons.

Tout d'abord, identifier les stratégies de trading et évaluer leur performance peut combler le fossé entre les découvertes récentes de Kelley and Tetlock [2013] et Kaniel, Liu, et al. [2012], qui apportent une lumière positive sur les compétences de trading des individus, et la littérature précédente qui, généralement, dépeint les investisseurs particuliers comme "bruiteurs naïfs".

En particulier, le fait que le flux d'ordres agrégé des retails a un certain pouvoir prédictif sur les futurs rendements boursiers autour des annonces de résultats n'implique pas nécessairement que certains investisseurs particuliers sont en mesure de réaliser des gains de trading autour de l'annonce (après coûts de transactions et ajustement pour le risque).

Deuxièmement, identifier les stratégies qui sont mises en œuvre autour de l'annonce peut aider à comprendre l'origine de plusieurs puzzles empiriques qui ont été décrits dans la littérature. Par exemple, à la fois le "*earnings announcement premium*" (Frazzini and Lamont [2007]) et le "*earnings announcement drift*" (Taylor [2010] et Kaniel, Liu, et al. [2012]) semblent être

dûs (du moins en partie) au comportement boursier des investisseurs individuels. Le problème, comme souvent, est que les données disponibles ne permettent pas aux auteurs de ces études de suivre les investisseurs individuels au cours du temps, et ainsi de confirmer que certaines stratégies implémentées par les retails peuvent effectivement exercer suffisamment de pression sur les prix pour générer les deux patterns ci-dessus.

## Méthodologie

Le but de ce chapitre est donc d'identifier directement les stratégies de trading mises en œuvre par les investisseurs individuels autour des annonces de résultats, et d'évaluer leur performance en contrôlant pour les effets fixes individuels.

J'adopte dans ce chapitre une méthodologie standard "d'évent study" ([MacKinlay \[1997\]](#), [Frank and Goyal \[2007\]](#)) et pour évaluer la rentabilité des stratégies de trading exécutés par des individus, j'adopte la méthodologie décrite dans [Schlarbaum, Lewellen, and Lease \[1978\]](#).

## Résultats

J'obtiens trois principaux résultats. Tout d'abord, je trouve que les investisseurs augmentent fortement à la fois la participation au marché et le volume des transactions avant, pendant et peu de temps après l'annonce. Ce résultat est cohérent avec les papiers de [Vieru et al. \[2006\]](#), [Taylor \[2010\]](#) et [Etheber, Hennen, and Meyer \[2012\]](#).

Deuxièmement, je montre que les investisseurs tentent activement d'exploiter une information privilégiée avant l'annonce en exécutant des round-trips à horizon très court terme autour de l'annonce. Je trouve en effet que pendant les quelques jours *qui précèdent* le jour de l'annonce, la durée observée (en jours) des allers-retours exécutés par des investisseurs individuels dans mon échantillon diminue fortement, ce qui suggère que les investisseurs individuels débouclent leurs positions peu de temps après l'annonce afin d'encaisser leurs gains.

Troisièmement, je montre que les allers-retours commencés un jour avant l'annonce sont plus rentables que ceux initiés en temps normal. Ce n'est pas le cas, en revanche, pour les allers-retours qui sont initiés *le jour même* et les jours qui suivent l'annonce. Cette différence de performance suggère que les investisseurs individuels peuvent profiter de certaines informations privées avant l'annonce officielle des résultats trimestriels, mais ils sont moins en mesure de réagir à temps et d'exploiter de l'information qui vient d'être diffusée le jour de l'annonce.

Ce comportement boursier peut, potentiellement, ralentir l'ajustement des prix boursiers aux nouvelles informations. Plus précisément, je trouve que les investisseurs sont fortement contrariants par rapport au *return intraday* du jour de l'annonce: ils vendent massivement quand il y a un return intraday très positif et ils achètent massivement quand il y a un return intraday très négatif le jour de l'annonce. Ce comportement peut se comprendre comme la conséquence directe d'une prise de bénéfices après l'annonce.

# 1 Chapter one

## **To See is To Know: Efficient Display of Market Data for Retail Investors**

I test whether the *display format* of market data affects the trading performance of retail investors. To do so, I exploit a large brokerage dataset covering a period during which the market information provided to the broker customers changed in format, but not in content. I find that a more efficient information display allows investors to increase returns on their limit orders, because it becomes easier for them to mitigate the risk of adverse selection when trading with those orders. Hence, the display format of market data matters for the individual investor.

## 1.1 Introduction

Regulators are increasingly concerned with not only how much financial information is provided to investors, but also *how* it is displayed to them, because the form of presentation can have first-order effects on individuals' investment choices.<sup>1</sup> For instance, in 2009, the Securities and Exchange Commission adopted a disclosure framework for mutual funds investment opportunities "*that is easier to use and more readily accessible*". Similarly, the European Commission's PRIIPS (2012) proposal was designed to "*improve the quality of information that is provided to consumers when considering investments*" using a format "*that is short and plain-speaking, and thus far more consumer-friendly*".<sup>2</sup>

In this paper, I test whether the *display format* of market data provided to retail investors affects their trading performance. This question is important because it is fundamental to identify the determinants of households' financial decisions (Campbell [2006]; Calvet, Campbell, and Sodini [2007]) and, in particular, the large observed heterogeneity in retail trading performance remains to be explained.<sup>3</sup> To understand why the display format of market data might matter for retail investors, recall that once investors have chosen a portfolio allocation, they must implement the trading strategies required to rebalance their holdings. These trading strategies involve a choice between market and limit orders.<sup>4</sup> If investors use limit orders, then they monitor and revise/cancel them, after submission, to mitigate adverse selection and non-execution risks.<sup>5</sup> That is, investors should monitor the market and their limit orders to improve their

---

<sup>1</sup>See Benartzi and Thaler [1999], Saez [2009], Looney and Hardin [2009] and Kaufmann, Weber, and Haisley [2012].

<sup>2</sup>See [www.sec.gov/rules/final/2009/33-8998.pdf](http://www.sec.gov/rules/final/2009/33-8998.pdf) and [europa.eu/rapid/press-release\\_IP-12-736\\_en.pdf](http://europa.eu/rapid/press-release_IP-12-736_en.pdf).

<sup>3</sup>Prior studies have shown that financial literacy (Lusardi and Mitchell [2011]; Bhattacharya et al. [2012]), financial sophistication (Calvet, Campbell, and Sodini [2007]) and financial wealth (Calvet and Sodini [Forthcoming]) are important determinants of the financial decisions of individuals. See also Barber, Lee, et al. [2012] and Grinblatt, Keloharju, and Linnainmaa [2011] for specific studies on investors' performance heterogeneity

<sup>4</sup>Limit orders specify a number of shares and a price: the maximum price at which an investor is willing to buy or the minimum price at which an investor is willing to sell. Limit orders accumulate in the limit order book in descending buy-price order or ascending sell-price order (price priority) and join the queue composed of other limit orders that have the same price (time priority). Market orders specify a number of shares to buy or sell, but no particular price. Market orders are automatically filled at the most attractive price posted by previous limit orders in the limit order book.

<sup>5</sup>Hence, limit orders must be monitored but market orders do not. "*After all, providing limit orders is, in fact, not*

trading performance on those orders. If a display format of market data makes such *market monitoring* more efficient then, everything else being constant, it should also improve investors' trade performance on their limit orders.

To test this hypothesis, I exploit a unique setting in which the quantity of market information provided to investors remained fixed, while the *display format* of that information exogenously changed at some point in time. Specifically, I use the introduction in June 2003, by a large brokerage house, of a trade order management software (hereafter, Trader+). The software displayed market data in a more efficient way because it *simultaneously* gathered all relevant information items (market data and investors' orders) into a user-customized screen. The simultaneous presentation of information items allows investors to understand how different stock prices movements are related to one another and helps investors better assimilate that information when monitoring their limit orders. For instance [Hodge, Kennedy, and Maines \[2004\]](#), who study a search-facilitating technology introduced by the SEC in 2003 (and adopted in 2009), argue that the "*simultaneous presentation of related information directs users' attention toward examining relations among the information items (Russo [1977]). Simultaneous presentation also reduces the cognitive costs of integrating the information*".<sup>6</sup> Importantly, the software left the *quantity* of the data being processed unchanged. Indeed, the same information was available to the brokerage customers, in a more dispersed form, on other trading channels (such as using the brokerage website to submit an order). As the software displayed the market data more efficiently than before, I expect Trader+ users to better monitor their limit orders, which should subsequently lead to higher trading performance on those orders. To test these implications, I use a difference-in-differences identification strategy in which treated investors, who switch to the new software, are matched (with a propensity-score algorithm) to similar control investors who do not use the software.

---

*costless since it requires some monitoring to insure that orders are not left exposed after, for example, a public information release" (Glosten [1994]).*

<sup>6</sup>This is the eXtensible Business Reporting Language (XBRL) technology. In 2003 the Securities and Exchange Commission introduced this search-facilitating technology to ease the acquisition and the processing of public companies' financial statements by non-professional users. See <http://xbrl.sec.gov> for technical details and [Maines and McDaniel \[2000\]](#) and [Hodge, Kennedy, and Maines \[2004\]](#) for a discussion.

I find that the display format of market data does matter for the individual investor. As in [Grinblatt, Keloharju, and Linnainmaa \[2011\]](#) and [Linnainmaa \[2010\]](#), for each buy and sell limit order on a given stock in my sample, I compute the signed return from the execution price to the closing price of that stock the same day. Indeed, this return has been used in the literature as a proxy for the risk of adverse selection faced by investors when trading with limit orders. It is therefore a performance measure of one order's execution quality and should capture "*the active management of individuals*".<sup>7</sup> Intuitively, the higher the risk of adverse selection, the lower this return. My difference-in-differences estimates show that following the introduction of Trader+, investors' intraday returns on their limit orders jump by 10 basis points.

The economic magnitude of this result is large in light of previous studies on the trading performance of investors. For instance [Grinblatt, Keloharju, and Linnainmaa \[2011\]](#) find that intraday returns of Finnish investors with high IQ (in the top 5% of the distribution) outperform those with low IQ by 11 basis points, which suggests that the display format of market data has an effect for trading performance similar to that of individual cognitive abilities. As other direct comparisons, [Coval, Hirshleifer, and Shumway \[2005\]](#) find a strategy based on previous successful/unsuccessful retail investors that earns a daily abnormal return of 5 basis points and, in Taiwan, [Barber, Lee, et al. \[2012\]](#) show that the spread in gross returns between the top and bottom performing investors is 70 bps per day.

My explanation for this increase in trade performance is that investors take advantage of the improved information display format of Trader+ to monitor their limit orders more efficiently than before, which, in turn, leads to higher returns on those orders. This increase in performance has two implications, according to market microstructure models of order choice in which investors' trade performance is affected by the "cognitive cost" of monitoring their limit orders.<sup>8</sup> First, treated investors should use limit orders more often than before, because the expected

---

<sup>7</sup>See [Grinblatt, Keloharju, and Linnainmaa \[2011\]](#). For instance, [Hollifield et al. \[2006\]](#) define the risk of adverse selection as the expected loss (or gain) due to future expected changes in stock value given execution. Accordingly, [Liu \[2009\]](#) obtain a proxy for this expectation by comparing the current price after the execution of a limit order to the price at which the order has been executed.

<sup>8</sup>This cognitive constraint is embedded in many leading market microstructure papers. See for instance [Foucault, Roell, and Sandas \[2003\]](#), [Hollifield et al. \[2006\]](#), [Liu \[2009\]](#) and [Biais, Hombert, and Weill \[2013\]](#).

utility of trading with a limit order increases relative to the utility of trading with a market order. In line with this intuition, I find that following the introduction of the software the *fraction* of limit orders executed by treated investors increases by 8 percentage points, relative to the control group. Second, I expect treated investors to be able to spot and monitor short-term trading opportunities that were likely to be too difficult to monitor before. This change in behavior should lead to a decrease in individual trading horizon. To test this implication, I identify all the round-trips executed by investors in my sample and I proxy for trading horizon of a given investor, in a given month, by computing the average duration (in days) of his round-trips initiated that month. I find that i) the average short-term trading horizon decreases by almost 4 days (approximately a 10% decrease relative to the pre-treatment trading horizon of the treated group) and ii) the number of round-trips opened and closed within the same day increases by 2 percentage points (from a pre-treatment percentage of about 5% for the treated group). Consequently the overall trading activity of the treated group nearly doubles, compared to that of the control group, only one month after the introduction of the software. In summary, the changes in trading behavior for the treated group, after the software is introduced, are thus consistent with an explanation based on limit order monitoring costs.

I also consider, and rule out, alternative explanations for my findings based on trading speed, overconfidence and investors' self-selection. A first possibility is that trading returns increase because the use of the software is associated with a greater trading speed that allows investors to place their orders faster than other investors. This would be the case, for instance, if investors switch to Trader+ when they increase the speed of their internet connection or if orders sent through Trader+ are routed more quickly to the market by the brokerage house. If this explanation is correct, performance should increase for both market *and* limit orders. In particular, this improvement should be stronger for market orders, because these orders benefit the most from a higher trading speed (see [Garvey and Wu \[2010\]](#)). Instead, I find the opposite in my sample: returns on limit orders improve, whereas returns on market orders *do not improve* and the use of limit orders increases for investors who switch to Trader+. These facts suggest that

my results are not driven by a change in trading speed.

Another possibility is that the trading software may have increased investors' overconfidence. A mechanism similar to the one in [Barber and Odean \[2002\]](#) would suggest that investors with past successes prior to the introduction of Trader+ are more likely to become overconfident and to adopt Trader+, because overconfidence is enhanced by the "illusion of control" enjoyed when using trading software. Therefore, investors may mistakenly believe that the software allows them to better manage their limit orders. This alternative explanation has two implications. First, one expects returns on limit orders to actually decrease for treated investors after the switch to Trader+, because the software should enhance investors' overconfidence. I find the opposite result. Alternatively, returns should be at least lower for more overconfident traders, because more overconfidence leads to larger and more frequent behavioral mistakes. Instead, using individual performance prior to Trader+ as a proxy for investor overconfidence (as suggested in [Barber and Odean \[2002\]](#)), I find that investors who are more likely to be overconfident benefit the most from the software. These findings suggest that overconfidence does not explain my results.

Finally, while the software's introduction was an event decided by the brokerage house, its adoption (or non-adoption) was left to the investor. Therefore self-selection issues may be a possible concern to establish causality.<sup>9</sup> Difference-in-differences estimates are robust to a particular form of selection: those who switch to the treatment are allowed to be those who benefit the most from it (see [Todd \[2007\]](#)), as long as selection is based on time-invariant (observable or unobservable) variables. Self-selection is only a concern in the case of dynamic selection into treatment, for which I do not find evidence in my data.<sup>10</sup>

This paper directly contributes to the few recent works that study how alternative presentation formats of financial information influence investors' risk-appetite ([Kaufmann, Weber, and](#)

---

<sup>9</sup>This is a well known problem in the program evaluation literature. See [Heckman et al. \[1996\]](#) and [Heckman et al. \[1998\]](#).

<sup>10</sup>In that case the most critical difference-in-differences identifying assumption, the "parallel trend assumption", would not be verified. For an example, see [Ashenfelter and Card \[1985\]](#). The trends in outcome between treated and control investors in my sample fully support this (untestable) assumption.

Haisley [2012]), investors' behavioral biases (Rubaltelli et al. [2005]) and investors' retirement portfolio choices (Looney and Hardin [2009] and Bateman et al. [2011]). To the best of my knowledge, this is the first paper to show that the *display format* of market data affects the trading choices and the performance of individual investors. A direct implication of this result is that once households' portfolio allocation choices are made, regulators can improve the portfolio-rebalancing execution strategies of retail investors by giving households access to market data that is easier to process. That is, a more efficient display format of market data can help retail investors to make better financial decisions, and thus can be seen as a practical way to improve their financial literacy (Lusardi and Mitchell [2011]).<sup>11</sup>

My results also contribute to a recent literature pioneered by Linnainmaa [2010], Barber, Lee, et al. [2012] and Kelley and Tetlock [2013] that uses insight from market microstructure theory to understand investors' decisions. This paper suggests that the display format of the data being processed by investors may explain part of the large cross-section variation in retail investors trading skills identified in the literature (Grinblatt, Keloharju, and Linnainmaa [2011]; Barber, Lee, et al. [2012]) and may be an indirect determinant of households' financial decisions.

This article proceeds as follows. Section 1.2 discusses my testable empirical hypothesis. Section 3.2 presents the brokerage dataset and Section 1.4 motivates and discusses my identification strategy. The results are given in Section 1.5 and Section 1.6. Before concluding, I discuss potential alternative explanations for my results in Section 1.7.

---

<sup>11</sup>Indeed, the term financial literacy “can encompass concepts ranging from financial awareness and knowledge, including of financial products, institutions, and concepts; financial skills, such as the ability to calculate compound interest payments; and financial capability more generally, in terms of money management and financial planning” (Xu and Zia [2012]).

## 1.2 Testable Hypothesis

Almost all major exchanges today are limit order book markets.<sup>12</sup> In a limit order book market there is no designated market-maker setting prices for market participants: investors can instead either provide liquidity to the market by using limit orders or consume liquidity by using market orders.

Therefore, to rebalance its portfolio holdings, an investor must implement a trading strategy that involves the choice between market and limit orders. [Figure 1.1](#) specifies the decision tree faced by an investor willing to trade, and helps to understand how I obtain my empirical implications. This tree is a modified version of the one in [Fong and Liu \[2010\]](#).

<Insert [Figure 1.1](#) Here>

The investor first needs to collect and process market data (step 1) and optimally decide which order type to use for his trade (step 2): market order or limit order. Limit orders specify a number of shares and the maximum price at which an investor is willing to buy or the minimum price at which he is willing to sell. Limit orders accumulate in the limit order book in descending buy-price order or ascending sell-price order (price priority) and join the queue of other limit orders with the same price (time priority). Market orders specify a number of shares to buy or sell, but do not specify a particular price: they are *immediately* filled at the most attractive price posted by limit orders in the limit order book.

Hence, market orders provide immediacy and guaranteed execution at the cost of the bid-ask spread. A limit order instead offers a better execution price than a market order, but involves two different risks. First, the execution of a limit order is not guaranteed. This risk is known as

---

<sup>12</sup>Such as the NYSE Euronext or the NASDAQ. As of January 2012, the worldwide Electronic Order Book (EOB) turnover value was about USD 49,000 billions (see <http://www.world-exchanges.org>).

the risk of non-execution. Second, if a trader does not monitor the market, the limit order may be executed in adverse market conditions. This risk is known as the adverse selection risk.<sup>13</sup>

Both of these risks can be mitigated by active order management, which is usually called *monitoring activity* in the market microstructure literature. Limit orders can indeed be cancelled and/or revised after their submission. A higher limit price priority will increase the execution probability and thus decrease the risk of non-execution. A lower limit price priority will instead decrease the likelihood of execution and thus decrease the risk of adverse selection (Fong and Liu [2010]).

Therefore, an investor must monitor the market and his limit orders to increase his trading performance on these orders. As monitoring is a costly cognitive process, investors cannot continuously monitor their limit orders. This reasoning implies that a more efficient monitoring activity should lead, everything else being equal, to a higher trading performance on limit orders. This increase in performance, in turn, should induce investors to use limit orders more often than they previously did before because it is more profitable to do so.<sup>14</sup>

Furthermore, short-term strategies in which investors trade very frequently (such as “day-trading”, where investors revert their positions at the daily level) are more difficult to monitor than long-term ones. If this is correct, a more efficient monitoring activity should allow investors to spot more short-term trading opportunities than before. This effect implies that the proportion of short-term trading strategies should increase (leading to higher trading activity) when investors monitor their limit orders more efficiently.<sup>15</sup>

In summary, I have the following testable hypothesis at the investor level:

---

<sup>13</sup>Also known as the risk of being picked-off, or the free-option risk. See Copeland and Galai [1983], Foucault [1999] or Hollifield et al. [2006].

<sup>14</sup>This reasoning is illustrated in step 2 of Figure 1.1: an investor submits a limit order if the *expected* utility of using a limit order is greater than the *certain* utility of using a market order. The investor sends a market order otherwise.

<sup>15</sup>This mechanism is actually mentioned in Barber and Odean [2002]: “Investors may trade more when they go online simply because of greater ease of access. For rational investors this implies that there were potentially profitable trades that the investors declined to make before going online because the expected profits did not warrant the effort of calling a broker”.

1. A more efficient monitoring activity increases the trading performance of limit orders, because investors better mitigate adverse selection risks.
2. A more efficient monitoring activity has no effect on the performance of market orders, because market orders do not need to be monitored (as their execution is immediate and certain).
3. A more efficient monitoring activity increases the probability of submitting a limit order, as limit orders become more profitable than market orders.
4. A more efficient monitoring activity increases trading volume and reduces trading horizon, because investors exploit short-term strategies that were too costly to execute and monitor before.

## 1.3 Data

In this section I describe the data and define the variables that I will use in order to test implications 1,2,3 and 4 described in the previous section.

### 1.3.1 The Brokerage Dataset

The data used in this paper comes from a leading French online broker.<sup>16</sup> The raw dataset contains at the daily level all of the executed trades sent by each of the 145801 customers of the broker from 1999 up to 2010, which represents more than 15 million trades.

Each trade comes with the following information: the asset type (equity, bonds, etc), the trading exchange identifier (the ISIN), the trading date, the quantity, the order type (limit, market and other minor orders types), the amount traded in euro, the trading place and the trading channel used to submit the order. I match the trades in my dataset with market data

---

<sup>16</sup>A subsample of this database covering the period 1999-2001 has already been used, to address another research question, in [Foucault, Sraer, and Thesmar \[2011\]](#).

provided by Eurofidai, the European financial data institute. Trades are matched by ISIN code, trading day and trading exchange code. Trades for which no information is available from Eurofidai are discarded from the sample.

Summary statistics of the raw brokerage dataset are provided in [Table 1.1](#). This table shows the corresponding number of trades, percentage and cumulative percentage for several categorical variables that describe the nature of my data well. Panel A shows for instance that trades are in the majority of cases limit orders (62%) and market orders (28%). The other minor orders types are rarely used by investors. The information on order type is completely missing for 1999 and 2000 and may be sometimes missing up to 2004. Therefore the previous figures are computed without taking into account the number of orders with missing order type information.

Investors could submit an order by using the telephone to speak to a broker official (Telephone), by calling a voicemail service and typing instruction using the telephone's keys (Phone+), by using a web navigator to connect to the broker website (Web), by using an old French Videotex online service accessible through the telephone lines (Minitel) or by using online basic computer software (Online+). Trader+ is a trading software that was introduced in June 2003. It will be fully discussed and presented in [Section 1.4](#). Panel B of the table highlights that a large majority of trades are submitted using the internet or trading software (Online+ and Trader+).

Panel C and Panel D of the table show that most of individual trading activity consists of buying or selling common stocks on the NYSE Euronext Paris trading exchange. This dataset, therefore, shows patterns of individual investment behavior that are similar to other recent databases used in the literature. For instance, Finnish investors in [Linnainmaa \[2010\]](#) also use limit orders for most of their trades on the Helsinki Stock Exchange.

<Insert [Table 1.1](#) Here>

### 1.3.2 Variable Definition

I describe in this section how I measure investors' trading performance ([Section 1.3.2](#)) and investor trading horizon ([Section 1.3.2](#)).

#### Measuring trading performance

I measure trading performance as follows. For each trade in my sample, I compute the signed difference between the closing price of the stock bought (or sold) and the execution price of the order, divided by the execution price of the order:

$$R_{i,t,k,s} = \text{sign}_{i,t,k,s} * \frac{\text{Close}_{t,s} - \text{Price}_{i,t,k,s}}{\text{Price}_{i,t,k,s}}$$

$\text{Close}_{t,s}$  is the closing price on day  $t$ , of the traded stock  $s$ , and  $\text{Price}_{i,t,k,s}$  is the execution price of order number  $k$  submitted on day  $t$ , for stock  $s$ , by individual  $i$ .  $\text{Sign}_{i,t,k,s}$  is a dummy variable that equals one for a buy order and minus one for a sell order.

This ex-post performance measure is well suited for assessing the performance of a limit order, because it captures the adverse selection risk faced by investors using those orders (see [Harris and Hasbrouck \[1996\]](#)). Indeed, [Hollifield et al. \[2006\]](#) define the risk of adverse selection as the expected loss (or gain) due to future expected changes in stock value given execution. Similarly, [Liu \[2009\]](#) obtain a proxy for this expectation by comparing the current price after the execution of a limit order to the price at which the order has been executed.

A similar approach is also adopted in [Grinblatt, Keloharju, and Linnainmaa \[2011\]](#) and [Linnainmaa \[2010\]](#). [Linnainmaa \[2010\]](#) computes intraday returns on investors' orders by also comparing the execution price to the close price on the same day. He also infers from those returns that investors' limit orders may suffer from adverse selection risks:

*“Investors may have limit orders in the book around earnings announcements if they are unaware of an earnings announcement or if the cost of monitoring a limit order exceeds the expected loss due to an earnings announcement.”* (pp.1488)

In [Grinblatt, Keloharju, and Linnainmaa \[2011\]](#) the authors compare the performance between high and low IQ investors and argue that this return is essentially a measure of one order’s execution quality and should capture “the active management of individuals”. “High IQ investors”, they say, “may be better or quicker at processing information into a useful signal, or excel at distinguishing useful information from noise”. Because the researchers are able to detect statistically significant differences between the returns of the high IQ and low IQ groups, their hypothesis cannot be rejected.

My approach follows the same reasoning. If a more efficient display of market data allows investors to monitor their limit orders more efficiently, then investors should improve upon the execution quality of their limit orders relative to investors using a less efficient data display. This improvement is captured by my intraday returns.

To test this hypothesis, I need an improvement in the display of market data for retail investors. This point will be discussed in [Section 1.4](#).

## Measuring investors’ trading horizon

To proxy for investors’ trading horizon (and test the fourth empirical prediction), I adopt a methodology first described in [Schlarbaum, Lewellen, and Lease \[1978\]](#). I aggregate investor i’s single trades into round-trip trades and I use the investor i’s average round-trip duration in a given month, as a proxy for trading horizon that month.

In order to do so, I first compute the daily net quantity traded by investor  $i$ , on stock  $s$  and day  $t$  as:

$$Q_{i,t,s} = Q_{i,t,s}^{buy} - Q_{i,t,s}^{sell}$$

where  $Q_{i,t,s}^{buy}$  (resp.  $Q_{i,t,s}^{sell}$ ) represents the actual quantity of stock  $s$  bought (resp. sold) by investor  $i$  on day  $t$ .

Then, I sort all the daily net quantities in my dataset by trader, stock and trading day and I keep track of the cumulative stock quantity held by a trader day after day. A round-trip starts and ends with a zero net cumulative quantity. Round-trips are thus trading positions that are fully unwound: stocks previously bought are entirely sold and stocks previously sold short are completely bought back.

The duration of a round-trip is thus the number of days between the start and the end of a round-trip. I obtain a proxy for investor  $i$ 's trading horizon, in a given month, by taking the average duration of all the investor  $i$ 's round-trips *started* that month.

## 1.4 Methodology and Identification Strategy

### 1.4.1 Trader+

I explain in this section why the display of market data in Trader+ should help investors more efficiently monitor their limit orders.

Monitoring activity is a costly cognitive process. Thus investors cannot achieve continuous monitoring of their limit orders. However, the idea that information technologies can help investors is pervasive in the literature. For instance, [Biais, Hombert, and Weill \[2013\]](#) derive equilibrium prices in a limit order book when new trading technologies such as algorithmic trading are used to alleviate the individual's limited cognitive capacities. Also, [Foucault, Roell, and Sandas \[2003\]](#) makes the distinction between news monitoring, such as monitoring public announcements, and quote monitoring. Quote monitoring is limited to monitoring other dealers' quote updates and, they argue, does not depend on the monitoring level because "*in practice, bandits and dealers [in a dealers' market] use software that alerts them to quote updates in different*

securities”. While they address other questions, the authors mention that the probability of observing quote updates depends on the fixed cost of the trading technology used.

Furthermore, [Fong and Liu \[2010\]](#) state that “*the option to revise or cancel limit orders is particularly important following recent advances in information technology [...] that allow traders to manage orders directly from their computers*”. This point is consistent with [Peng and Xiong \[2006\]](#), who study the effects of limited attention on asset prices and show that “as information technology advances over time, investors’ attention constraints become less binding”.

Building on these ideas, I argue that the new information display of market data, available on Trader+, allowed Trader+ users to monitor their limit orders more efficiently than before. Indeed, prior to June 2003, the broker customers submitted their orders mainly through two trading channels: using the existing trading software available (Online+) and connecting to the brokerage web interface (Web). [Figure 1.2](#) provides a screenshot of each of these trading channels, and highlights how identical information items (such as the current state of the limit order book, or the recent market movements statistics) are displayed differently on those trading channels. In both cases, one can see that the market data are dispersed through different webpages (for Web) or through different tabs (for Online+).

<Insert [Figure 1.2](#) Here>

Instead, Trader+ displayed market data much more efficiently than those trading channels because it *simultaneously* gathered all relevant information items into a user-customized screen. A screenshot of Trader+ is shown in [Figure 1.3](#). As one can see, investors using Trader+ could see on their computer screen, not only the limit order book or the most recent market statistics, but also their pending orders, the stock intraday graphics and market data at the stock-level.<sup>17</sup> The key point is that these information items were also available, in a more dispersed form, on the other trading channels. In other words, the quantity of information remains the same, while the display of that information varies when Trader+ is introduced.

---

<sup>17</sup>Additionally, investors can fully customize their Trader+ graphical user interface according to their needs and preferences, choosing the relative ordering and appearance of each information items (display color, size, etc).

<Insert [Figure 1.3](#) Here>

The simultaneous presentation of information items is the critical feature that allows investors to better understand how stock prices and market movements are related to one another, and helps investors better assimilate that information when they monitor their limit orders. I therefore use the introduction of Trader+ as a positive shock to the investors' limited monitoring capacities. This assertion is consistent with [Hodge, Kennedy, and Maines \[2004\]](#), who consider a search-facilitating technology (the eXtensible Business Reporting Language (XBRL) technology) introduced by the SEC in 2003 and state that:

*“A key attribute of XBRL-enhanced search engines is that they facilitate users simultaneously viewing similarly tagged financial information. This simultaneous presentation helps users to evaluate items in relation to each other and to integrate the related information when making decisions (Russo 1977). This feature of search-facilitating technology facilitates users integrating dispersed, but related, information in ways they may not have considered in a more traditional, non-searchable, environment ”*

In summary, the display of information of Trader+ should allow investors to monitor more efficiently their limit orders, which gives me the opportunity to test the empirical implications derived in [Section 1.2](#).

### **1.4.2 Identification Strategy**

To test my empirical predictions, I use in this paper a difference-in-differences methodology (hereafter "DiD"). DiD estimates recover average treatment effects on the treated group by essentially comparing the outcomes between treated and control investors both before and after exposure to the treatment (Trader+).

The key identifying assumption in this setup (see [Angrist and Pischke \[2008\]](#) or [Wooldridge \[2010\]](#)) is that trends in outcomes for treated and control investors would have been the same

in the absence of treatment. This fundamental, but untestable, identifying assumption can nonetheless be evaluated by comparing graphically the trends in outcomes for both groups in multiple periods before and after the treatment exposure. I use the 10-year range of my dataset to support this “parallel trend” assumption.

It is well known that difference-in-differences estimates can be obtained using an OLS regression (see [Angrist and Pischke \[2008\]](#)), which has the advantage of allowing for a correct specification of standard errors (see [Bertrand, Duflo, and Mullainathan \[2004\]](#)). I therefore run the following difference-in-differences regression for outcomes at the investor-day-trade level:

$$Y_{i,t,k} = \alpha + \beta * Monitoring_{i,t} + \gamma * Month_t + \delta_i + \epsilon_{i,t,k}$$

Similarly, the difference-in-differences regression at the investor-month level is:

$$Y_{i,m} = \alpha + \beta * Monitoring_{i,m} + \gamma * Month_m + \delta_i + \epsilon_{i,m}$$

$Monitoring_{i,t}$  (resp.  $Monitoring_{i,m}$ ) is the variable of interest. It is a dummy variable that equals one, for investor  $i$ , for all days  $t$  (resp. months  $m$ ) that follows his first trade executed through Trader+. This specification thus allows for a staggered entry into treatment at the investor level, and is more precise than a standard difference-in-differences framework that imposes the same treatment period for every treated investor. As a robustness check, I verify (and obtain) in [Section 1.6](#) that my results still hold in a standard difference-in-differences regression.

The regression specification includes a full range of individual fixed effects  $\delta_i$  and monthly fixed effects for every month from 2002 to 2005. Standard errors are robust and clustered at the individual level (see [White \[1980\]](#) and [Rogers \[1994\]](#)).

The fixed effects  $\delta_i$  captures the differences between investors that are fixed over time, whereas the monthly dummy variables capture the time factors that are common to both treated

and control investors. The difference-in-differences estimate is the coefficient  $\beta$ . It shows the effect of a more effective monitoring activity (induced by the aggregated display of information in Trader+) on the outcome  $Y$ .

The difference-in-differences framework is a powerful tool for causal inference but my estimates may be biased if investors are allowed to enter or exit the market after (or before) the treatment period. To control for attrition, I require that both treated and control investors have an open common stock position at least between 2002 and 2005, at least a single trade between March and the end of May 2003 and at least one trade after June 2003. Treated investors have submitted at least one trade with Trader+ after June 2003, while control investors have not.

### 1.4.3 The Control Group

Investors who switch to Trader+ may be different, in many ways, from other investors who never use the software. To obtain a control group that provides a credible counterfactual, I use a nearest-neighbor propensity score algorithm in order to find the best control investor for each treated investor in my sample. The propensity-score algorithm matches treated and control investors who share the same probability (estimated *using only pre-treatment data*) of switching to Trader+ after June 2003.

This approach relies on finding the right set of covariates that determines the switch to treatment. In this respect, the seminal paper on online investors by [Barber and Odean \[2002\]](#) is very useful. The authors find that young active male traders with high incomes are more likely to switch to online trading than other traders. Those who switch also report more trading experience and a particular preference for speculative trades. Prior to going online, moreover, investors experienced unusually strong performance. [Choi, Laibson, and Metrick \[2002\]](#) study in a different setting the impact of a web-based trading channel on two large corporate 401(k) plans. They obtain that traders that are used to phone trading are less likely to try the web.

While these results may not be completely generalizable to this study, as the majority of investors in my sample *are already using the internet* for the majority of their trades, these papers emphasize that trading behavior before the treatment period can help identify the variables key to the selection process.

From a more technical point of view, [Stuart \[2010\]](#) also remarks that:

*“it is important to include in the matching procedure all variables known to be related to both treatment assignment and the outcome [...]. When matching using propensity scores [...] there is little cost to including variables that are actually unassociated with treatment assignment, as they will be of little influence in the propensity score model. Including variables that are actually unassociated with the outcome can yield slight increases in variance. However, excluding a potentially important confounder can be very costly in terms of increased bias.”*

I therefore compute various covariates according to investor demographics, general trading behavior, account size and trading channel preferences and I estimate, with pre-treatment data only, a cross-section logit regression where the dependent variable is one for a treated investor and zero otherwise. All these covariates are obtained using individual trading data from March, April and May 2003.

These covariates are grouped in four classes: Demographics, Trading Behavior, Account Size and Trading Channels. I use two demographic variables: Female is a dummy variable that is one for females and zero for males, and Age corresponds to the investor’s age in 2003. The variables that capture investor trading behavior, over the three months pre-treatment period, are the individual number of orders executed ("Nb of executed trades"), the individual number of different asset classes traded ("Nb. of asset types"), the individual mean euro amount traded ("Mean amount traded"), the individual mean daily-return on executed orders ("Mean daily return"), the individual percentage of orders that are limit orders ("% Limit orders") and the individual percentage of orders submitted with a margin account ("% Margin"). Account size

variables provide information about the total euro value of an investor's common stock holdings on both its spot-market account and his margin account (in ten of thousands euros) in May 2003. Finally, trading channels variables provide, for each trading channel, the corresponding individual percentages of trades submitted by an investor through that channel.

Results of the logit regression are given in [Table 1.2](#). What determines the switch to the softwares appears clearly: active traders with high past performance have a larger probability of switching to treatment. Furthermore, investors who use online trading channels, such as the web or the old computer software that has been available since 1999 (variable "Online+", highly significant), are also more likely to use Trader+.

<Insert [Table 1.2](#) Here>

The matching algorithm, based on the above logit estimation, performs well as [Table 1.3](#) shows. [Table 1.3](#) compares the mean covariate values between treated and matched control investors in my sample. It appears that all the variables that determine the investor propensity to use Trader+ (according to [Table 1.2](#)) are very similar between treated and matched control investors. Indeed, treated and matched controls have a similar pre-treatment performance, submit on average the same number of trades between March and May 2003 (25 vs. 22), the average order amount is almost equal (3400€ vs. 3200€) and they manage their orders using the existing online trading tools (online+ : 55% vs 56%, web : 34% vs 33%).

<Insert [Table 1.3](#) Here>

Last, [Figure 1.4](#) shows, for both the treatment and the control group, the total number of orders (each year from 1999 to 2010) broken down by order type and trading channel. These figures summarizes well the outcome of the matching procedure. For instance, one can see that from the Figure that: i) the number of orders submitted each year until 2003 is very similar between treated and control investors and ii) both groups submit limit orders and use online tools most of the time.

<Insert [Figure 1.4](#) Here>

## 1.5 Results

### 1.5.1 Trading performance

I test in this section my first two empirical predictions. Does a more effective display of market data (due to Trader+) allow investors to better monitor their limit orders and increase, in turn, their returns on those orders?

Panel A in [Table 1.4](#) provides a first answer. For each trading channel, this table provides a statistical description of the distribution of investors' intraday returns. I use all the orders submitted by both treated and control investors after June 2003. What emerges from the table is that returns obtained by traders using Trader+ are greater, in each quantile, than returns achieved through other trading channels. Moreover the table shows a progressive shift in returns from negative values towards positive ones as investors move from hard-to-read displays of market data (e.g.: minitel) to more efficient ones (online trading channels). These facts are thus consistent with the idea that online tools display market data in a way that is "more processable" for investors ([Russo \[1977\]](#)).

<Insert [Table 1.4](#) Here>

The formal test of the first and second empirical implication is shown in [Table 1.5](#). I use three different samples of orders. I report in the column labeled "All Orders", the difference-in-differences coefficient of interest (see [Section 1.4](#)), when I include in my sample all the orders of both treated and control investors. In the columns "Limit" and "Market", the sample is restricted to limit orders only or market orders only. In all cases, the sample period is between 2002 and 2005, and standard errors are clustered at the individual level.

The difference-in-difference coefficient, which is the coefficient on the variable is *Monitoring*, captures the effect, on investors' trading performance, induced by a more efficient monitoring activity. As predicted, I find a strong increase in the trading performance of executed limit orders (10 basis points) and a positive, but non-significant, effect on market orders.

<Insert [Table 1.5](#) Here>

These results confirm the first two empirical implications and suggest two remarks. First, this is evidence that the returns I compute are indeed a good proxy for the risk of adverse selection, because the "falsification test" provided by the sample of market orders only is fulfilled: no effect is found on those orders. Second, market orders are not exposed to adverse selection and non-execution risk but they certainly benefit from a good market timing, which is also likely affected by *Trader+* (see Step 1 in [Figure 1.1](#)). This explains why the coefficient is slightly positive. In essence, market and limit orders thus allow me to disentangle between market timing and pure monitoring activity.

Additionally, this table also rules out possible concerns coming from *marketable limit orders* in my sample. Marketable limit orders are limit orders whose price is above the best ask (for buy orders) or below the best bid (for sell orders). Because of their price, they are therefore immediately executed and are thus nearly identical to market orders. I cannot separate in my database between pure limit orders and marketable limit orders because I do not have the timestamp of the order. Therefore, I cannot compare the original limit order price to the bid-ask spread at the time of submission. However, the possibility that the majority of orders in my database are actually marketable limit orders is rejected by the data. In such a case, indeed, there should be not difference between estimates on market orders and limit orders.

Monthly trends in returns for both treated and control investors, before and after June 2003, are shown in [Figure 1.5](#) and [Figure 1.6](#). [Figure 1.5](#) refers to the sample containing all the orders of both the treatment and control group, while [Figure 1.6](#) shows the trend in returns separately for the restricted samples of market orders only and limit orders only. In both cases

these Figures support the parallel trend assumption. Recall that the DiD estimates could be biased if, for instance, treated investors had experienced a transitory shock in returns (positive or negative) before treatment. Indeed, my regression in this case would erroneously attribute to the treatment an effect due to mean-reversion. Instead, [Figure 1.6](#) suggests that investors' returns between the treatment and the control group are visually (and statistically) indistinguishable prior to the introduction of Trader+, which goes against a dynamic selection threat.

<Insert [Figure 1.5](#) Here>

<Insert [Figure 1.6](#) Here>

## 1.5.2 Order Choice

I test in this section the third hypothesis: do investors use limit orders more often when they are able to monitor their limit orders more efficiently?

As in the previous section, Panel B in [Table 1.4](#) provides a first empirical answer. Panel B is a frequency table that gives the corresponding number of trades and row and column percentages for each possible combination of trading channel and order type. I use all the orders submitted by both treated and control investors after June 2003.

The table shows that there is a strong association between the use of limit orders and the use of trade order management software. Investors submit and execute 36% of all their limit orders using Trader+, the rest being submitted with Online+ (34%) or using the Web (28%). It is interesting to see how the choice between limit and market orders is completely reversed when one compares orders submitted by Telephone with orders submitted by Trader+ : 59% of the orders submitted by Telephone are market orders while this figure drops to 32% for Trader+. This empirical pattern illustrates well the idea that telephone users may incur very high monitoring costs that render limit orders too difficult to monitor, and therefore less profitable, on average, than market orders.

Turning to a more formal test, I report in column "LO Dummy" of [Table 1.5](#), the estimates of a DiD regression where the dependent variable is a dummy variable that equals one for a limit order and zero otherwise. The effect of a more efficient monitoring activity is identified, again, by the difference-in-differences coefficient on the variable *Monitoring*. The coefficient shows that following their switch to Trader+, the probability of submitting a limit order increases for treated investors by about 8 percentage points, relative to control investors.

In terms of economic significance, I find that the effect is strong enough to be visible at the week level after June 2003. [Figure 1.7](#) shows the proportion of limit orders executed by treated and control investors as the proportion of trades submitted by treated investors with Trader+ increases. As expected, limit orders become more and more important in the order flow of the treated group after June 2003, which confirms visually the previous results.

<Insert [Figure 1.7](#) Here>

### 1.5.3 Trading Volume

Last, I test the fourth implication in this section. How does the efficiency of monitoring activity affects investors' trading activity? As stated before, I expect an increase in trading activity for investors who switch to Trader+, as they should be able to implement short-term trading strategies that are unprofitable without efficient monitoring of limit orders.

Results are presented in Panel B of [Table 1.5](#). Variables in this panel are count variables at the investor-month level. I count the number of executed orders submitted by a given individual in a given month (column "Nb. Trades") and I separately count the number of limit or market orders executed that month. To capture another dimension of trading activity, I also compute the number of unique stocks traded by a given investor in a given month (column "Nb. Unique"). This variable captures whether investors have a larger "consideration set" (see [Barber and Odean \[2007\]](#)), when they better monitor the market and their orders. As discussed before, I

compute the average duration of investor i's round-trips *started* in a given month, as a proxy for his trading horizon that month (column "Horizon"). To further assess whether investors are more likely to engage in daily-trading, I also generate a dummy variable that equals one if a round-trip is fully completed within a day (column "Round-trip") and zero otherwise.

On average, both treated and control investors execute 9.7 orders per month (the baseline constant of my regression). The highly significant coefficient of 4.5 on the variable *Monitoring* suggests that trading activity at the individual-month level for treated investors increases by nearly 50%, relative to the control investors, when monitoring costs are lower. As with intraday returns, and consistent with my previous results on order choice, the increase in trading activity is due to investors using limit orders more often. Indeed, the table shows that both increases in the number of limit and market orders are highly significant, but the coefficient on the number of limit orders (4.5 additional orders per month, column "Nb. Limit") is about 7.5 times the coefficient on the number of market orders.

The coefficient of 0.4 (a 10% increase relative to the baseline constant) on the number of different stocks traded by investors each month is also of interest. It suggests that investors facing lower monitoring costs are able to manage more different trading positions *at the same time* than before. Investors trade more of the same stocks, but also more of different stocks.

Last, I turn to the effects of limited attention on trading horizon. The baseline constant in the column labeled "Horizon" indicates that the average short-term round-trip duration in both groups is about one month (32 days).<sup>18</sup> This result confirms that my sample of investors is populated by active traders. The coefficient on *Monitoring* is -3.6. This means that after the switch to Trader+, treated investors reduce their trading horizon by almost a trading week (a 10% decrease). Trading horizon decreases because investors use shorter trading strategies: the proportion of daily round-trips increases by 2% points after the switch, and relative to the control investor group.

---

<sup>18</sup>As the distribution of trading horizons is extremely skewed to the left, this regression is estimated on round-trips whose horizon is less than 100 days.

Trading activity increases, but it is important to know whether investors actually trade higher or smaller amounts than they previously did. It could be the case that treated investors split their orders so that trading frequency increases but the overall amount traded stays the same. I find that treated investors trade larger amounts than before (see Panel A, column "Amount"). This increase is explained by the use of leverage (see Panel A, column "Margin"): in order to benefit from short-term opportunities, treated investors increase leverage so that they can trade larger amounts.

Figure 1.8 provides a direct illustration of the findings above. I plot the number of trades executed, the median amount traded, the percentage of levered orders executed and the number of unique stocks traded each month by treated and control investors. Subfigure a (labeled "Nb. of Trades") shows how sudden the rise in trading activity is right after the introduction of Trader+. On aggregate, the monthly volume of orders executed by treated investors almost doubles relative to that of the control group. The patterns are very similar across the other measures of trading volume. This effect is interesting. Indeed, as a comparison, it takes 18 months for investors in Choi, Laibson, and Metrick [2002] to increase trading by 50%, relative to control investors, after their switch to a web-based trading channel.

<Insert Figure 1.8 Here>

Figure 1.9 illustrates more precisely the changes in trading horizon. I recover the empirical distribution of the duration of investors' round-trips before and after June 2003 in both groups. The densities are shown in the top subfigure in Figure 1.9. Before treatment, the distribution of round-trips lengths between treated and control investor is very similar. However, after treatment, the number of *short-term* trading strategies increases dramatically for treated investors relative to the control investors. The bottom subfigure focuses on round-trips whose duration is less than 100 days. It shows the median monthly duration for both groups. Again, and similarly to the patterns identified above, one can see that the median horizon of the treated group falls, relative to the control group, just after June 2003.

<Insert Figure 1.9 Here>

In summary, the results discussed in this section suggest that allowing investors to monitor their limit orders more efficiently, thanks to a better display of market data, significantly affects trading volume in terms of i) the frequency of trading (investors submit more orders and especially more limit orders) and ii) the investor horizon (investors execute more short-term strategies than they previously did).

## 1.6 Robustness Analyses

### 1.6.1 Standard difference-in-differences estimation

As fixed-effects regressions are sometimes “black-boxes” (Duflo [2002]), I verify in this section that my results are robust to a standard difference-in-differences analysis that does not take into account the staggered nature of the switch to Trader+ (the variable *Monitoring* defined in the previous sections). A standard DiD regression in my setting would mean that treated investors become treated after June 2003, whenever their first trade with Trader+ occurred. I therefore run the following difference-in-differences regression:

$$Y_{i,t,k} = \alpha + \beta * Treated_i + \gamma * Post_t + \lambda * Treated_i * Post_t + \epsilon_{i,t,k} \quad (1.1)$$

$$Y_{i,m} = \alpha + \beta * Treated_i + \gamma * Post_m + \lambda * Treated_i * Post_m + \epsilon_{i,m} \quad (1.2)$$

This standard specification is similar to the one used in this paper.  $Y_{i,t,k}$  (resp.  $Y_{i,m}$ ) is the outcome of interest computed from trade  $k$  submitted by individual  $i$  on day  $t$  (resp. the outcome of interest computed at the investor-month level). *Treated* is a dummy variable that equals one for a treated investor and *Post* is a dummy variable that equals one if a trade is executed after

June 2003.  $\epsilon_{i,t,k}$  is an error term clustered at the individual level. The variable *Treated* captures the differences between treated and control that are fixed over time, whereas the variable *Post* captures the time factors that affect the trends in outcomes after June 2003 common to both treated and control investors. The difference-in-difference estimate is captured by the coefficient on the interaction variable *Treated \* Post*.

Results are shown in [Table 1.6](#). All the previous results are confirmed: returns improve on limit orders only, investors use limit orders more and trading horizon decreases.

<Insert [Table 1.6](#) Here>

## 1.6.2 Placebo Treatment

As a additional robustness analysis, I can take advantage of the simplicity of the previous specification to perform a placebo test. To do so, I estimate a standard difference-in-differences regression using only data from April 2002 to April 2003. I then split the sample in two and consider that a (placebo) treatment has occurred after October 2002. To do so I generate a variable *FakePost*, that is equal to one between October 2002 and April 2003.<sup>19</sup>

Of course, no particular shock has hit the treated group during this period. Thus I expect the difference-in-difference estimates to be statistically non-significant.

Results are provided in [Table 1.7](#). As expected, this placebo treatment has no effect. For each outcome, the coefficient on the variable *Treated \* FakePost* is always non-statistically different from zero.

<Insert [Table 1.7](#) Here>

---

<sup>19</sup>Results are similar if I choose another placebo date.

## 1.7 Alternative Explanations

I find in this paper that retail traders who switch to a new order management software subsequently earn higher returns on their limit orders, use limit orders more, trade more and have a shorter trading horizon. I propose an explanation in which investors have bounded rationality. The introduction of Trader+ allows them to better monitor the market and their limit orders thanks to a more efficient display of market data. The objective of this section is to discuss alternative explanations that may also explain my results.

### 1.7.1 Dynamic Self-Selection

A possible concern in my identification strategy is self-selection. As stated before, difference-in-differences are robust to self-selection issues if selection into treatment is determined by time-invariable variables (observable or unobservable). However, differences-in-differences are biased in case of dynamic self-selection because the “parallel trend assumption” would not be verified in that case. In my setting this would be the case, for instance, if investors suddenly understood the importance of monitoring their limit orders, as a reaction to the brokers’ ad in favor of the software, and subsequently began to monitor them more efficiently by simply exerting more attentional effort. Investors eventually switched to Trader+ simply because they had the opportunity to do so, but the effect of Trader+ is nil. In other words, treated investors in my sample would have obtained higher returns even without using Trader+.

Dynamic self-selection seems not to be a serious issue in my setting for two reasons. The first reason is provided by the dynamics of individual entry into treatment. [Figure 1.10](#) provides, each month starting from June 2003, the number of investors who switch into treatment status that month (i.e. investors submit their first Trader+ trade that month). From the graph, one can see that almost all treated investors switch immediately to Trader+ in June 2003. This pattern, and the jump in returns (shown in [Figure 1.5](#)) that immediately follows the introduction of

Trader+, are thus only consistent with self-selection dynamics that *contemporaneously affect* the treatment group *exactly* in June 2003. While I cannot definitely reject this possibility, this phenomenon seems hard to justify.

<Insert [Figure 1.10](#) Here>

Second, if the use of Trader+ has no effect, then there should be no specific pattern between investors' increases in returns and investors' switches to Trader+. Instead, the data shows that improvements in trading performance *systematically follow* the switch to Trader+. This is causality in the sense of Granger: consequences follow causes, and not the opposite.

These patterns can be seen in [Figure 1.11](#). I consider here a larger sample of investors, which also includes those that opened a trading account after 2003. I then classify then all the investors who eventually use Trader+ in my dataset into various cohorts, according to the year in which they submitted their first trade using Trader+. I obtain 8 cohorts from 2003 to 2010. In [Figure 1.11](#), I compare the trends in outcomes between the 2003 cohort, the 2005 cohort and the matched control investors who never use Trader+ (defined in [Section 1.4](#)).<sup>20</sup>

Subfigure a in [Figure 1.11](#) shows two interesting patterns: i) investors who will eventually switch in 2005 to Trader+ earn, before 2005, intraday returns that are very similar to those of the control group and ii) the jump in returns of the 2005 cohort only occurs in 2005, not before. A similar reasoning applies to all the other subfigures: for instance, the trading activity of the 2005 cohort increase only after January 2005.

<Insert [Figure 1.11](#) Here>

In summary, I find no evidence that dynamic self-selection is a serious concern in my paper.

---

<sup>20</sup>Results are similar with other cohorts.

## 1.7.2 Trading Speed

Another possibility is that investors achieve higher trading returns because the decision to use the software is associated with a positive variation in trading speed, that allows investors to place their order faster than other investors.

This possibility is related to several recent papers that focus on automation and speed in financial markets (see [Garvey and Wu \[2010\]](#) and [Hendershott and Moulton \[2011\]](#)).<sup>21</sup> Trading speed could play a role in my setting if investors switch to Trader+ when the quality of their Internet connection increases or if orders sent through Trader+ are routed more quickly to the market.

If this alternative explanation is correct, then the performance should increase for both market and limit orders. More specifically, the improvement in performance should even be greater for market orders than for limit orders, because market orders benefit the most from trading speed increases. Indeed, investors with a speed-advantage should use market orders more often than limit orders, in order to gain from speed-related trading strategies, such as picking off stale limit orders not being monitored by other slow investors ([Garvey and Wu \[2010\]](#)).

[Table 1.5](#) shows instead that for the treated group, after the switch to Trader+, i) the lion's share of improvements in returns comes from limit orders, ii) returns on market orders *do not improve* significantly and iii) investors use limit orders more than before.

These changes in behavior thus suggest that my results are not driven by variations in trading speed.

---

<sup>21</sup>It should be recalled that Trader+ *is not* an algorithmic trading tool: it takes no trading decisions alone. Behind each single order in my database there is the voluntarily mouse-click of an investor, so the improvements in order execution shown in the data are the outcomes of the investors' trade decisions.

### 1.7.3 Overconfidence

One last possible alternative explanation for my results is that Trader+ may have increased investors' overconfidence. Indeed, overconfidence has become the leading behavioral explanation for the sharp increase in trading, and the lower subsequent performance, of retail investors that switched to online trading at the beginning of 2000.<sup>22</sup> A mechanism similar to the one in Barber and Odean [2002] would suggest that investors with past successes prior to the introduction of Trader+ are more likely to become overconfident and also more likely to switch to the software because overconfidence is enhanced by the “feeling of empowerment” (or the “illusion of control”) enjoyed when using a trading software.

Therefore, treated investors in this paper could become overconfident about their ability to monitor their limit orders. This reasoning has two implications. First, overconfidence arguments predict a decrease in trading performance. Table 1.5 shows the opposite result. Alternatively, more overconfident investors switching to the software should achieve even lower returns than less overconfident investors, because investors more prone to overconfidence are more likely to commit behavioral trading mistakes.

To test this hypothesis, I compute the Sharpe ratio of past trading returns, prior to June 2003, for each investor in my sample. I then classify each investor into three groups defined by the terciles of the distribution of Sharpe ratios and I run a difference-in-differences regression for each group. This specification allows me to compare the changes in trading performance of (more or less) overconfident treated investors, after their switch to Trader+, compared to control investors with the same level of overconfidence.

Results are shown in Table 1.8. I find that investors with higher past returns (in the top tercile group 3) are those who benefit *the most* from the software (18 basis points per limit order trade), which contradicts the previous prediction based on overconfidence arguments.

---

<sup>22</sup>See Choi, Laibson, and Metrick [2002] and Barber and Odean [2002]

Interestingly, one can also see from the table that improving the display of information seems to be only beneficial for retail investors (at least in terms of trading returns).

Therefore, overconfidence does not seem to play a role in this paper.

<Insert [Table 1.8](#) Here>

## 1.8 Conclusion

I test whether the display of market data affects the trading performance of retail investors. I use the introduction of a trade order management software, that gathered all the relevant (and otherwise dispersed) market information into a single screen, as a shock that allows investors to better monitor their limit orders.

I find a strong increase in limit order trading performance for investors who use the software, suggesting that for retail investors the display format of financial information is critical. This paper therefore implies that regulators and policymakers should be concerned with *how* retail investors collect and process market data when they trade to rebalance their portfolio holdings.

## 1.8 References

- Angrist, J. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Ashenfelter, O. and Card, D. (1985). "Using the longitudinal structure of earnings to estimate the effect of training programs". *The Review of Economics and Statistics*.
- Barber, B. and Odean, T. (Dec. 2007). "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors". *Review of Financial Studies* 21.2.
- Barber, B., Lee, Y., Liu, Y., and Odean, T. (2012). "The Cross-Section of Speculator Skill Evidence from Day-Trading". *Working Paper*.
- Barber, B. and Odean, T. (2002). "Online Investors: Do the Slow Die First?" *Review of Financial Studies* 15.2.
- Bateman, H., Eckert, C., Geweke, J., Louviere, J., Satchell, S., and Thorp, S. (2011). *Financial Competence, Risk Presentation and Retirement Portfolio Preferences*. Working Papers 201120. ARC Centre of Excellence in Population Ageing Research (CEPAR), Australian School of Business, University of New South Wales.
- Benartzi, S. and Thaler, R. H. (Mar. 1999). "Risk Aversion or Myopia? Choices in Repeated Gambles and Retirement Investments". *Management Science* 45.3.
- Bertrand, M., Duflo, E., and Mullainathan, S. (2004). "How much should we trust differences-in-differences estimates?" *The Quarterly Journal of Economics* February.
- Bhattacharya, U., Hackethal, a., Kaesler, S., Loos, B., and Meyer, S. (Jan. 2012). "Is Unbiased Financial Advice to Retail Investors Sufficient? Answers from a Large Field Study". *Review of Financial Studies* 25.4.
- Biais, B., Hombert, J., and Weill, P.-O. (July 2013). *Equilibrium Pricing and Trading Volume under Preference Uncertainty*. IDEI Working Papers 787. Institut d'Économie Industrielle (IDEI), Toulouse.

- Calvet, L. E., Campbell, J. Y., and Sodini, P. (Oct. 2007). “Down or Out: Assessing the Welfare Costs of Household Investment Mistakes”. *Journal of Political Economy* 115.5.
- Calvet, L. E. and Sodini, P. (Forthcoming). “Twin Picks : Disentangling the Determinants of Risk-Taking in Household Portfolios”. *The Journal of Finance*.
- Campbell, J. (2006). “Household Finance”. *Journal of Finance* 61.4.
- Choi, J., Laibson, D., and Metrick, A. (2002). “How does the Internet affect trading? Evidence from investor behavior in 401(k) plans”. *Journal of Financial Economics* 64.3.
- Copeland, T. E. and Galai, D. (1983). “Information Effects on the Bid-Ask Spread”. *Journal of Finance* 38.5.
- Coval, J., Hirshleifer, D., and Shumway, T. (2005). “Can Individual Investors Beat the Market?” *HBS Finance Working Paper No. 04-025*.
- Duflo, E. (2002). “Empirical methods”. *MIT 14.771 / Harvard 2390b course*.
- Fong, K. Y. and Liu, W.-M. (2010). “Limit order revisions”. *Journal of Banking & Finance* 34.8.
- Foucault, T. (1999). “Order flow composition and trading costs in a dynamic limit order market”. *Journal of Financial Markets* 2.2.
- Foucault, T., Roell, A., and Sandas, P. (2003). “Market Making with Costly Monitoring: An Analysis of the SOES Controversy”. *Review of Financial Studies* 16.2.
- Foucault, T., Sraer, D., and Thesmar, D. J. (2011). “Individual Investors and Volatility”. *Journal of Finance* LXVI.4.
- Garvey, R. and Wu (Nov. 2010). “Speed, distance, and electronic trading: New evidence on why location matters”. *Journal of Financial Markets* 13.4.
- Glosten, L. (1994). “Is the electronic open limit order book inevitable?” *The Journal of Finance* XLIX.4.
- Grinblatt, M., Keloharju, M., and Linnainmaa, J. (2011). “IQ, trading behavior, and performance”. *Journal of Financial Economics*.
- Harris, L. and Hasbrouck, J. (1996). “Market vs. Limit Orders: The SuperDOT Evidence on Order Submission Strategy”. *The Journal of Financial and Quantitative Analysis* 31.2.

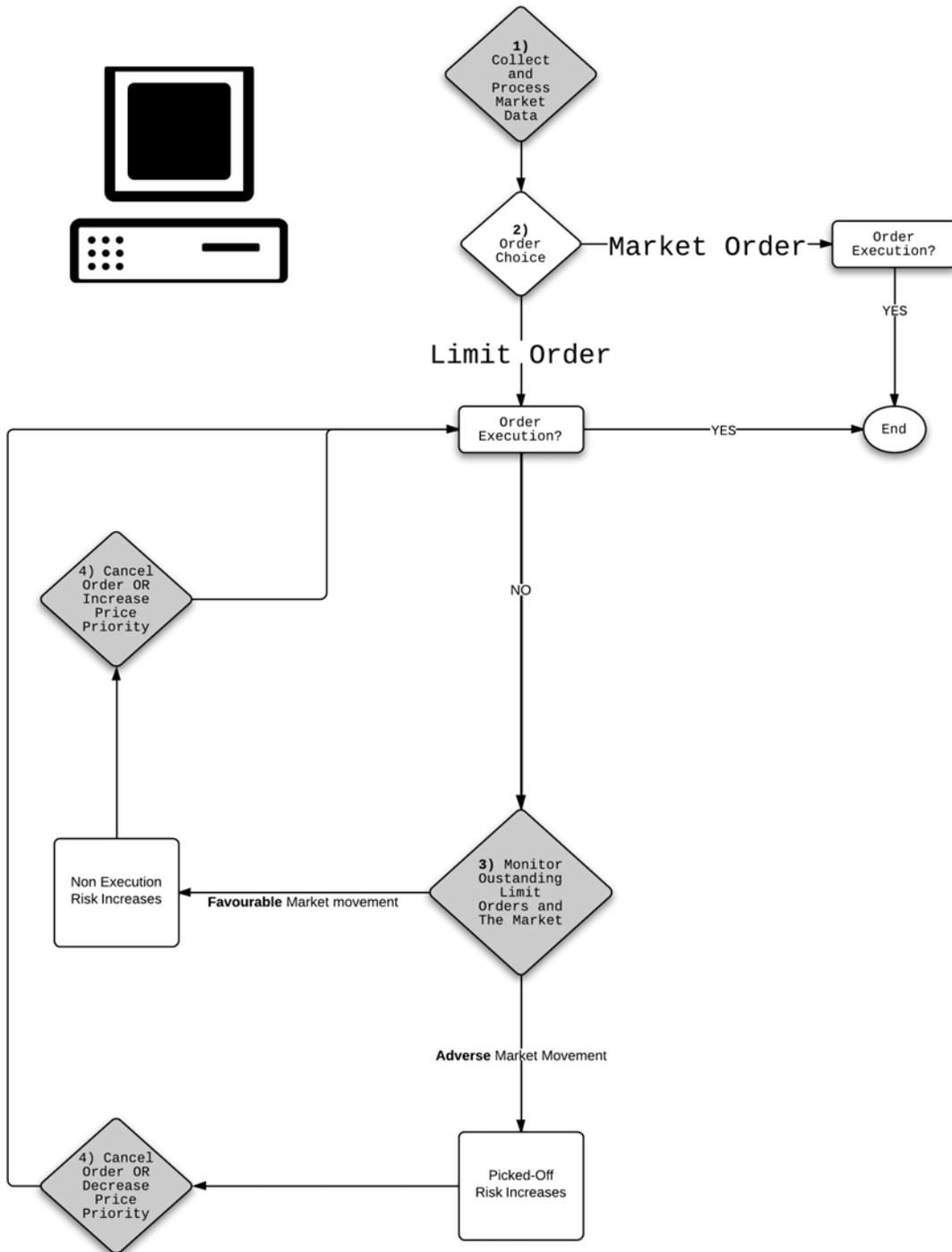
- Heckman, J., Ichimura, H., Smith, J., and Todd, P. (1998). "Characterizing selection bias using experimental data". *Econometrica*.
- Heckman, J., Ichimura, H., Smith, J., and Todd, P. (1996). "Sources of selection bias in evaluating social programs: An interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation method". *Proceedings of the National Academy of Sciences of the United States of America* 93.23.
- Hendershott, T. and Moulton, P. C. (Nov. 2011). "Automation, speed, and stock market quality: The NYSE's Hybrid". *Journal of Financial Markets* 14.4.
- Hodge, F., Kennedy, J., and Maines (2004). "Does search-facilitating technology improve the transparency of financial reporting?" *The Accounting Review*.
- Hollifield, B., Miller, R. A., Sandas, P., and Slive, J. (2006). "Estimating the Gains from Trade in Limit-Order Markets". *The Journal of Finance* 61.6.
- Kaufmann, C., Weber, M., and Haisley, E. (Dec. 2012). "The Role of Experience Sampling and Graphical Displays on One's Investment Risk Appetite". *Management Science* 59.2.
- Kelley, K. and Tetlock, P. (2013). "How wise are crowds? Insights from retail orders and stock returns". *The Journal of Finance* January.
- Linnainmaa, J. (2010). "Do limit orders alter inferences about investor performance and behavior?" *The Journal of Finance* LXV.4.
- Liu, W. (Feb. 2009). "Monitoring and limit order submission risks". *Journal of Financial Markets* 12.1.
- Looney, C. A. and Hardin, A. M. (Aug. 2009). "Decision Support for Retirement Portfolio Management: Overcoming Myopic Loss Aversion via Technology Design". *Management Science* 55.10.
- Lusardi, A. and Mitchell, O. (2011). "Financial literacy around the world: an overview". *Journal of Pension Economics and Finance* 10.4.
- Maines, L. and McDaniel, L. (2000). "Effects of Comprehensive-Income Characteristics on Nonprofessional Investors' Judgments: The Role of Financial-Statement Presentation Format." *Accounting Review* 75.2.

- Peng, L. and Xiong, W. (June 2006). "Investor attention, overconfidence and category learning". *Journal of Financial Economics* 80.3.
- Rogers, W. (1994). "Regression standard errors in clustered samples". *Stata technical bulletin*.
- Rubaltelli, E., Rubichi, S., Savadori, L., Tedeschi, M., and Ferretti, R. (Mar. 2005). "Numerical Information Format and Investment Decisions: Implications for the Disposition Effect and the Status Quo Bias". *Journal of Behavioral Finance* 6.1.
- Russo, J. (1977). "The value of unit price information". *Journal of Marketing Research*.
- Saez, E. (Jan. 2009). "Details Matter: The Impact of Presentation and Information on the Take-up of Financial Incentives for Retirement Saving". *American Economic Journal: Economic Policy* 1.1.
- Schlarbaum, G., Lewellen, W., and Lease, R. (1978). "Realized returns on common stock investments: The experience of individual investors". *Journal of Business* 51.2.
- Stuart, E. A. (Feb. 2010). "Matching methods for causal inference: A review and a look forward." *Statistical Science* 25.1.
- Todd, P. (2007). "Evaluating social programs with endogenous program placement and selection of the treated". *Handbook of development economics*.
- White, H. (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity". *Econometrica* 48.4.
- Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data: Second Edition*. Mit Press.
- Xu, L. and Zia, B. (2012). "Financial literacy around the world: an overview of the evidence with practical suggestions for the way forward". *World Bank Policy Research Working Paper* June.

## 1.9 Figures

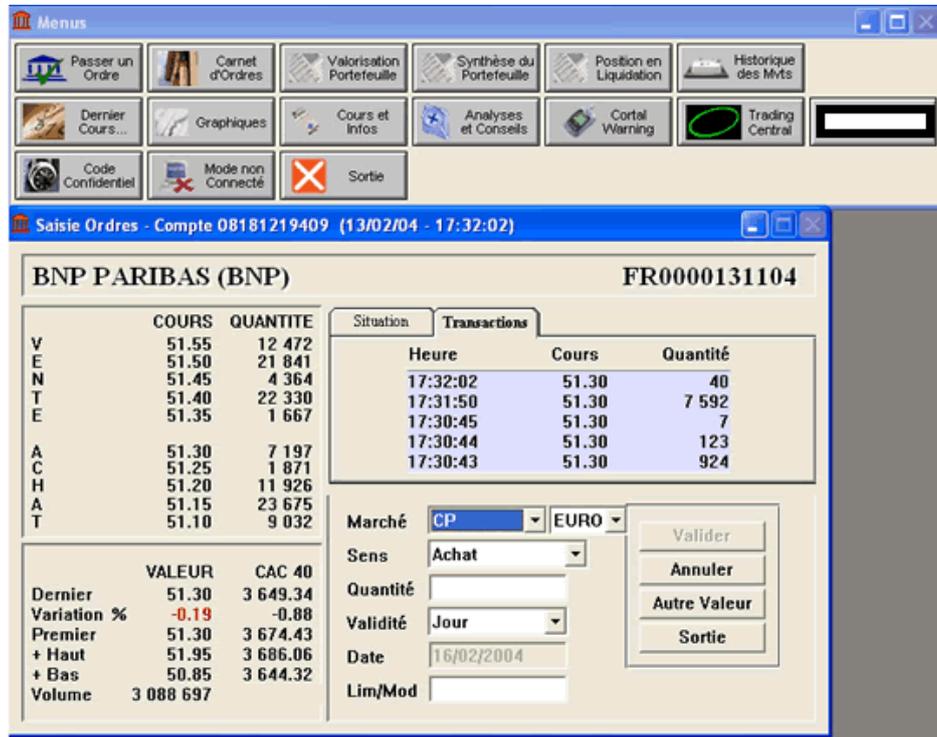
Figure 1.1: Investors' decision tree

This is a decision tree of order choice, revision and cancellation (adapted from Fong and Liu [2010]). Investors need to collect and process market data (step 1) and optimally decide which order type to use for his trade (step 2: market order or limit order). If the investor chooses a market order then the order is immediately executed and the trade is terminated. If the investor chose a limit order then, depending on the limit order price, the order may or may not be executed. A limit order not executed joins the limit order book according to time and price priorities. As the order may be picked-off by other investors and face the risk of non-execution, the investor must monitor the market and eventually revise or cancel his orders (steps 3 and 4). I marked in gray all the steps improved by Trader+.

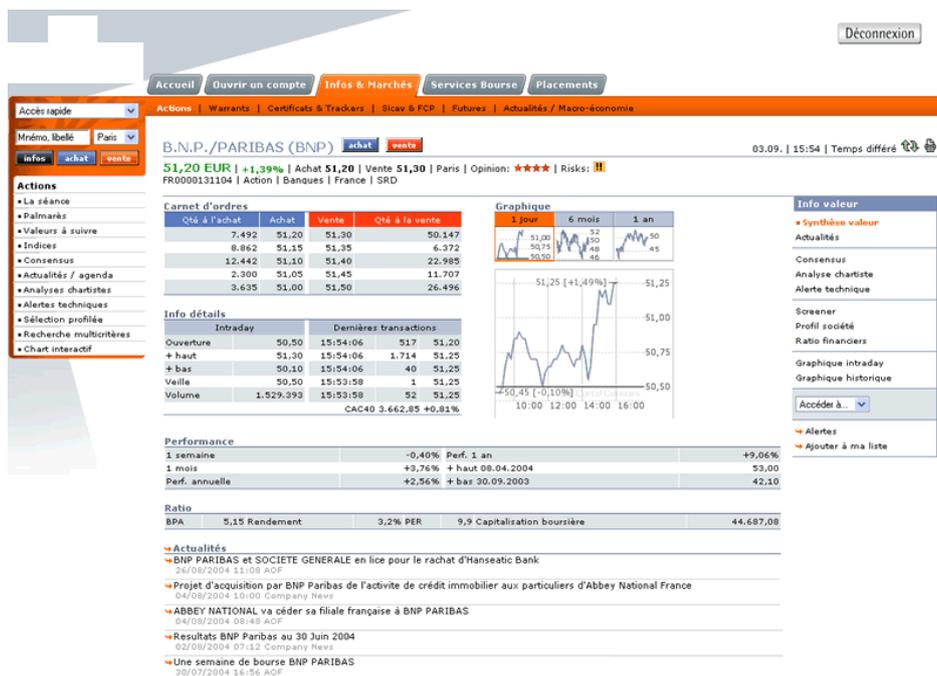


## Figure 1.2: Main trading channels used by the broker customers before Trader+

This Figure shows two screenshots of the main trading channels used by investors in my sample, before the introduction of Trader+. Investors could send their orders using Online+, a basic trading software that has been available since 1999, or using the Web brokerage interface (Web). The same amount of market information is provided on each trading channel.



(a) Online+



(b) Web

Figure 1.3: Screenshot of Trader+

This picture is taken from the official Trader+ user guide made available to the broker customers in June 2003. It is a screenshot of the different windows available when using Trader+: the pre-orders window, the limit order book, the intraday graphics windows, the market newsfeed, the market data window, the equity research box and the trade order forms.

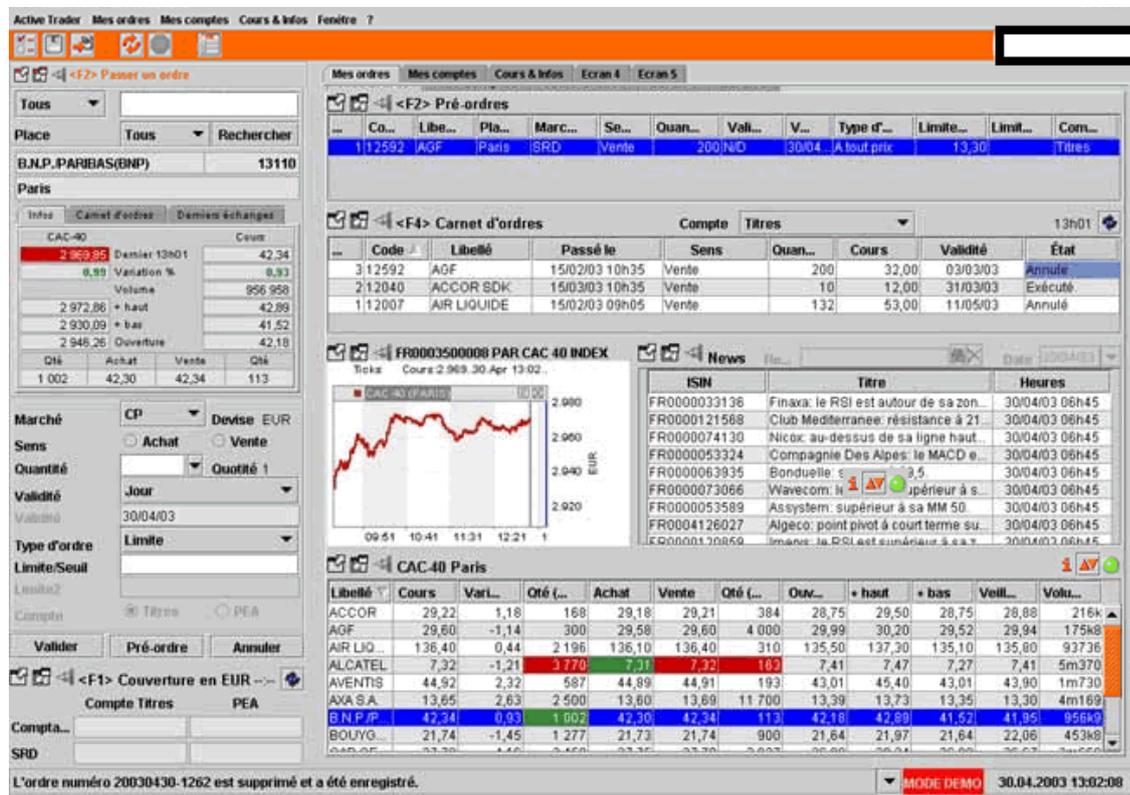
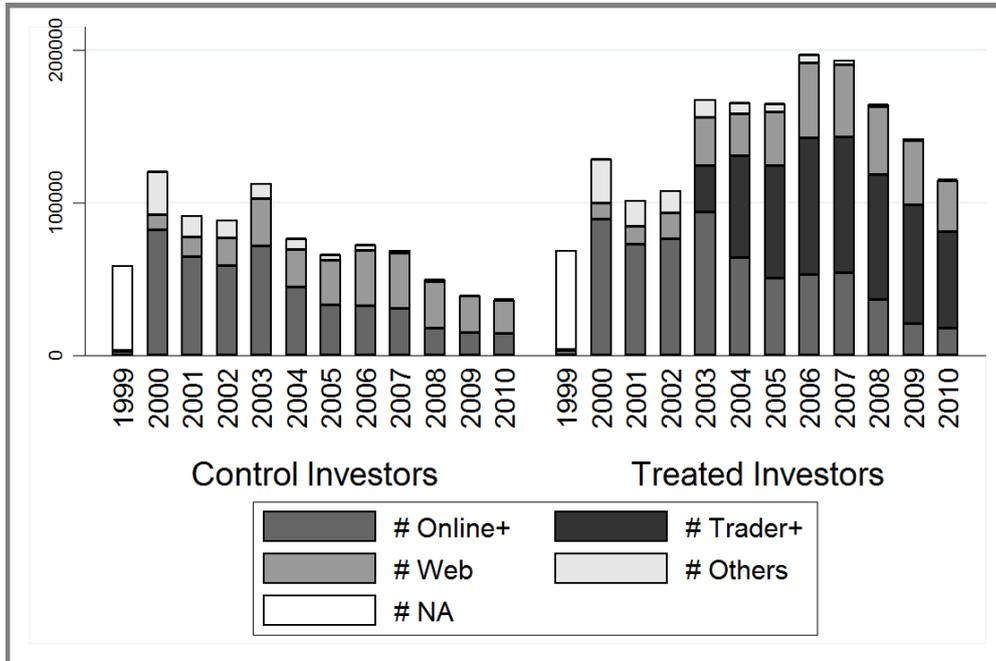
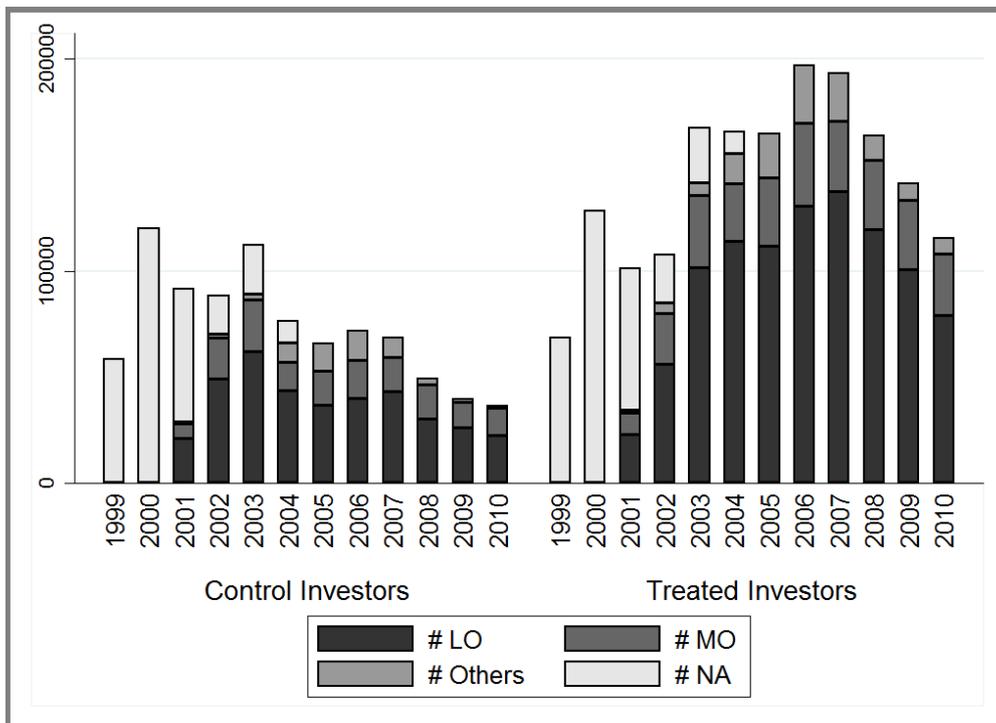


Figure 1.4: Orders submitted by treated and matched control investors, 1999-2010

For each year (on the x axis) this Figure shows (on the Y axis) the total number of trades submitted by treated or matched control investors that year. This number is then broken down according to order type or trading channel. Each bar thus is a sum of stacked bars that give the corresponding number of orders of a given order type or trading channel. “LO” stands for limit orders and “MO” for market orders. “Other” refers to minor other order types. Missing values are labeled “NA”.



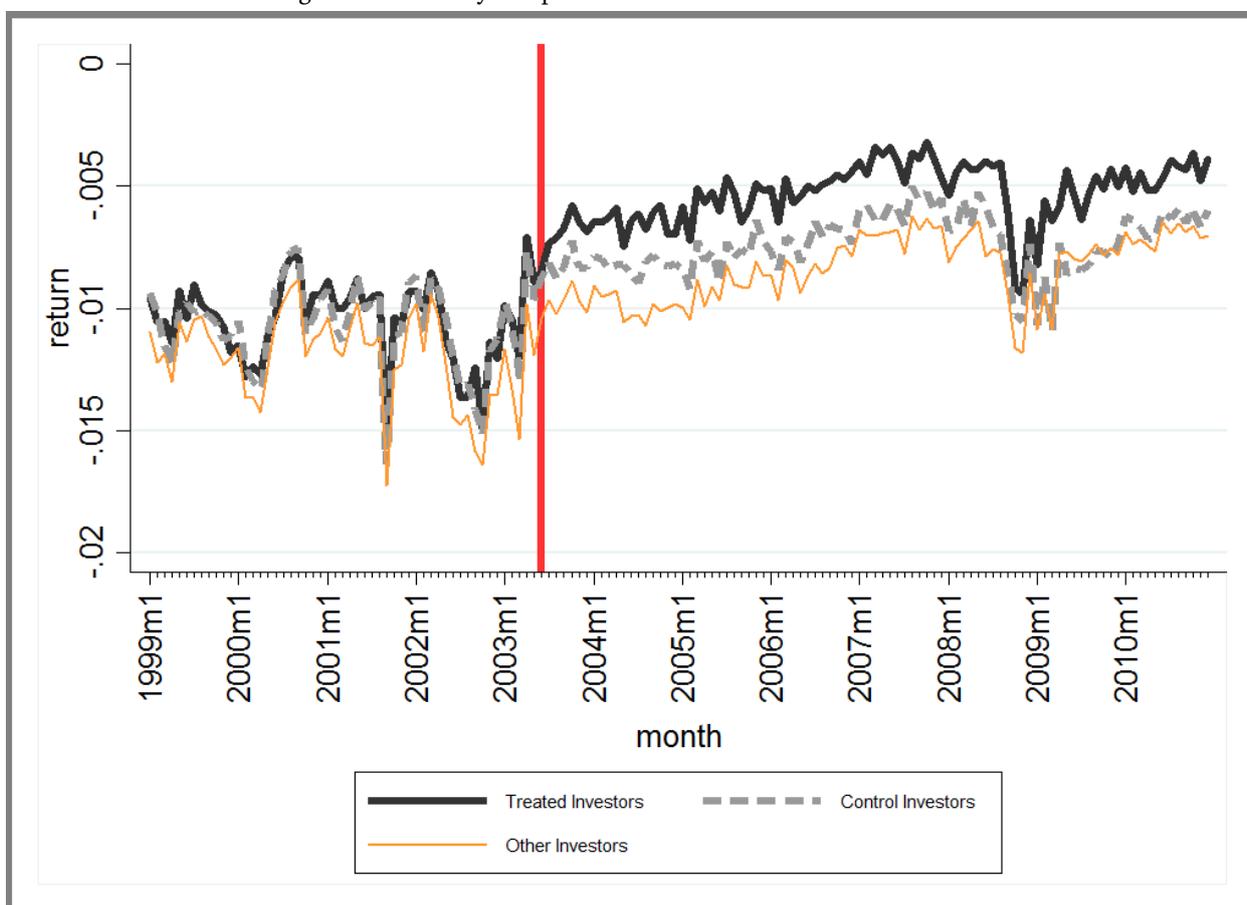
(a) Trading Channel



(b) Order Type

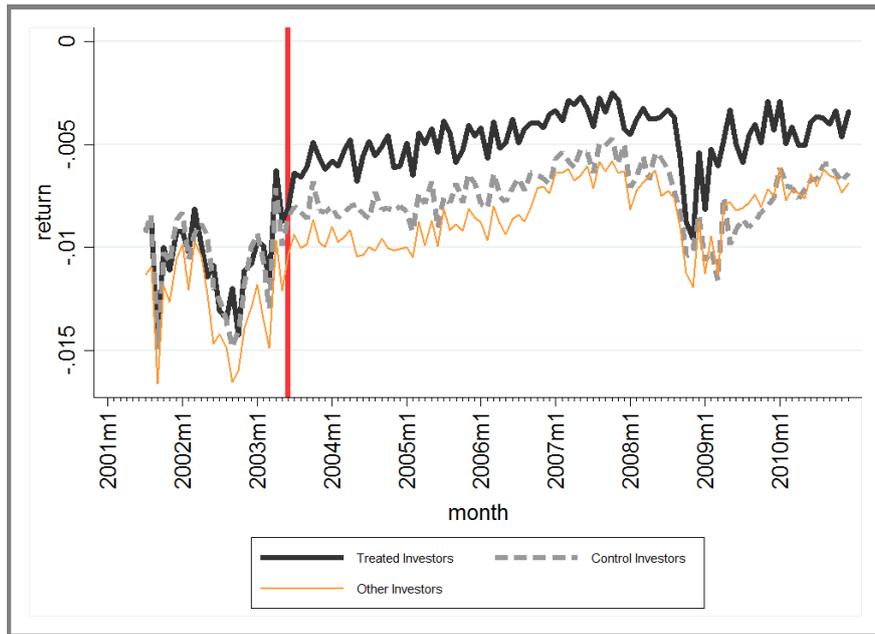
### Figure 1.5: Monitoring activity and performance

I compute and plot the monthly average of daily returns achieved by treated, control and other investors in my sample. The vertical red line identifies the introduction of Trader+ by the brokerage house in June 2003. All investors have submitted at least one trade between March 2003 and the end of May 2003, have an open common stock position between 2002 and 2005 and have submitted at least one trade after June 2003. Treated investors are investors who submitted at least one trade using Trader+ after June 2003. Control investors have never used Trader+ and are matched one-to-one to treated investors with a propensity score methodology. Other control investors are the remaining investors in my sample.

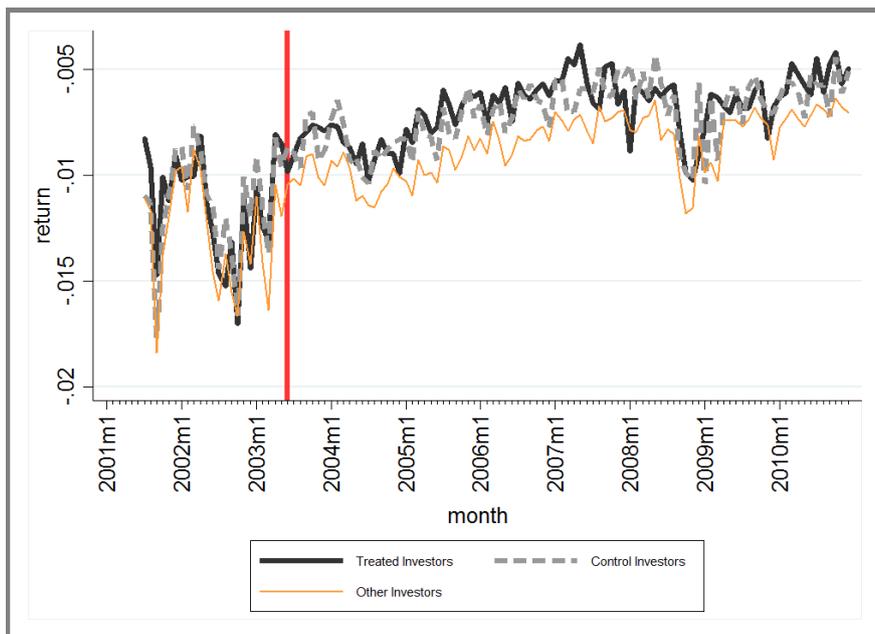


## Figure 1.6: Monitoring activity and performance: market orders vs. limit orders

I compute and plot the monthly average of daily returns achieved by treated, control and other investors in my sample. The computation only includes executed limit orders in the top Figure, and only includes executed market orders in the bottom Figure. The vertical red line identifies the introduction of Trader+ by the brokerage house in June 2003. All investors have submitted at least one trade between March 2003 and the end of May 2003, have an open common stock position between 2002 and 2005 and have submitted at least one trade after June 2003. Treated investors are investors who submitted at least one trade using Trader+ after June 2003. Control investors have never used Trader+ and are matched one-to-one to treated investors with a propensity score methodology. Other control investors are the remaining investors in my sample.



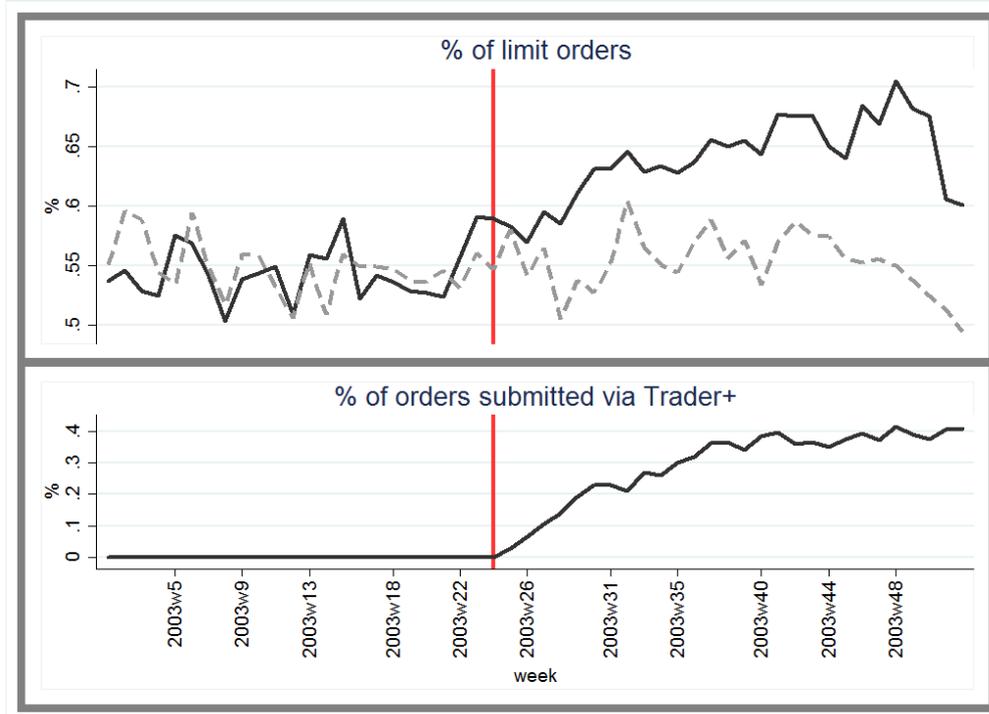
(a) Returns - Limit Orders Only



(b) Returns - Market Orders Only

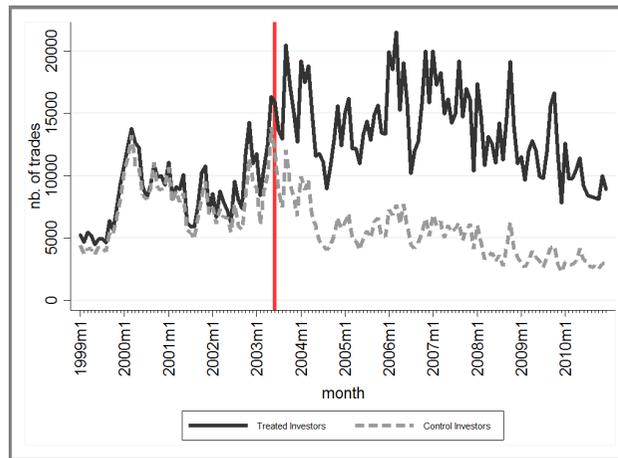
## Figure 1.7: Monitoring activity and order choice

I compute and plot the weekly percentage of executed orders that are limit orders submitted by treated and control investors in my sample. The vertical red line identifies the introduction of Trader+ by the brokerage house in June 2003. All investors have submitted at least one trade between March 2003 and the end of May 2003, have an open common stock position between 2002 and 2005 and have submitted at least one trade after June 2003. Treated investors are investors that have submitted at least one trade using Trader+ after June 2003. Control investors have never used Trader+ and are matched one-to-one to treated investors with a propensity score methodology.

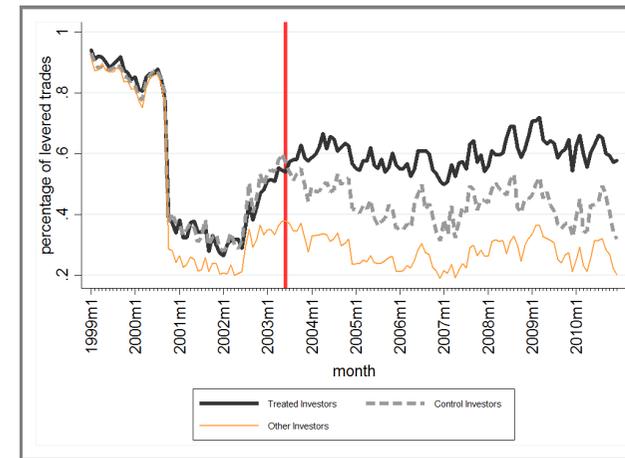


## Figure 1.8: Monitoring activity and trading volume

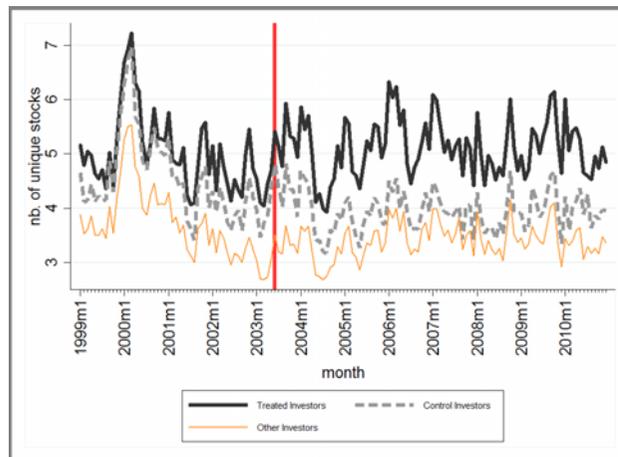
I compute and plot a) the monthly number of trades, b) the monthly percentage of leveraged orders (through margin account), c) the number of distinct stocks and d) the monthly median amount traded by treated, control and other investors in my sample. The vertical red line identifies the introduction of Trader+ by the brokerage house in June 2003. All investors have submitted at least one trade between March 2003 and the end of May 2003, have an open common stock position between 2002 and 2005 and have submitted at least one trade after June 2003. Treated investors are investors who submitted at least one trade using Trader+ after June 2003. Control investors have never used Trader+ and are matched one-to-one to treated investors with a propensity score methodology. Other control investors are the remaining investors in my sample.



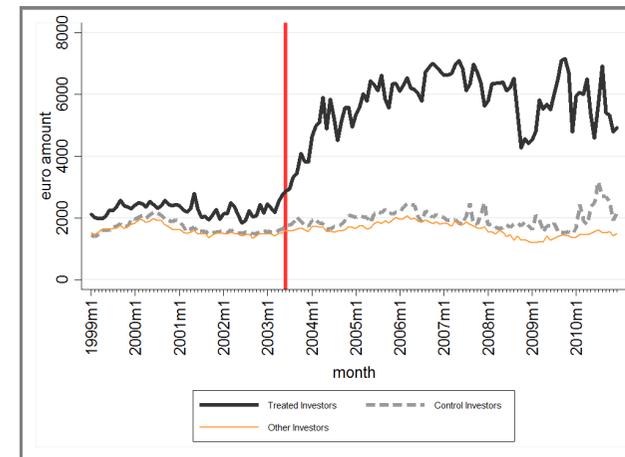
(a) Nb. of Trades



(b) Percentage of Leveraged Orders



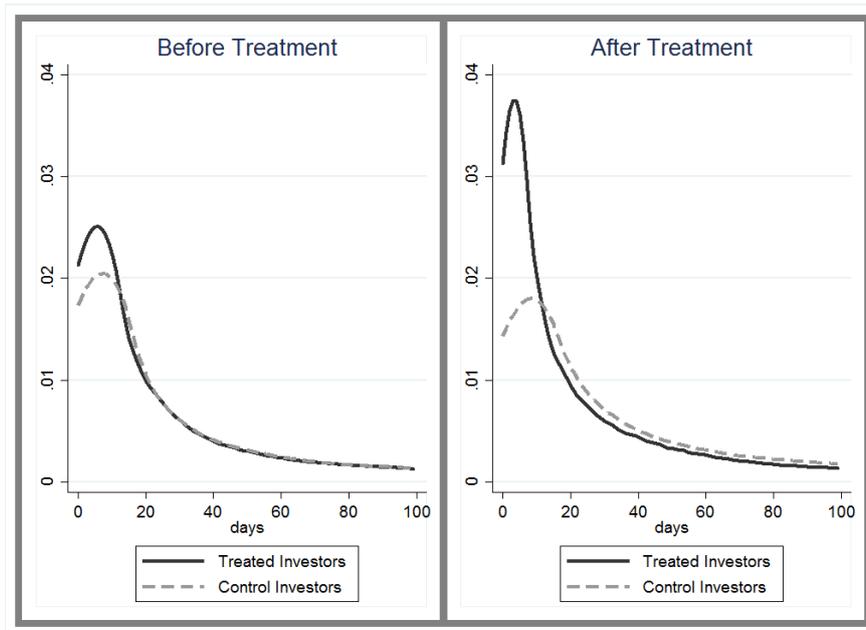
(c) Average Nb. of Distinct Stocks Traded



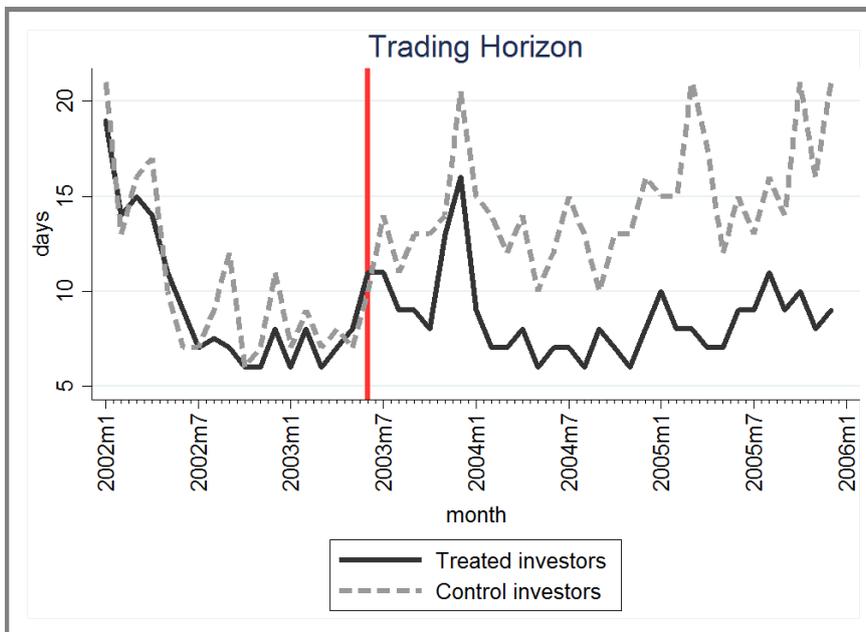
(d) Median Euro Trade Amount

## Figure 1.9: Monitoring activity and trading horizon

I sort all the trades in my database by investor, stock and trading day and I keep track of the number of stocks held by an investor at any time. A round-trip on a given stock starts and ends with a zero net inventory position on that stock. The duration of a round-trip is then the number of days between those two events. Using data in the pre-treatment period (January 2001-June 2003) and the post-treatment period (June 2003 - December 2006), I generate in Subfigure a the densities of investors' round-trip durations (in days). In Subfigure b I compute the monthly median duration of round-trips executed by treated and control investors. The sample in Subfigure b contains all the round-trips whose duration is lower than 100 days.



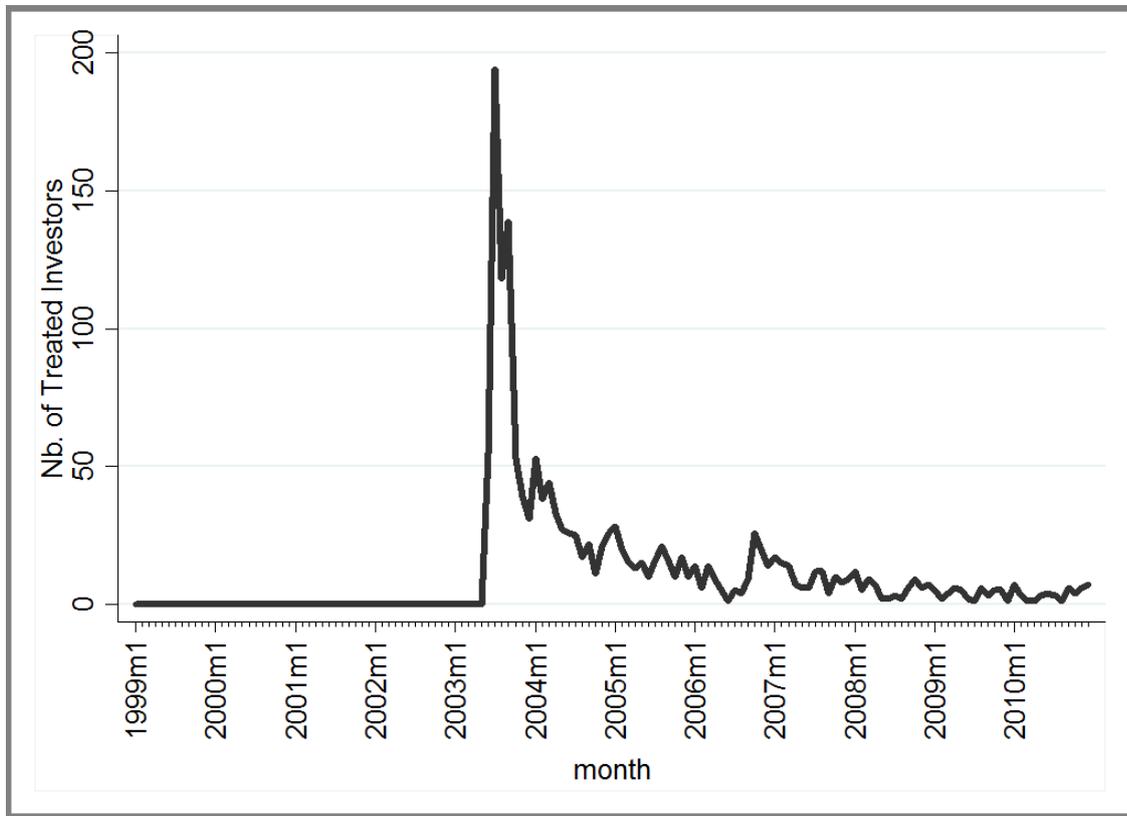
(a) Density of Trading Horizon in Days



(b) Median Length in Days for Short-Term Strategies

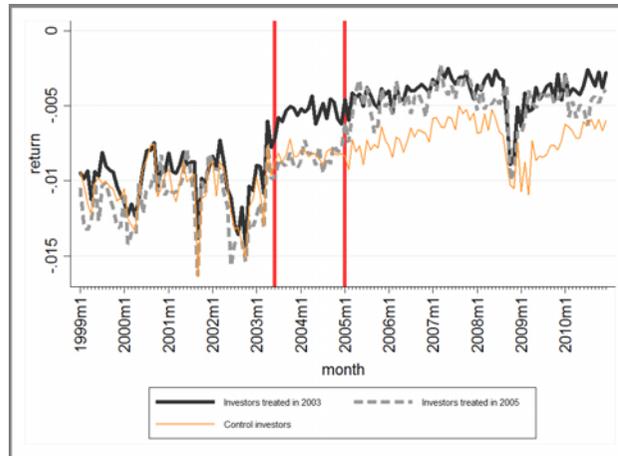
Figure 1.10: Dynamics of entry into treatment

This Figure shows the total number of (treated) investors in a given month that have submitted their first trade with Trader+ that month.



## Figure 1.11: Causes before consequences, or consequences before causes?

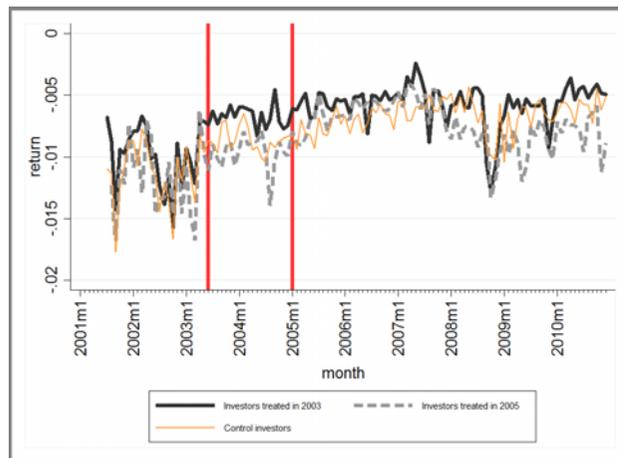
I identify year-cohorts of treated investors, according to the year in which investors submit their first trade using Trader+. This figure compares the trends in outcomes between the first cohort of treated investors (that switched to Trader+ in 2003) and investors treated two years later, in 2005.



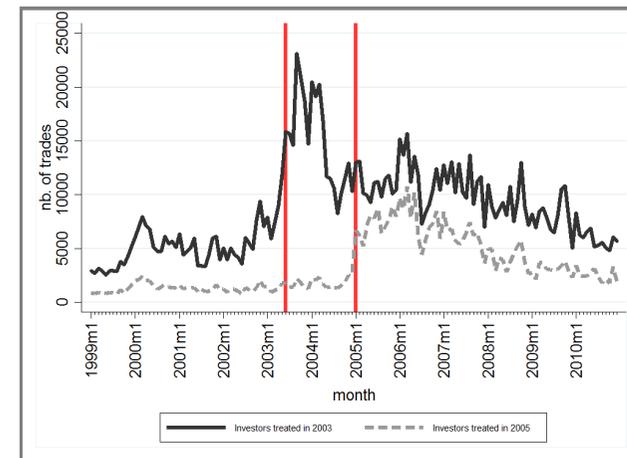
(a) Daily Returns - All Orders



(b) Daily Returns - Limit Orders only



(c) Daily Returns - Market Orders only



(d) Nb. of trades

## 1.10 Tables

## Table 1.1: Summary statistics of the brokerage dataset

This table describes the main characteristics of the brokerage dataset used in this paper. The table gives the corresponding number of trades, percentage and cumulative percentage for each category of information (investors' order choices, investors' use of trading channels, investors' use of asset classes and investors' use of trading exchanges). The total number of trades may be different from one category to another due to missing values. The sample period is from 1999 to 2010.

Panel A: Order Choice			
	Frequency	Percentage	Cum. Percentage
Limit Order	6434646	62.6	62.6
Market Order	2918369	28.4	91.0
Others/Missing	920520	8.96	100
Total	10273535	100	

Panel B: Trading Channels			
	Frequency	Percentage	Cum. Percentage
Web	5785047	42.1	42.1
Online+	4103445	29.8	71.9
Trader+	2049605	14.9	86.8
Minitel	782961	5.69	92.5
Telephone	686728	4.99	97.5
Phone+	181141	1.32	98.8
Manual	165430	1.20	100
Total	13754357	100	

Panel C: Asset Classes			
	Frequency	Percentage	Cum. Percentage
Common Stocks	12986031	86.2	86.2
Others	824604	5.47	91.6
Equity Warrants	488490	3.24	94.9
Index Warrants	365357	2.42	97.3
Turbo Warrants	270680	1.80	99.1
ETF	134270	0.89	100.0
Total	15069432	100.0	

Panel D: Trading Exchanges			
	Frequency	Percentage	Cum. Percentage
NYSE Euronext Paris	13888529	92.16	92.16
Tradegate	669061	4.44	96.60
Others	437689	2.90	99.51
Nasdaq	74153	0.49	100.00
Total	15069432	100.00	

Nb. of distinct investors	145801
Nb. of buy order	7910252
Nb. of sell orders	7159180

**Table 1.2: Estimation of the probability of using Trader+ at least once after June 2003 (using pre-treatment data only)**

This table gives the estimates from a cross-section logistic regression where the dependent variable is a dummy variable that equals 1 if an investor is considered as treated: he has submitted at least one trade using Trader+ during the period June 2003-December 2010. Investors in my sample are required to have an open common stock position at least between 2002 and 2005, they have submitted at least one trade between March and the end of May 2003, and at least one trade after June 2003. I compute various covariates according to investor demographics, general trading behavior, account size and trading channel preferences in March, April and May 2003. I use two demographic variables: Female is a dummy variable that equals one for females and 0 for males, and Age correspond to the investor's age in 2003. The variables that capture investor trading behavior, over the three month pre-treatment period, are the number of orders executed ("Nb of executed trades"), the number of different asset classes traded ("Nb. of asset types"), the mean euro amount traded ("Mean amount traded"), the mean daily-return on executed orders ("Mean daily return"), the percentage of orders that are limit orders ("% Limit orders") and the percentage of orders submitted with a margin account ("% Margin"). Account size variables give information about the total euro value of an investor's common stock holdings on both his spot-market account and his margin account (in tens of thousands of euros) in May 2003. Lastly, trading channels variables provide, for each trading channel, the corresponding percentages of trades submitted by an investor through that channel.

	Logit	
<b><u>Demographics</u></b>		
Female	-0.349***	(0.000)
Age in 2003	0.003	(0.163)
<b><u>Trading Behavior (March-May 2003)</u></b>		
Nb. of executed trades	0.008***	(0.000)
Nb. of asset types	-0.013	(0.412)
Mean amount traded	0.042***	(0.000)
Mean daily return	4.319***	(0.001)
% Limit orders	-0.054	(0.427)
% Margin	0.959***	(0.000)
<b><u>Account size in May 2003</u></b>		
market value (spot)	0.028***	(0.000)
market value (margin)	0.003	(0.826)
<b><u>Use of Trading Channels (March-May 2003)</u></b>		
% Phone+	0.358*	(0.091)
% Minitel	-0.239	(0.288)
% Online+	1.572***	(0.000)
% Web	0.875***	(0.000)
Constant	-3.334***	(0.000)
<i>N</i>	11282	
pseudo $R^2$	0.103	

**Table 1.3: Comparisons between treated investors and control investors obtained via propensity matching**

This table compares the averages covariates between treated and control investors before and after the matching. All computations are performed with trading data coming from the pre-treatment period March-May 2003. All variables are defined in [Table 1.2](#). Standard errors are between parentheses.

	<u>Before Matching</u>		<u>After Matching</u>	
	Raw Control	Raw Treated	Matched Control	Matched Treated
<b><u>Demographics</u></b>				
Female	0.186 (0.389)	0.126 (0.332)	0.155 (0.362)	0.126 (0.332)
Age in 2003	47.97 (13.68)	47.72 (12.43)	48.10 (13.39)	47.72 (12.43)
<b><u>Trading Behavior (March-May 2003)</u></b>				
Nb. of executed trades	9.536 (23.90)	25.57 (50.42)	21.23 (49.00)	25.57 (50.42)
Nb. of asset types	2.502 (1.712)	2.707 (1.908)	2.667 (1.823)	2.707 (1.908)
Mean amount traded	2.043 (3.235)	3.461 (7.447)	3.186 (5.956)	3.461 (7.447)
Mean daily return	-0.0155 (0.0275)	-0.0103 (0.0186)	-0.0104 (0.0222)	-0.0103 (0.0186)
% Limit orders	0.572 (0.444)	0.589 (0.408)	0.583 (0.424)	0.589 (0.408)
% Margin	0.160 (0.326)	0.338 (0.420)	0.354 (0.428)	0.338 (0.420)
<b><u>Account size in May 2003</u></b>				
market value (spot)	2.342 (4.538)	3.236 (5.422)	3.226 (7.290)	3.236 (5.422)
market value (margin)	0.177 (1.481)	0.795 (5.265)	0.567 (3.071)	0.795 (5.265)
<b><u>Use of Trading Channels (March-May 2003)</u></b>				
% Phone+	0.0744 (0.248)	0.0400 (0.177)	0.0431 (0.189)	0.0400 (0.177)
% Minitel	0.0862 (0.272)	0.0295 (0.156)	0.0262 (0.150)	0.0295 (0.156)
% Online+	0.327 (0.457)	0.550 (0.473)	0.564 (0.478)	0.550 (0.473)
% Web	0.409 (0.475)	0.340 (0.448)	0.330 (0.451)	0.340 (0.448)
Nb. unique investors	9668	1618	1618	1618

Table 1.4: Summary statistics on investors' trading channels and investors' order choice

Panel A shows the main quantiles, the mean and the standard deviation (sd) of investors' intraday returns, tabulated by trading channel. Panel B is a two-way table that provides, for each trading channel, the corresponding row percentage (top), column percentage (middle) and frequency of orders of a given order type (bottom). The sample contains all the orders of both treated and control investors after June 2003.

Panel A: Intraday Returns vs. Trading Channel							
	p10	p25	p50	mean	p75	p90	sd
Trader+	-0.024	-0.011	-0.003	-0.005	0.004	0.014	0.019
Online+	-0.026	-0.013	-0.005	-0.006	0.003	0.012	0.019
Web	-0.028	-0.014	-0.005	-0.007	0.002	0.013	0.020
Minitel	-0.024	-0.014	-0.006	-0.007	0.001	0.010	0.017
Phone+	-0.026	-0.015	-0.007	-0.008	-0.000	0.009	0.018
Telephone	-0.034	-0.018	-0.008	-0.011	0.000	0.010	0.022
Total	-0.026	-0.013	-0.004	-0.006	0.003	0.013	0.019

Panel B: Trading Channel vs. Order Choice					
	Limit Order	Market Order	Others	Missing	Total
Trader+	71.2	20.6	8.0	0.2	100.0
	36.2	32.6	27.3	2.3	33.6
	406675	117751	45999	1007	571432
Online+	67.7	14.1	12.2	6.1	100.0
	34.0	21.9	40.8	77.7	33.2
	381804	79343	68737	34468	564352
Web	61.1	29.0	9.3	0.7	100.0
	27.9	41.1	28.4	7.6	30.2
	313338	148755	47708	3369	513170
Minitel	60.0	5.2	19.8	15.0	100.0
	0.8	0.2	1.9	5.4	0.9
	9480	818	3121	2374	15793
Phone+	42.1	28.6	13.5	15.7	100.0
	0.6	1.3	1.3	5.8	1.0
	6950	4720	2230	2591	16491
Telephone	35.1	58.9	2.8	3.2	100.0
	0.5	2.8	0.3	1.2	1.0
	6045	10161	487	550	17243
Total	66.2	21.3	9.9	2.6	100.0
	100.0	100.0	100.0	100.0	100.0
	1124292	361548	168282	44359	1698481

Table 1.5: Does the display format of market data matters for retail investors?

This table show the results of the following difference-in-difference OLS regression :

$$Y_{i,t,k} = \alpha + \beta * Monitoring_{i,t} + \gamma * Month_t + \delta_i + \epsilon_{i,t,k} \quad (1.3)$$

$$Y_{i,m} = \alpha + \beta * Monitoring_{i,m} + \gamma * Month_m + \delta_i + \epsilon_{i,m} \quad (1.4)$$

$Monitoring_{i,t}$  (resp.  $Monitoring_{i,m}$ ) is a (treatment) dummy variable that equals one, for investor  $i$ , for all days  $t$  (resp. months  $m$ ) that follows his first trade executed through Trader+ (including the first day of use).  $\delta_i$  is an individual fixed effect and  $Month$  is a monthly fixed effect. Definitions of variables are provided in Table 1.9. The sample contains the orders of both the treated and the control group. The sample period is from January 2002 to January 2006. Returns are multiplied by 100 and standard errors (in parenthesis) are clustered at the individual level. \*, \*\*, \*\*\* denote significance at the 10%, 5% and 1% levels.

Panel A						
	All Orders	Limit	Market	LO dummy	Amount	Margin
Monitoring	0.0690*** (0.021)	0.0961*** (0.024)	0.00641 (0.055)	0.0782*** (0.013)	752.5*** (217.305)	0.0756*** (0.014)
Constant	-0.844*** (0.023)	-0.785*** (0.032)	-0.956*** (0.042)	0.573*** (0.012)	3525.5*** (200.585)	0.402*** (0.012)
Nb. obs	947726	571635	190442	947726	947726	947726
R-Square	0.00667	0.00562	0.00845	0.00845	0.00816	0.0128
Nb. investors	3236	3112	2487	3236	3236	3236
Avg. obs per investor	292.9	183.7	76.57	292.9	292.9	292.9
Panel B						
	Nb. Trades	LO Trades	MO Trades	Nb. Unique	Horizon	Round-Trip
Monitoring	4.466*** (1.261)	4.567*** (1.229)	0.656*** (0.245)	0.451*** (0.123)	-3.671*** (0.547)	0.0235*** (0.004)
Constant	9.791*** (0.461)	5.481*** (0.431)	1.942*** (0.114)	4.963*** (0.096)	32.65*** (0.844)	0.0215*** (0.004)
Nb. obs	88756	88756	88756	88756	39449	39449
R-Square	0.0115	0.00909	0.00566	0.0254	0.0291	0.00985
Nb. investors	3236	3236	3236	3236	2787	2787
Avg. obs per investor	27.43	27.43	27.43	27.43	14.15	14.15

**Table 1.6: Robustness: standard difference-in-difference estimation**

This table show the results of the following difference-in-difference OLS regression :

$$Y_{i,t,k} = \alpha + \beta * Treated_i + \gamma * Post_t + \lambda * Treated_i * Post_t + \epsilon_{i,t,k} \quad (1.5)$$

$$Y_{i,m} = \alpha + \beta * Treated_i + \gamma * Post_m + \lambda * Treated_i * Post_m + \epsilon_{i,m} \quad (1.6)$$

Treated is a dummy variable that equals one for a treated investor and Post is a dummy variable that equals one if a trade is executed after June 2003. The difference-in-difference estimate is captured by the coefficient on the interaction variable Treated\*Post. The sample period is from January 2002 to January 2006. Returns are multiplied by 100 and standard errors (in parenthesis) are clustered at the individual level. \*, \*\*, \*\*\* denote significance at the 10%, 5% and 1% levels.

Panel A	All Orders	Limit	Market	LO dummy	Amount	Margin
Post	0.282*** (0.036)	0.267*** (0.058)	0.272*** (0.047)	0.00845 (0.019)	761.6*** (169.033)	0.00722 (0.025)
Treated	0.0139 (0.062)	0.0293 (0.095)	-0.0725 (0.104)	-0.0183 (0.032)	371.7 (232.349)	-0.0250 (0.032)
Post*Treated	0.171*** (0.051)	0.228*** (0.070)	0.0956 (0.083)	0.134*** (0.028)	1535.0*** (362.403)	0.148*** (0.033)
Constant	-1.090*** (0.053)	-1.064*** (0.091)	-1.089*** (0.047)	0.549*** (0.027)	3099.8*** (193.176)	0.462*** (0.025)
Nb observations	947726	571635	190442	947726	947726	947726
R-Square	0.008	0.010	0.005	0.017	0.018	0.019

Panel B	Nb. Trades	LO Trades	MO Trades	Nb. Unique	Horizon	Round-Trip
Post	-0.729** (0.362)	-0.336 (0.289)	-0.210 (0.160)	0.00643 (0.085)	5.723*** (0.510)	-0.0211*** (0.003)
Treated	1.501** (0.710)	0.594 (0.572)	0.292 (0.273)	0.603*** (0.165)	-0.526 (0.628)	0.00938** (0.004)
Post*Treated	3.463*** (0.846)	3.684*** (0.789)	0.417* (0.244)	0.435*** (0.133)	-3.313*** (0.681)	0.0176*** (0.005)
Constant	9.041*** (0.566)	4.965*** (0.483)	1.967*** (0.188)	4.001*** (0.109)	23.52*** (0.451)	0.0537*** (0.003)
Nb observations	88756	88756	88756	88756	39449	39449
R-Square	0.006	0.006	0.001	0.008	0.009	0.005

## Table 1.7: Robustness: placebo test

This table show the results of the following difference-in-difference OLS regression :

$$Y_{i,t,k} = \alpha + \beta * Treated_i + \gamma * FakePost_t + \lambda * Treated_i * FakePost_t + \epsilon_{i,t,k} \quad (1.7)$$

$$Y_{i,m} = \alpha + \beta * Treated_i + \gamma * FakePost_m + \lambda * Treated_i * FakePost_m + \epsilon_{i,m} \quad (1.8)$$

Treated is a dummy variable that equals one for a treated investor and FakePost is a dummy variable that equals one if a trade is executed after October 2002. The difference-in-difference estimate is captured by the coefficient on the interaction variable Treated\*FakePost. The sample period is from April 2002 to April 2003. Standard errors in Panel A (in parenthesis) are clustered at the individual level. \*, \*\*, \*\*\* denote significance at the 10%, 5% and 1% levels.

Panel A	All Orders	Limit	Market	LO dummy	Amount	Margin
FakePost	0.00171 (0.052)	0.0309 (0.085)	-0.0144 (0.066)	-0.000537 (0.024)	540.3** (222.362)	0.152*** (0.024)
Treated	-0.0162 (0.097)	-0.0116 (0.158)	-0.0693 (0.119)	-0.0428 (0.042)	398.5* (209.522)	-0.0119 (0.039)
FakePost*Treated	0.0294 (0.064)	0.0273 (0.094)	-0.0334 (0.121)	0.0312 (0.028)	6.635 (262.950)	-0.0232 (0.028)
Constant	-1.197*** (0.087)	-1.176*** (0.151)	-1.191*** (0.062)	0.551*** (0.036)	2631.3*** (170.072)	0.366*** (0.033)
Nb observations	205404	110317	44967	205404	205404	205404
R-Square	0.000	0.000	0.000	0.001	0.003	0.020

Panel B	Nb. Trades	LO Trades	MO Trades	Nb. Unique	Horizon	Round-Trip
FakePost	1.909*** (0.444)	1.046*** (0.348)	0.333* (0.184)	0.208** (0.091)	-2.497*** (0.799)	0.00783 (0.006)
Treated	1.346* (0.795)	0.347 (0.652)	0.374 (0.304)	0.607*** (0.189)	0.150 (0.963)	0.00738 (0.007)
FakePost*Treated	0.701 (0.584)	0.642 (0.439)	-0.107 (0.235)	0.0226 (0.136)	-0.981 (1.059)	0.00508 (0.008)
Constant	7.869*** (0.668)	4.335*** (0.577)	1.798*** (0.222)	3.807*** (0.125)	24.18*** (0.704)	0.0559*** (0.005)
Nb observations	21043	21043	21043	21043	8758	8758
R-Square	0.005	0.003	0.001	0.005	0.004	0.002

## Table 1.8: Heterogeneous treatment effects

I classify each investor in my sample (including both treated and control investors) into three groups defined by the terciles of the distribution of individual Sharpe ratios (computed from trading returns) before June 2003. This table shows the results of the following difference-in-difference OLS regression, estimated separately for each tercile group:

$$Y_{i,t,k} = \alpha + \beta * Monitoring_{i,t} + \gamma * Month_t + \delta_i + \epsilon_{i,t,k} \quad (1.9)$$

The dependent variables are defined in the previous tables.  $Monitoring_{i,t}$  is a (treatment) dummy variable that equals one, for investor  $i$ , for all days  $t$  that follows his first trade executed through Trader+.  $\delta_i$  is an individual fixed effect and  $Month_t$  is a monthly fixed effect. In each regression the sample period is from January 2002 to January 2006. Returns are multiplied by 100 and standard errors (in parenthesis) are clustered at the individual level. \*, \*\*, \*\*\* denote significance at the 10%, 5% and 1% levels.

Investor Group 1	All Orders	Limit	Market	LO dummy	Amount	Margin
Monitoring	0.0142 (0.054)	0.0907* (0.049)	-0.209 (0.179)	0.0854*** (0.026)	570.9* (319.771)	0.0669*** (0.019)
Constant	-1.265*** (0.058)	-1.110*** (0.063)	-1.607*** (0.129)	0.565*** (0.019)	2658.7*** (126.201)	0.353*** (0.017)
Nb. obs	185316	101425	38455	185316	185316	185316
R-Square	0.0258	0.0211	0.0389	0.00778	0.0106	0.0103
Nb. investors	631	620	509	631	631	631
Avg. obs per investor	293.7	163.6	75.55	293.7	293.7	293.7
Investor Group 2	All Orders	Limit	Market	LO dummy	Amount	Margin
Monitoring	0.127*** (0.027)	0.0942*** (0.034)	0.101* (0.052)	0.0820*** (0.024)	1060.7*** (313.762)	0.123*** (0.019)
Constant	-0.817*** (0.038)	-0.814*** (0.058)	-0.832*** (0.051)	0.564*** (0.016)	3038.6*** (369.219)	0.379*** (0.015)
Nb. obs	255871	135770	61191	255871	255871	255871
R-Square	0.00842	0.00757	0.00750	0.00531	0.00806	0.0190
Nb. investors	631	620	533	631	631	631
Avg. obs per investor	405.5	219.0	114.8	405.5	405.5	405.5
Investor Group 3	All Orders	Limit	Market	LO dummy	Amount	Margin
Monitoring	0.161*** (0.030)	0.186*** (0.037)	0.127* (0.071)	0.0676*** (0.025)	594.8 (440.234)	0.0425 (0.032)
Constant	-0.648*** (0.033)	-0.595*** (0.051)	-0.749*** (0.059)	0.534*** (0.025)	4247.5*** (397.355)	0.423*** (0.024)
Nb. obs	358834	231085	66158	358834	358834	358834
R-Square	0.00272	0.00284	0.00475	0.0223	0.0103	0.0179
Nb. investors	630	618	517	630	630	630
Avg. obs per investor	569.6	373.9	128.0	569.6	569.6	569.6

## Table 1.9: Dependent variables

This table describes the dependent variable used in this paper.  
The dependent variables on panel A are:

- 1)  $Y_{i,t,k} = R_{i,t,k}$ : the daily-return of trade  $k$  submitted on day  $t$  by individual  $i$ . For a buy (sell) orders it is (minus) the return from the execution price of the order to the closing price of the stock bought (sold) that day. "**All Orders**" refers to the full sample of orders, "**Limit**" refers to the sample of limit orders only and "**Market**" to the sample of market orders only.
- 3)  $Y_{i,t,k} = LO_{i,t,k}$ : a dummy variable that equals one if order  $k$  submitted by individual  $i$  on day  $t$  is a limit order and zero otherwise (column "**Order Choice**").
- 4)  $Y_{i,t,k} = Amount_{i,t,k}$ : the amount traded on trade  $k$  submitted on day  $t$  by individual  $i$  (column "**Amount**").
- 5)  $Y_{i,t,k} = Margin_{i,t,k}$ : a dummy variable that equals one if order  $k$  submitted by individual  $i$  on day  $t$  is a leveraged order and zero otherwise (column "**Margin**").

The dependent variables on panel B are:

- 1)  $Y_{i,m} = Count_{i,m}$ : the number of trades executed during month  $m$  by individual  $i$ . "**Nb. Orders**" refers to the full sample of orders. "**Nb. limit**" refers to the sample of limit orders only and "**Nb. market**" to the sample of market orders only.
- 2)  $Y_{i,m} = Unique_{i,m}$ : the number of unique stocks traded by individual  $i$  during month  $m$  (column "**Nb. Unique**").
- 3)  $Y_{i,m} = Horizon_{i,m}$ : the average duration of short-term round-trips initiated by individual  $i$  during month  $m$  (column "**Horizon**").
- 4)  $Y_{i,m} = RoundTrip_{i,m}$ : the percentage of round-trips initiated during month  $m$  by individual  $i$  that are initiated and closed the same day (column "**Round-Trip**").

## 2 Chapter Two

## **Tiny Arbitrageurs**

Can liquidity provision to the market be a profitable trading strategy for active retail investors? While a majority of retail investors provide liquidity at its own expense, I find that active investors ranked in the top performance deciles seem to persistently beat the market using these strategies. These top traders earn positive abnormal returns *net of fees* using more limit orders and executing more contrarian strategies than other individual investors. In particular, these traders generate their most profitable round-trip trades at times when arbitrage risk is higher, and thus when institutional investors are likely to be more constrained. Therefore, these traders seem able to identify arbitrage opportunities that would normally be seized by professional arbitrageurs.

## 2.1 Introduction

Recent findings suggest that retail investors use limit orders to provide liquidity to institutional investors who demand immediacy (Kaniel, Saar, and Titman [2008], Kelley and Tetlock [2013]), but according to this view institutional investors should offer price concessions (which generate subsequent return reversals) to retail investors in order to induce them to act as counterparts (Grossman and Miller [1988], Campbell, Grossman, and Wang [1993]).

However, it remains an open question whether retail investors actually profit from these liquidity provision trading strategies, *after transaction costs*. For instance, Kelley and Tetlock [2013] find that buy-sell limit order imbalances of retail investors predict the cross-section of stock returns at monthly horizon but acknowledge that “[*their*] data do not reveal individuals’ portfolios, holding periods, or transactions costs, precluding any direct analysis of gross or net performance of retail investors in [*their*] sample”. Similarly, Kaniel, Saar, and Titman [2008] state that although individual investors trades can forecast stock returns over short horizon, “*this does not necessarily imply that individual investors, who have much longer holding periods, realize abnormal returns*”. These questions are nonetheless important because it is difficult to explain what drives the high degree of heterogeneity in retail trading performance that has been identified in the literature (Grinblatt, Keloharju, and Linnainmaa [2011], Barber et al. [2013]). Moreover, from an investor protection point of view, it is fundamental to know whether retail investors engage or not in trading strategies that almost surely lead to large financial losses.

The focus of this paper is thus to directly test *whether* liquidity provision strategies can be a source of net profits for active retail investors. To do so, I exploit two unique features of a novel and large brokerage dataset that spans over ten years of trading records for more than 140,000 individual investors in France. First, the dataset mostly contains the transactions fees incurred by retail investors at the investor-trade level. Second, active investors in my database seem to behave as “market-makers”: they extensively trade with limit orders and have much shorter holding periods than the other investors in my sample. Hence, by focusing on active investors I

consider the ideal population to study whether providing liquidity to the market can actually benefit to some retail investors, after transaction costs.<sup>1</sup>

To tackle this research question, I need to accurately assess the profitability of the individual trading strategies, which is known to be in itself a challenging problem for researchers. For instance, given the high degree of heterogeneity in trading horizons, motivations for trading and even strategies pursued, many papers impose strong ad-hoc rules such as evaluating the performance of a trade over a fixed horizon defined by the researcher. These assumptions can potentially hide many important features of investor behavior. To overcome these difficulties, I take advantage of the panel structure of my dataset by adapting a methodology first described in [Schlarbaum, Lewellen, and Lease \[1978\]](#). The methodology is straightforward, yet powerful: I identify all the round-trips realized by investors in my database by sorting all their trades by investor, stock and trading day and by keeping track of the number of stocks held by an investor at any time. For a given investor, a round-trip on a given stock thus starts and ends with a zero net inventory position on that stock, and the duration of a round-trip is the number of days between those two events.

By evaluating investors' trading performance in this manner, I am therefore able to compute individual (gross and net) profit and losses that are almost assumption free. Indeed, [Schlarbaum, Lewellen, and Lease \[1978\]](#) argue that *“the dates, prices, and commissions recorded in the transactions file are “hard” data; they represent actual investment activities. As such, they eliminate any need to choose an arbitrary periodic-return measurement interval, to adjust for portfolio changes during that interval, or to allocate transactions costs across intervals by some amortization rule. A realized-return, investment-round-trip orientation, therefore, is persuasively clean.”*. Moreover, as in [Shapira and Venezia \[2001\]](#) and [Chakrabarty, Moulton, and Trzcinka](#)

---

<sup>1</sup>In a very recent working paper, [Barrot, Sraer, and Kaniel \[2014\]](#) also examine whether retail investors are compensated for providing liquidity. Their paper provides an interesting and complementary analysis to my work. They use the same brokerage dataset as me, but consider a different sample of individuals (not only active traders) and use a different methodology that, contrary to my approach, neither use the order type information (market vs. limit orders) nor the trading fees incurred when trading. They find three very interesting results: i) individuals seem to provide liquidity when institutional liquidity is scarce, ii) individuals with a low trading frequency and slow reversal of trades cannot benefit from liquidity provision, iii) fast and more experienced traders tend to earn better returns from liquidity provision relative to slow traders.

[2013], I compute risk-adjusted returns for each individual round-trip by subtracting the return that the individual would have obtained from a passive investment in the market index (over the same duration period) to the (gross or net) return of the round-trip.

Can retail investors profit from liquidity providing strategies? I rank investors in my sample according to the Sharpe ratios of their trading strategies and I find that only retail investors ranked in the top three performance deciles are able to earn positive abnormal returns *net of fees* by providing liquidity to the market. For instance, retail investors in the tenth decile (the highest in terms of performance) earn on average 10 basis points per day (60 basis points more than the worst performing group of investors) using round-trip strategies that only involve the use of limit orders. The economic magnitude of their performance is large: as a comparison the top 500 day traders in Taiwan earn net abnormal returns of approximately 38 basis points (Barber et al. [2013]). Interestingly, these top-traders are significantly more contrarians than any other group of active retail investors. This behavior suggests that these traders behave as market-makers by buying when there is a downward sell price pressure and selling when there is an upward buy price pressure.<sup>2</sup>

To understand whether the results above are driven by skill or luck, I test for persistence in trading performance across my sample of retail investors. I document a significant level of individual performance persistence. Specifically, as in Coval, Hirshleifer, and Shumway [2005] and Seru et al. [2009], I divide my sample in two disjoint time periods (2000-2005 and 2005-2010) and I compute the Spearman rank correlation between the individual Sharpe ratios from the first period to those of the second period. I obtain a highly significant correlation coefficient of 0.20 that is in line with prior research on retail performance persistence. Additionally, using a non-parametric methodology adapted from the mutual fund literature (Brown and Goetzmann [1995]), I show that “winners” (investors with a positive Sharpe ratio) in a given year have a

---

<sup>2</sup> Consistent with my findings, Grinblatt, Keloharju, and Linnainmaa [2011] also document that high IQ investors are more contrarian than low IQ investors, but can only speculate whether this behavior is rational or indicative of a graded behavioral bias. They state that “*by selling stocks at monthly highs and holding stocks at monthly lows, high-IQ investors are more likely to be following a rational liquidity provision strategy than a psychological bias that diminishes returns*”.

higher probability, compared to losing investors, of winning the next year, which is consistent with the recent findings of [Linnainmaa \[2011\]](#) and [Barber et al. \[2013\]](#). Hence, it therefore likely that some investors in this paper have genuine trading skills.

In summary, I provide evidence that a large majority of individual investors seem to *passively* provide liquidity to the market (by leaving for instance their limit orders unattended in the limit order book), while a minority of them seem to have the required skills to *actively* engage in profitable liquidity-providing strategies (after the inclusion of transaction costs).

Next, I study whether some particular market frictions allow top traders in my sample to persistently profit from providing liquidity. Indeed, it may seem surprising that a subset of retail investors is able to capture arbitrage opportunities *before* other more sophisticated arbitrageurs in the market completely exploit them, such as hedge funds or institutional investors. One possibility is that the most successful investors in my sample provide liquidity to the market when institutional arbitrageurs are more constrained, and thus not able to exploit the arbitrage opportunities available.

I provide evidence in favor of this “limits-to-arbitrage” explanation ([Shleifer and Vishny \[1997\]](#)) in two different ways. First, for each round-trip in my sample I recover the level of intraday volatility of the stock traded during the day the round-trip was initiated.<sup>3</sup> Indeed, by conditioning on the starting date of a round-trip it is likely that I identify when investors *actively* seek to exploit a trading opportunity. I use the intraday volatility because it is a proxy for arbitrage risk: the higher the volatility, the higher the arbitrage risk ([Pontiff \[2006\]](#) and [Lam and Wei \[2011\]](#)) and the lower the liquidity ([Hameed, Kang, and Viswanathan \[2010\]](#) and [Nagel \[2012\]](#)).<sup>4</sup> I find that top-traders initiate their most profitable round-trips when arbitrage risk is extremely high, and earn on average 16 additional basis points on their (net) abnormal returns relative to the days where the volatility is low. Importantly, this positive relationship between intraday volatility and profitability does not hold for bottom traders, which confirms

---

<sup>3</sup>Therefore before this starting date the net inventory position on that stock for a given investor is exactly zero.

<sup>4</sup>See also [Mashruwala, Rajgopal, and Shevlin \[2006\]](#) for a discussion in the accounting literature.

that only top-traders are able to identify and exploit the moments when the market needs liquidity.

Second, I look at the time series variation of the abnormal returns earned, on average, by each decile group of investors. I find that over the last decade, the top traders in my sample earn significantly higher profits during the periods of market turmoil than during periods of ordinary market conditions. For instance, while the top-group earns average daily abnormal returns (net of fees) non-statistically different from zero in 2005, they gain on average 16 basis points in 2001 (dot-com bubble) and 35 basis points in 2008 (during the recent financial crisis). On the opposite, the bottom group experiences sharp losses during during the same periods. As the recent financial crisis was a period where arbitrageurs were extremely constrained and liquidity very scarce (Ben-David, Franzoni, and Moussawi [2011] and Nagel [2012]), my previous results are thus consistent with some retail investors acting as liquidity-providers of last resort.

This article proceeds as follows. [Section 2.2](#) presents my methodology and my brokerage dataset. Results are given and discussed in [Section 2.3](#). Before concluding, I study in [Section 2.4](#) whether limits-to-arbitrage allow some traders to profit from liquidity providing strategies.

## 2.2 Methodology

I describe in this section the data used in this paper, and my methodology for assessing the investors' trading performance.

### 2.2.1 The brokerage dataset

The data used in this paper comes from a leading French online broker.<sup>5</sup> The raw dataset contains at the daily level all the executed trades sent by the 145801 distinct broker's customers

---

<sup>5</sup>A subsample of this database covering the period 1999-2001 has already been used, to address another research question, in [Foucault, Sraer, and Thesmar \[2011\]](#).

from 1999 up to 2010, which represents more than 15 million trades.<sup>6</sup>

Each trade comes with the following information: the asset type (equity, bonds, etc), the trading exchange identifier (the ISIN), the trading date, the quantity, the order type (limit, market and other minor orders types), the amount traded in euro, the trading place, the transaction fees incurred (from 2006 onward) and the trading channel used to submit the order. I match the trades in my dataset with market data provided by Eurofidai, the European financial data institute. Trades are matched by ISIN code, trading day and trading exchange code. Trades for which no information is available from Eurofidai are discarded from the sample.

Summary statistics of the raw brokerage dataset are provided in Table 2.1. This table shows the corresponding number of trades, percentage and cumulative percentage for several categorical variables that describe the nature of my data well. Panel A shows for instance that trades are in the majority of cases limit orders (62%) and market orders (28%). The other minor orders types are only used marginally by investors. The information on order type is completely missing for 1999 and 2000 and may be marginally missing for up to 2004. Therefore these figures do not take into account the missing values.

Investors could submit an order by using the telephone to speak to a broker official (Telephone), by calling a voicemail service and typing instruction using the telephone's keys (Phone+), by using a web navigator to connect to the broker website (Web), by using an old French Videotex online service accessible through the telephone lines (Minitel) or by using online basic computer software (Online+ and Trader+). Panel B of the table highlights that a large majority of trades are submitted using the internet or trading software (Online+ and Trader+).

Panel C and Panel D of the table show that most of individual trading activity consists of buying or selling common stocks on the NYSE Euronext Paris trading exchange. This dataset, therefore, shows patterns of individual investment behavior that are similar to other recent databases used in the literature. For instance, Finnish investors in Linnainmaa [2010] also use limit orders for most of their trades on the Helsinki Stock Exchange.

---

<sup>6</sup>This section is taken from the first chapter.

<Insert Table 2.1 here>

## 2.2.2 Investors' round-trips

To understand whether retail investors are able to profit from liquidity providing strategies, I need to accurately assess their trading performance. This performance evaluation is in itself a difficult methodological issue.

Indeed, investors may trade for various reasons and may have very heterogeneous trading horizons that are not captured by the standard fixed-horizon performance measures used in the literature. I overcome this methodological challenge in this paper by following [Schlarbaum, Lewellen, and Lease \[1978\]](#): I aggregate investors' single trades into round-trip trades that fully reflect the actual cash-flows received or paid.

To do so, I first compute the daily net quantity traded by investor  $i$ , on stock  $s$  and day  $t$  as:

$$Q_{i,t,s} = Qtit_{i,t,s}^{buy} - Qtit_{i,t,s}^{sell},$$

where  $Qtit_{i,t,s}^{buy}$  (resp.  $Qtit_{i,t,s}^{sell}$ ) represents the actual quantity of stock  $s$  bought (resp. sold) by investor  $i$  on day  $t$ .

Then, I sort all the daily net quantities in my dataset by trader, stock and trading day and I keep track of the cumulative stock quantity held by a trader day after day. A round-trip starts and ends with a zero net cumulative quantity. Round-trips are thus trading positions that are fully unwound: stocks previously bought are entirely sold and stocks previously sold short are completely bought back. I provide in [Table 2.2](#) a detailed example of this procedure.

<Insert Table 2.2 here>

The main advantage of computing round-trips over a traditional fixed-horizon approach is clearly stated in [Schlarbaum, Lewellen, and Lease \[1978\]](#):

*The dates, prices, and commissions recorded in the transactions file are hard data; they represent actual investment activities. As such, they eliminate any need to choose an arbitrary periodic-return measurement interval, to adjust for portfolio changes during that interval, or to allocate transactions costs across intervals by some amortization rule. A realized-return, investment-round-trip orientation, therefore, is persuasively clean*

This methodology is still rare in the literature because it requires brokerage data that have a panel structure at the individual level, which is in general not possible in the US (see for instance [Kaniel, Saar, and Titman \[2008\]](#) or [Kelley and Tetlock \[2013\]](#)). My dataset thus provides an opportunity to enrich the inference previously drawn on investor trading behavior and performance.

Once a round-trip is identified, I compute several measures that will serve to assess the profitability of investors' trades. The duration of a round-trip is the number of days between the start and the end of a round-trip. I thus can obtain a proxy for investor i's trading horizon, in a given month, by taking the average duration of all the investor i's round-trips *started* that month.

The gross profit generated by a round trip is the (signed) sum of all the cash-flows arising within a given round-trip. A buy order represents a negative cash-flow corresponding to the amount paid, while a sell orders generate a positive cash-flow corresponding to the amount received from the sell. I thus compute:

$$\text{Gross Profit}_{i,s}(t, T) = \text{Sell}_{i,s}(t, T) - \text{Buy}_{i,s}(t, T),$$

where  $\text{Gross Profit}_{i,s}(t, T)$  stands for the gross profit of a round-trip initiated on day t, closed on day T, on stock s, by investor i.  $\text{Sell}_{i,j}(t, T)$  corresponds to the sum of all the sell orders' euro amount executed within that round-trip while  $\text{Buy}_{i,j}(t, T)$  corresponds to the sum of all the buy

orders' euro amount executed within the same round-trip. I compute net profits by subtracting the amount of transactions fees paid by the traders to total buy and sell cash-flows.

$$Sell_{i,s}^{net}(t, T) = Sell_{i,s}(t, T) - Fee_{i,s}^{sell}(t, T)$$

$$Buy_{i,s}^{net}(t, T) = Buy_{i,s}(t, T) + Fee_{i,s}^{Buy}(t, T)$$

Similarly, I compute the (gross and net) returns for individual's  $i$  round-trip on stock  $j$ , started on day  $t$  and closed on day  $T$  as:

$$R_{i,j}^{gross}(t, T) = \frac{Sell_{i,j}(t, T) - Buy_{i,j}(t, T)}{Buy_{i,j}(t, T)},$$

$$R_{i,j}^{net}(t, T) = \frac{Sell_{i,j}^{net}(t, T) - Buy_{i,j}^{net}(t, T)}{Buy_{i,j}^{net}(t, T)},$$

Importantly, following [Shapira and Venezia \[2001\]](#) and [Chakrabarty, Moulton, and Trzcinka \[2013\]](#), I keep track of the starting and ending date of each round-trip to compute risk-adjusted returns. To do so, I compare the (gross or net) return earned on a given round-trip to the return that would have been *passively* achieved over the *same* holding period by investing in the market index. :

$$R_{i,j}^{abnormal}(t, T) = R_{i,j}(t, T) - \frac{Index(T) - Index(t)}{Index(t)},$$

$$R_{i,j}^{abnormal-net}(t, T) = R_{i,j}^{net}(t, T) - \frac{Index(T) - Index(t)}{Index(t)},$$

where  $Index(t)$  stands for the general French market index computed by Eurofidai, or the French CAC40 index value at the end of day  $t$ . Adjusting returns using French indexes as market benchmarks is appropriate in my setting because the large majority of trades in the database are executed on French stocks.

Last, as different round-trips may have different durations, I follow [Shapira and Venezia](#)

[2001], and I convert each (gross or net) abnormal return into an equivalent *daily-return* using the formula:<sup>7</sup>

$$R_{i,j}^{daily}(t, T) = (1 + R_{i,j}^{abnormal-net}(t, T))^{(1/T)} - 1$$

## 2.3 Results

### 2.3.1 Ranking investors by trading performance

I describe in this section how I classify investors in my sample according to their trading performance. As in [Coval, Hirshleifer, and Shumway \[2005\]](#), I rank investors into deciles of trading performance.

Of course, assessing the performance of a trader over its whole career is per se a difficult question. I take nonetheless a straightforward approach: I compute the gross (net) Sharpe ratio of each trader, obtained as his average gross (net) round-trip profit divided by the standard deviation his gross (net) profits. My approach is thus similar to [Barber et al. \[2013\]](#) who state that they “*include dollar profits [in the computation of the investors’ Sharpe ratios] to capture investors who might consistently earn low returns on a large dollar value of trades.*”

I then classify investors into deciles of Sharpe ratios. By doing this I restrict my sample to investors having at least 25 round-trips over their whole career in order to have a precise measurement of their performance. As I classify investors only according to Sharpe ratios, then the top deciles are simply the best performers among the traders in my sample.

I provide in [Table 2.3](#) the main summary statistics on the individual Sharpe ratios, for investors in a given decile group.

---

<sup>7</sup>Instead, annualizing return could be problematic as it could lead to extremely high figures.

<Insert Table 2.3 here>

Next, for each measure of trading performance, I compare each decile group (from the second to the tenth) to the worst performing group (decile 1) by running the following regression

$$Y_{i,s,t} = \alpha + \sum_{d=2}^{10} Decile_{d,i} + \epsilon_{s,t},$$

where the round-trip performance measure  $Y$  (one of those defined in the previous section) is stacked by decile  $i$ , stock  $s$  and starting date  $t$ .  $Decile_i$  is a dummy variable that equals one only if the return is earned by an investor belonging to decile  $i$ .  $\epsilon_{s,t}$  is a error term that is doubly clustered by stock  $s$  and starting date  $t$ , as recommended in [Thompson \[2011\]](#).

Results are presented in [Table 2.4](#). The different columns corresponds to different measures of performance: gross and net profits in euro (column 1 and 2) and gross and net abnormal returns when the benchmark is either the Eurofidai French market index (columns 3 and 4) or the CAC40 index (columns 5 and 6).

Each column contains two subcolumns (labeled “ALL” and “LO”). The label “ALL” simply corresponds to the whole sample of round-trips available in my sample, whereas the label “LO” stands for a restricted subsample of round-trips that *only contain* limit order trades. This separation allows me to test whether there is a difference in profitability across groups when investors execute round-trip strategies that provide liquidity to the market.

<Insert Table 2.4 here>

It emerges from [Table 2.4](#) that only traders that are ranked in the top three deciles are able to earn statistically positive profits (and risk-adjusted returns) from trading and in particular from liquidity-providing strategies. For instance, the bottom group loses on average EUR 125 per round-trip (a loss of EUR 140 when I restrict my sample to round-trips that are limit-orders only) and their performance mechanically worsens when transactions costs are included (net

losses of approximately EUR 215). In contrast, the top group earn an additional EUR 550 profit relative to the bottom group, leading to a total average net profit of about EUR 270 (EUR 450-180). The pattern is similar when I consider net abnormal returns. For instance, consider column 4 where the market benchmark is the Eurofidai French index. It appears that the bottom group loses 33 basis points *per day* (a loss of 49 basis points for limit-orders only round-trips) whereas the top group is able to earn a net daily abnormal profit of about 10 basis points (that is 60 basis points more than the bottom group).

This 60 basis points spread between top and bottom group in my sample is economically large, and in line with similar results in other countries. For instance [Barber et al. \[2013\]](#) find that top-trades in Taiwan outperform bottom-traders by 73 basis points per day, while [Grinblatt, Keloharju, and Linnainmaa \[2011\]](#) find that the buys of high-IQ Finnish investors outperform those with low-IQ by 4.4 basis points.

### 2.3.2 Differences in trading behavior across investors

To understand what may drive the large heterogeneity in trading performance that I have identified in the previous section, I study in this section the observable characteristics of the trading strategies executed by each decile group.

[Table 2.5](#) gives the proportion of limit-orders, the average amount, the average level of leverage and the average fees paid by each decile (from the 2nd to the 10th) relative to the bottom group.

Interestingly, there is an almost monotonic relationship between investor's performance and the use of limit orders. All investors in my sample use limit orders for most of their trades (as with Finnish investors in [Linnainmaa \[2011\]](#)), but top-traders use them much more than the bottom ones: the difference between those two groups is more than 14 percentage points.

Next, while there is not much variability in the average amount traded across the decile groups (bottom traders' trades sizes are approximately EUR 3600), one can see that top-traders

use leverage significantly less and pay smaller fees (20 basis points less) than bottom traders. The fact that the spread in average fees paid between investors is very large confirms that it is critical to use real transaction costs incurred by retail investors, instead of estimating them. Furthermore, these results suggest that top-traders try to minimize the transaction costs of their strategies by avoiding excessive (costly) leverage. Indeed leverage can be done in France through the French Deferred Settlement Service, which is a paid service provided by brokerage houses to their customers.

<Insert [Table 2.5](#) here>

To further understand what kind of strategies these top traders use, I examine now their propensity to be contrarians. There is indeed evidence that contrarian strategies are a form of liquidity-providing strategies because *“buying and selling in reversal strategies resembles the trading of a market maker who sells when the public buys (which tends to coincide with rising prices) and who buys when the public sells (which tends to coincide with falling prices)”* (see [Nagel \[2012\]](#)).

To do so, I pair each daily individual quantity imbalance  $Q_{i,t,s}$  with the corresponding intraday return on stock  $s$  on the same day, defined as

$$R_{s,t}^{intraday} = \frac{P_{s,t}^{closing} - P_{s,t}^{opening}}{P_{s,t}^{opening}},$$

where  $P_{s,t}^{opening}$  and  $P_{s,t}^{closing}$  are the opening and closing price of security  $s$  on day  $t$ .

This methodology, which is adapted from [Foucault, Sraer, and Thesmar \[2011\]](#), allows me to classify each individual day-stock imbalance as a momentum, contrarian or round-trip trade. An individual day-stock imbalance is classified as contrarian trade if:

$$Q_{i,t,s} * R_{s,t}^{intraday} < 0.$$

Conversely an imbalance is classified as a momentum trade if:

$$Q_{i,t,s} * R_{s,t}^{intraday} > 0.$$

Trades for which  $Q_{i,t,s} * R_{s,t}^{intraday} = 0$  are defined as round-trip trades.

I present in [Table 2.6](#) estimations from the three following regressions

$$Type_{i,s,t} = \alpha + \sum_{d=2}^{10} Decile_{d,i} + \epsilon_i,$$

where  $Type_{i,s,t}$  is one of the possible three dummy variables  $Contr_{i,s,t}$ ,  $Mom_{i,s,t}$  or  $Round_{i,s,t}$ ,  $Decile_{d,i}$  is a dummy variable that equals one only if individual  $i$  belongs to decile  $d$  and  $\epsilon_i$  is a error term that is clustered at the individual level. Again, the bottom decile constitutes the omitted category.

<Insert [Table 2.6](#) here>

[Table 2.6](#) shows two patterns. First, it confirms that retail investors are, on average, contrarian (see for instance [Grinblatt and Keloharju \[2000\]](#) or [Kaniel, Saar, and Titman \[2008\]](#)): 58% of the individual day-stock imbalances in the bottom group are marked as contrarian. Second, as for limit orders, performance seems to increase with the contrarian tendency of individuals. But more importantly top performers in the highest decile (decile 10) are significantly even more contrarian than individuals in decile 9. In other words, the top-traders in my sample that earn the largest net abnormal returns are those who follow contrarian strategies the most, that is more than any other retail speculator in the market.

This result may seem surprising given that the aggregate contrarian tendency of individual is often described as a form of irrationality. However, [Grinblatt, Keloharju, and Linnainmaa \[2011\]](#) also document that high IQ investors in their sample appear to be more contrarian than low IQ investors but can only speculate whether this behavior is rational or indicative of a greater

behavioral bias. They suggest that “by selling stocks at monthly highs and holding stocks at monthly lows, high-IQ investors *are more likely* to be following a rational liquidity provision strategy than a psychological bias that diminishes returns”. I provide a graphical illustration in favor of this explanation: I show in [Figure 2.1](#) the daily imbalance of the top-group on ALCATEL-LUCENT in 2010. When I overlay the stock price, it appears indeed that the top-group buys at lows and sell at highs.

<Insert [Figure 2.1](#) here>

This explanation is also confirmed by [Table 2.7](#), where I show the main summary statistics on the durations of the round-trips implemented on average by each decile group. Specifically, one can see that the median duration of the round-trips for the top-group is about a month, which is exactly in line with the usual horizons of the contrarian strategies identified in the literature ([Jegadeesh \[1990\]](#), [Lehmann \[1990\]](#) and [Avramov, Chordia, and Goyal \[2006\]](#)).

<Insert [Table 2.7](#) here>

### 2.3.3 Is there persistence in trading performance?

At this point, and given the large sample size of individuals in my database, one can argue that even if all the traders in my sample are unskilled, some may appear to overperform by just being lucky enough to trade at the right moment. Is the performance of the top-group due to skill or mere luck? In this section, I test for persistence in trading performance using two standard non-parametric tests borrowed from the mutual fund performance literature.

First, I use a direct methodology from [Coval, Hirshleifer, and Shumway \[2005\]](#) and [Seru et al. \[2009\]](#). I divide my sample in two disjoint time periods (2000-2005 and 2005-2010) and I compute the Spearman rank correlation between the individual Sharpe ratios from the first period to those of the second period. Using the Spearman rank correlation coefficient is

appropriate here because this measure of association is very robust to potential outliers in the data. I obtain a highly significant correlation coefficient of 0.20 (t-stat of 22) that is in line with prior research on retail performance persistence. For instance, [Seru et al. \[2009\]](#) report a coefficient of 0.164 and consider this result “*quite statistically and economically significant*”.

Second, do winners (losers) in one period tend to be winners (losers) in the next period? To specifically answer this question I adapt the methodology described in [Brown and Goetzmann \[1995\]](#), because it makes makes the assessment of performance persistence very clear and intuitive. Their approach is based on contingency tables. [Table 2.8](#) displays from 2003 to 2010 what [Brown and Goetzmann \[1995\]](#) call “repeat winners” and “repeat losers” among traders. I do the following: each year I compute the Sharpe ratio (of the net profits) for each investor in my database. I restrict the sample to investors who submitted at least 10 round trips in a single year in order to obtain statistically valid (yearly) Sharpe ratio estimates.

An investor is classified as a winner in given year if his net Sharpe ratio during the same year is positive and is classified as a loser if his net Sharpe ratio is negative. The table is obtained as follows: rows indicates winning or losing investors in the *current* year, and columns indicates, conditional of their status in the current year, the number of investors that are either winners or losers the year after. For instance, there are 1180 losers and 908 winners out of 2036 (ranked) investors in 2004. I also differentiate between investors who decide to close their account in the next year (columns "Winner Quit" and "Loser Quit") and those who still have an open trading account (columns "Winner Stay" and "Loser Stay"). These investors are thus still “alive” in the next year, irrespective of their winning or losing status.

[Table 2.8](#) shows that, year after year, there is a remarkable persistence in both winning and losing status. The effect, however, seems stronger for losers. For example, compared to a losing investor in 2005, an investor that is winning in 2005 has twice the probability of being a winner in 2006 (23% vs. 51%). Similarly a losing investor in 2009 has a 84% probability of losing again in 2010, while the probability of losing in 2010 for a winning investor in 2009 is 51%).

<Insert [Table 2.8](#) here>

In summary there is evidence that luck cannot explain all my results, and top-traders in my sample do seem to possess genuine trading skills.

## 2.4 Limits-to-arbitrage and liquidity provision

I have shown in the previous section that the subset of retail investors in my sample, who seem to profit from liquidity provision to the market, are those who use limit orders and contrarian strategies more than any other investor. I study in this section whether this performance can be explained by some particular market frictions, because it is not clear why some retail investors may be able to capture arbitrage opportunities *before* other more sophisticated arbitrageurs in the market. Indeed, these opportunities should disappear after institutional arbitrageurs start to exploit them.

One possibility is that top traders in my sample may provide liquidity to the market when institutional arbitrageurs are more constrained, and thus not able to exploit the arbitrage opportunities available. For instance, [Foucault, Sraer, and Thesmar \[2011\]](#) discuss at the end of their paper the possible effect of retail contrarian trades on stock volatility. They argue that contrarian retail trades seem to enhance volatility but add:

*but reverse causality remains a possibility as sophisticated retail investors may enter the market and act as liquidity providers only in periods of high volatility (maybe because high volatility signals that the market lacks liquidity).*

[Figure 2.2](#) provides a visual illustration in favor of this explanation. To obtain the graph, I first compute a proxy for the intraday volatility of a stock on day  $t$  as:

$$\sigma_{s,t} = \frac{P_{s,t}^{max} - P_{s,t}^{min}}{P_{s,t}^{min} + P_{s,t}^{max}},$$

where  $P_{s,t}^{min}$  and  $P_{s,t}^{max}$  are the minimum and maximum Euronext traded price for stock  $s$  on day  $t$ . This is thus a common range-proxy for a stock intraday volatility.

Then, I plot in [Figure 2.2](#) this volatility against the daily imbalance of a trader randomly selected in the top decile group. Consistent with a liquidity-provision strategy and the arguments above, the individual seems to trade when the intraday volatility spikes.

<Insert [Figure 2.2](#) here>

I test more formally this possibility in two different ways. First, for each round-trip in my sample I recover the level of intraday volatility of the stock traded during the day the round-trip was initiated.<sup>8</sup> Indeed, by conditioning on the starting date of a round-trip is it likely that I identify when investors *actively* seek to exploit a trading opportunity. Using the intraday volatility an explanatory variable seems natural for this test because it is an illiquidity measure that captures arbitrage risk ([Pontiff \[2006\]](#), [Lam and Wei \[2011\]](#) and [Hameed, Kang, and Viswanathan \[2010\]](#)).

Specifically, I run the following regression separately for each decile of investors:

$$R_{s,t,k} = \alpha_s + \sum_{i=2}^{10} DecileVol_{s,t} + Year_t + \epsilon_{s,t},$$

Where  $R_{s,t,k}$  is the abnormal return (net of fees) computed on the  $k$ -th round-trip started on day  $t$  and stock  $s$  by an investor in a given decile.  $\alpha_s$  is a stock fixed effect,  $Year_t$  is a year fixed-effect and  $DecileVol_{s,t}$  is a dummy variable that classifies the round-trips into deciles of intraday volatility according to the volatility level of stock  $s$  on day  $t$ .<sup>9</sup>

I find that top-traders initiate their most profitable round-trips when arbitrage risk is extremely high (decile 10), and earn on average 16 additional basis points on their (net) abnormal returns relative to the days where the volatility is low. Interestingly this pattern is reversed

---

<sup>8</sup>In other words, before this starting date a given investor has a zero net inventory position on that stock.

<sup>9</sup>Recall that  $t$  stands for the starting date of the round-trip.  $\epsilon_{s,t}$  is thus an error term that is doubly clustered by stock  $s$  and starting date  $t$ .

for bottom traders (they lose about 10 basis points more when the volatility is high), which confirms the previous evidence that only top-traders are able to identify and *actively* exploit the profitable opportunities to provide liquidity, while the majority of retail investors provide liquidity at their own expenses.

<Insert [Table 2.9](#) here>

As a second piece of evidence, I look at the *time series variation* of the abnormal returns earned, on average, by each decile group of investors. To do so, I regress the abnormal returns earned by each decile group of investors on a set of year-dummy variables (from 2000 to 2010) and I double-cluster the standard errors at the day-stock level. Results are shown in [Table 2.10](#). To better understand the dynamics of those returns over time, I also plot in [Figure 2.3](#) the point estimates from the regression (i.e. the yearly mean abnormal returns) and the corresponding 95% confidence interval bands each year for three groups: the bottom-group, the 5th decile group and the top group.

<Insert [Table 2.10](#) here>

<Insert [Figure 2.3](#) here>

I find that over the last decade, the returns earned by the top-group spike significantly more during the periods of market turmoil than during periods of ordinary market conditions. For instance, while the top-traders earn average abnormal returns non-statistically different from zero in 2005, they gain on average 16 basis points in 2001 (dot-com bubble) and 35 basis points in 2008 (during the recent financial crisis). As in the previous set of regression, volatility does not translate into additional gains for retail investors in the bottom group. For instance, bottom-traders lose on average 58 basis points in 2006 and these losses further drop to an average loss of 66 basis points in 2008.

As the recent financial crisis was a period where liquidity was scarce and arbitrageurs were extremely constrained ([Ben-David, Franzoni, and Moussawi \[2011\]](#) and [Nagel \[2012\]](#)), my

previous results are thus consistent with some retail investors acting as liquidity-providers of last resort.

<Insert [Table 2.10](#) here>

## **2.5 Conclusion**

I identify a subset of retail investors that is able to beat the market by using liquidity-providing strategies. I find that their performance, net of fees, is large and significant. Interestingly, these traders seem to profit from arbitrage opportunities left out in the market by other classical arbitrageurs who are temporarily more constrained.

In conclusion, this paper sheds a positive light on a subset of remarkably skilled individuals and contributes to our understanding of the heterogeneity in investors' performance.

## 2.5 References

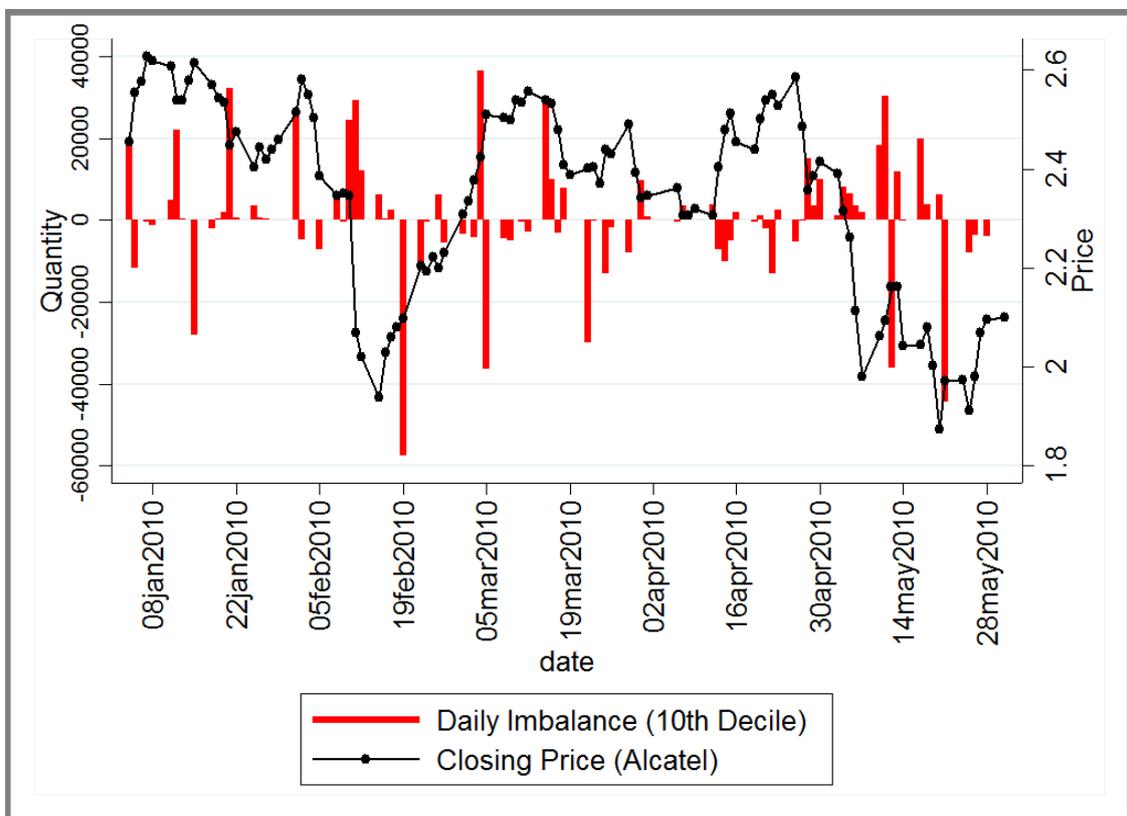
- Avramov, D., Chordia, T., and Goyal, A. (2006). "Liquidity and Autocorrelations in Individual". *Journal of Finance* LXI.5.
- Barber, B. M., Lee, Y.-T., Liu, Y.-J., and Odean, T. (July 2013). "The cross-section of speculator skill: Evidence from day trading". *Journal of Financial Markets*.
- Barrot, J., Sraer, D., and Kaniel, R. (June 2014). "Are retail traders compensated for providing liquidity?"
- Ben-David, I., Franzoni, F., and Moussawi, R. (Nov. 2011). "Hedge Fund Stock Trading in the Financial Crisis of 2007-2009". *Review of Financial Studies* 25.1.
- Brown, S. and Goetzmann, W. (1995). "Performance persistence". *The Journal of Finance*.
- Campbell, J., Grossman, S., and Wang, J. (1993). "Trading volume and serial correlation in stock returns". *The Quarterly Journal of Economics* 108.November.
- Chakrabarty, B., Moulton, P., and Trzcinka, C. (2013). "Institutional Holding Periods". *Working Paper*.
- Coval, J., Hirshleifer, D., and Shumway, T. (2005). "Can Individual Investors Beat the Market?" *HBS Finance Working Paper No. 04-025*.
- Foucault, T., Sraer, D., and Thesmar, D. J. (2011). "Individual Investors and Volatility". *Journal of Finance* LXVI.4.
- Grinblatt, M., Keloharju, M., and Linnainmaa, J. (2011). "IQ, trading behavior, and performance". *Journal of Financial Economics*.
- Grinblatt, M. and Keloharju, M. (2000). "The investment behavior and performance of various investor types : a study of Finland ' s unique data set". *Journal of Financial Economics* 55.1.
- Grossman, S. and Miller, M. (1988). "Liquidity and market structure". *the Journal of Finance* XLIII.3.
- Hameed, A., Kang, W., and Viswanathan, S. (2010). "Stock market declines and liquidity". *The Journal of Finance* LXV.1.

- Jegadeesh, N. (1990). "Evidence of predictable behavior of security returns". *Journal of Finance* 45.3.
- Kaniel, R., Saar, G., and Titman, S. (2008). "Individual investor trading and stock returns". *The Journal of Finance* LXIII.1.
- Kelley, K. and Tetlock, P. (2013). "How wise are crowds? Insights from retail orders and stock returns". *The Journal of Finance* January.
- Lam, F. E. C. and Wei, K. J. (Oct. 2011). "Limits-to-arbitrage, investment frictions, and the asset growth anomaly". *Journal of Financial Economics* 102.1.
- Lehmann, B. N. (1990). "Fads, Martingales, and Market Efficiency". *Quarterly Journal of Economics* 105.1.
- Linnainmaa, J. (2010). "Do limit orders alter inferences about investor performance and behavior?" *The Journal of Finance* LXV.4.
- (2011). "Why Do (Some) Households Trade So Much?" *Review of Financial Studies* 24.5.
- Mashruwala, C., Rajgopal, S., and Shevlin, T. (Oct. 2006). "Why is the accrual anomaly not arbitrated away? The role of idiosyncratic risk and transaction costs". *Journal of Accounting and Economics* 42.1-2.
- Nagel, S. (June 2012). "Evaporating Liquidity". *Review of Financial Studies* 25.7.
- Pontiff, J. (Oct. 2006). "Costly arbitrage and the myth of idiosyncratic risk". *Journal of Accounting and Economics* 42.1-2.
- Schlarbaum, G., Lewellen, W., and Lease, R. (1978). "Realized returns on common stock investments: The experience of individual investors". *Journal of Business* 51.2.
- Seru, A., Shumway, T., Stoffman, N., and Ross, S. M. (Sept. 2009). "Learning by Trading". *Review of Financial Studies* 23.2.
- Shapira, Z. and Venezia, I. (2001). "Patterns of behavior of professionally managed and independent investors". *Journal of Banking & Finance* 25.
- Shleifer, A. and Vishny, R. (1997). "The limits of arbitrage". *The Journal of Finance* LII.1.
- Thompson, S. B. (Jan. 2011). "Simple formulas for standard errors that cluster by both firm and time". *Journal of Financial Economics* 99.1.

## 2.6 Figures

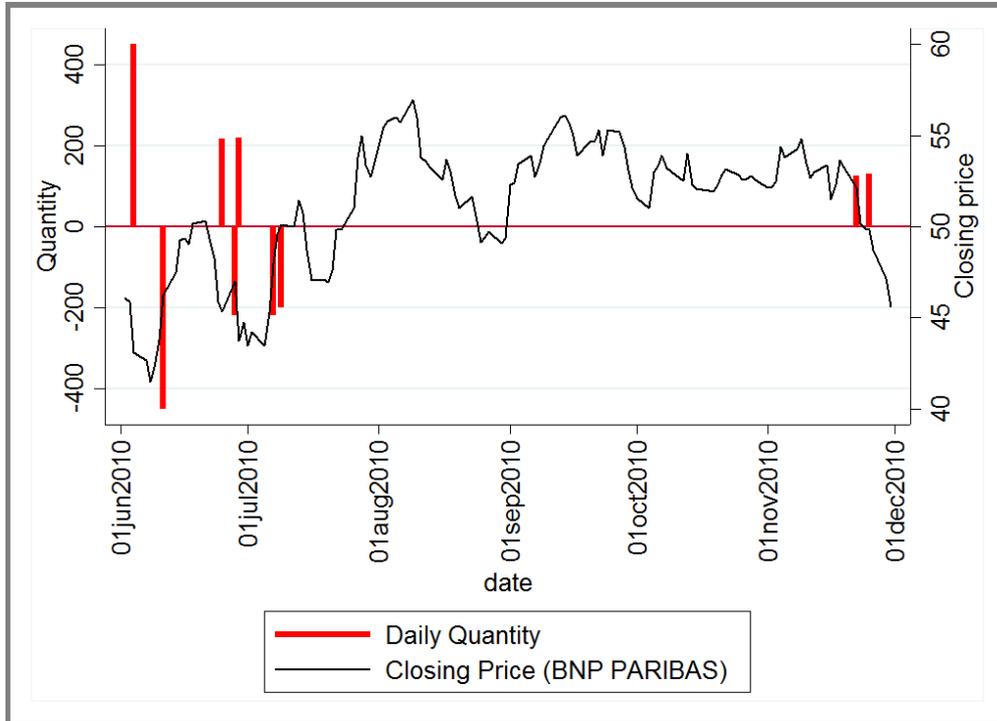
### Figure 2.1: Aggregate Net Imbalance: Top Decile

This Figure shows the daily imbalance (quantity bought on day  $t$  - quantity sold on day  $t$ ) on Alcatel-Lucent stocks in 2010 for the top-group in my sample. To do so, the trades from each trader in the top group are signed (positive for a buy and negative for a sell) and added together by trading day. I also plot on the same graph the daily closing price of the stock.

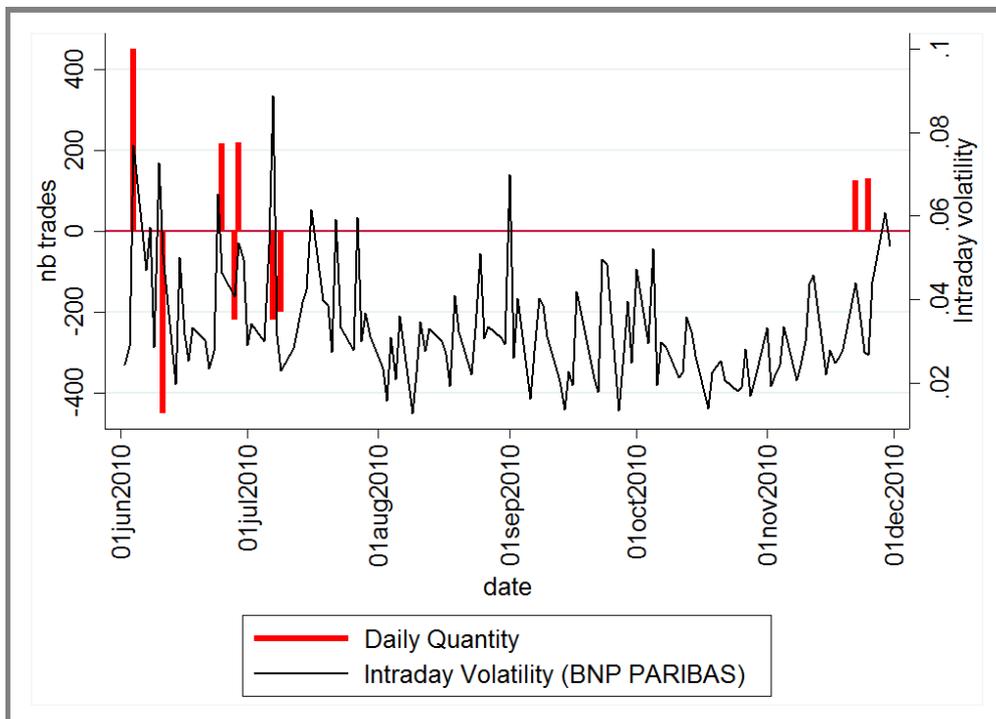


## Figure 2.2: A tiny market maker

This figure shows the daily imbalance (quantity bought on day  $t$  - quantity sold on day  $t$ ) for an investor randomly selected from the top decile group. The traded security is BNP Paribas in 2010. I plot the net individual daily imbalance (quantity bought on day  $t$  - quantity sold on day  $t$ ) against the closing price of that stock (top Figure) and against the intraday volatility of that stock (bottom Figure).



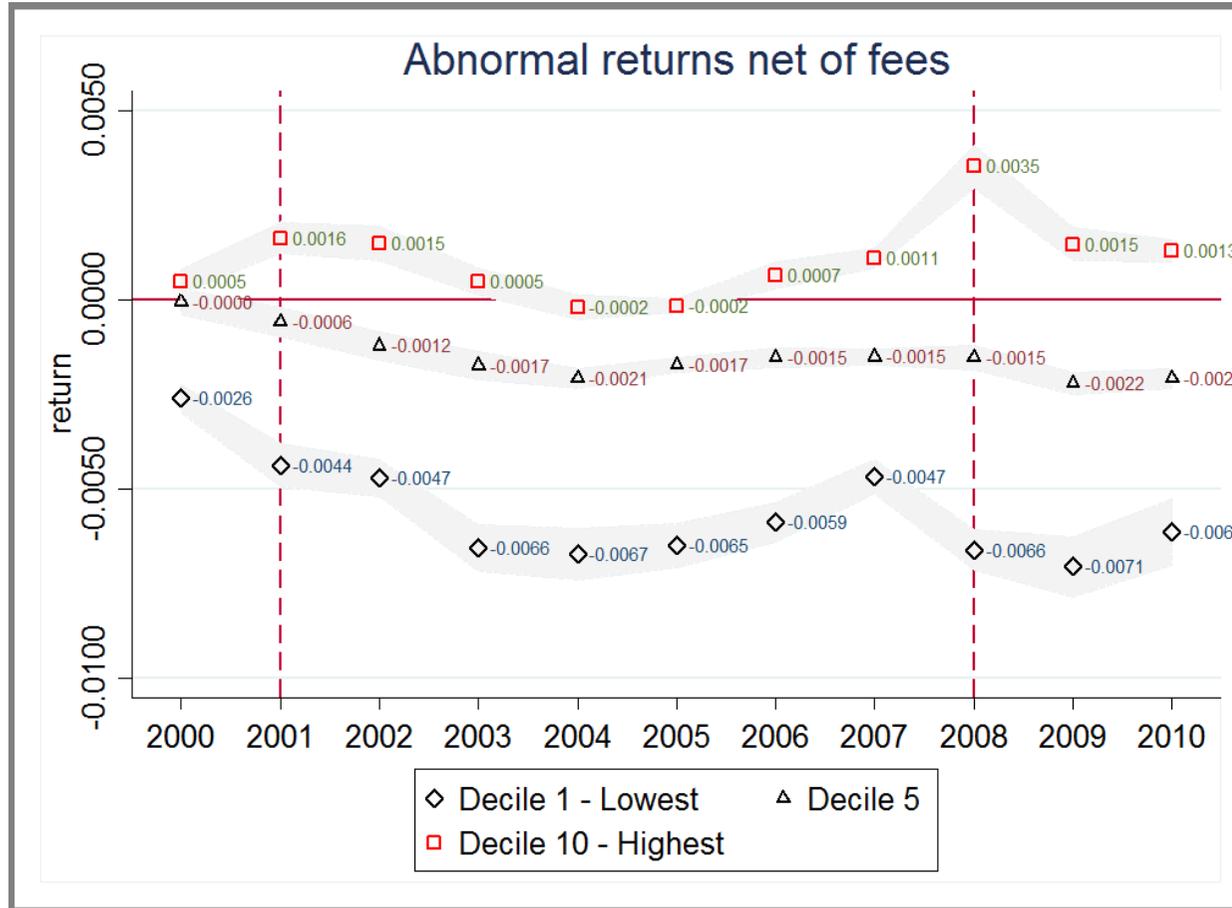
(a) Daily Order Flow vs Closing Price



(b) Daily Order Flow vs Intraday Volatility

Figure 2.3: Abnormal returns dynamics over the last decade

This figures shows points estimates and 95% confidence intervals bands (in gray) for the daily abnormal returns *net of fees* earned on average, each year, by the bottom, the middle and the top-group of investors in my sample. Standard errors are clustered at the stock-day level.



## 2.7 Tables

## Table 2.1: Summary statistics of the brokerage dataset

This table describes the main characteristics of the brokerage dataset used in this paper. The table gives the corresponding number of trades, percentage and cumulative percentage for each category of information (investors' order choices, investors' use of trading channels, investors' use of asset classes and investors' use of trading exchanges). The total number of trades may be different from one category to another due to missing values. The sample period is from 1999 to 2010.

Panel A: Order Choice			
	Frequency	Percentage	Cum. Percentage
Limit Order	6434646	62.6	62.6
Market Order	2918369	28.4	91.0
Others/Missing	920520	8.96	100
Total	10273535	100	

Panel B: Trading Channels			
	Frequency	Percentage	Cum. Percentage
Web	5785047	42.1	42.1
Online+	4103445	29.8	71.9
Trader+	2049605	14.9	86.8
Minitel	782961	5.69	92.5
Telephone	686728	4.99	97.5
Phone+	181141	1.32	98.8
Manual	165430	1.20	100
Total	13754357	100	

Panel C: Asset Classes			
	Frequency	Percentage	Cum. Percentage
Common Stocks	12986031	86.2	86.2
Others	824604	5.47	91.6
Equity Warrants	488490	3.24	94.9
Index Warrants	365357	2.42	97.3
Turbo Warrants	270680	1.80	99.1
ETF	134270	0.89	100.0
Total	15069432	100.0	

Panel D: Trading Exchanges			
	Frequency	Percentage	Cum. Percentage
NYSE Euronext Paris	13888529	92.16	92.16
Tradegate	669061	4.44	96.60
Others	437689	2.90	99.51
Nasdaq	74153	0.49	100.00
Total	15069432	100.00	

Nb. of distinct investors	145801
Nb. of buy order	7910252
Nb. of sell orders	7159180

## Table 2.2: Identification of Trading Strategies

I provide here an illustration of the round-trip identification used in this paper. In the example below, the investor completes two round-trips on the same security ABC. The first one started on January 1 and ended in January 18. The second one consists of a buy of 5 ABC on February 20 followed by a sell of 5 ABC the next day. Quantity stands for the quantity of stocks purchased or sold, Cash-Flow stand for the amount of money (in euro) actually received (in case of a sell) or paid (in case of a buy) by the investor. Fees is the daily amount of commissions fees incurred by the trader. CumQ keeps track of the inventory level on security ABC. As I identify two round-trips, the inventory level (or cumulative position held by the trader) revert accordingly to 0 two times.

Id	Stock	Date	Order	Quantity	Cash Flow	Fees	Daily Delta	CumQ
1	ABC	01/01/2007	BUY	10	-500	4,5	10	10
1	ABC	15/01/2007	BUY	11	-510	4,5	11	21
1	ABC	16/01/2007	SELL	10	550	4,5	-10	11
1	ABC	18/01/2007	SELL	11	560	4,5	-11	0
1	ABC	20/01/2007	BUY	5	-200	5	5	5
1	ABC	21/02/2007	SELL	5	170	4,5	-5	0

## Table 2.3: Sharpe ratios summary statistics

This table shows the mean and the main quantiles of the individual Sharpe ratios (within each decile group).

Sharpe ratios by deciles					
	Nb. Investors	Mean	p25	p50	p75
Decile 1 (Lowest)	1846	-1.12	-0.79	-0.62	-0.53
Decile 2	1845	-0.40	-0.43	-0.40	-0.37
Decile 3	1846	-0.30	-0.32	-0.30	-0.28
Decile 4	1845	-0.23	-0.25	-0.23	-0.22
Decile 5	1846	-0.18	-0.19	-0.18	-0.16
Decile 6	1845	-0.12	-0.14	-0.12	-0.11
Decile 7	1846	-0.06	-0.08	-0.06	-0.04
Decile 8	1845	0.03	0.00	0.02	0.05
Decile 9	1846	0.17	0.12	0.16	0.22
Decile 10 (Highest)	1845	0.83	0.39	0.53	0.77
Total	18455	-0.14	-0.30	-0.15	0.02

## Table 2.4: Trading performance and liquidity provision

This table shows the results of the following regression:

$$Y_{i,s,t} = \alpha + \sum_{d=2}^{10} Decile_{d,i} + \epsilon_{s,t},$$

where the round-trip performance measure Y (Gross and Net profits, Gross and Net abnormal returns) is stacked by decile i, stock s and starting date t.  $Decile_i$  is a dummy variable that equals one only if the return is earned by an investor belonging to decile i.  $\epsilon_{s,t}$  is a error term that is doubly clustered by stock s and starting date t (Thompson [2011]). The label ALL stands for all the round-trips in my sample, whereas LO stands for the round-trips that only include limit orders. \*, \*\*, \*\*\* denote significance at the 10%, 5% and 1% levels.

	Gross Profit (EUR)		Net Profit (EUR)		Gross Ab. return (INDEX)		Net Ab. Return (INDEX)		Gross Ab. return (CAC)		Net Ab. Return (CAC)	
	ALL	LO	ALL	LO	ALL	LO	ALL	LO	ALL	LO	ALL	LO
Decile 2	5.293 (0.88)	47.99*** (6.50)	35.71*** (5.49)	42.24*** (5.49)	0.000126* (1.90)	0.00178*** (14.66)	0.000599*** (9.49)	0.00207*** (14.00)	0.000151** (2.27)	0.00165*** (13.85)	0.000570*** (8.84)	0.00190*** (13.74)
Decile 3	26.86*** (3.36)	77.99*** (10.30)	66.08*** (7.87)	71.89*** (9.21)	0.000286*** (3.29)	0.00248*** (17.79)	0.000969*** (12.11)	0.00307*** (18.82)	0.000298*** (3.57)	0.00228*** (17.64)	0.000921*** (11.57)	0.00280*** (18.67)
Decile 4	45.49*** (5.12)	106.3*** (12.98)	83.73*** (9.09)	95.99*** (11.41)	0.000695*** (7.05)	0.00314*** (23.02)	0.00162*** (19.26)	0.00404*** (25.44)	0.000673*** (7.23)	0.00286*** (22.70)	0.00152*** (17.85)	0.00371*** (24.89)
Decile 5	69.24*** (7.44)	136.6*** (17.28)	107.5*** (11.38)	126.8*** (15.80)	0.00102*** (9.15)	0.00352*** (25.77)	0.00207*** (22.21)	0.00454*** (28.79)	0.000979*** (9.54)	0.00322*** (25.40)	0.00195*** (21.75)	0.00419*** (28.30)
Decile 6	95.13*** (9.44)	166.4*** (19.41)	130.6*** (12.89)	152.1*** (18.05)	0.00121*** (10.50)	0.00360*** (24.59)	0.00240*** (25.23)	0.00483*** (28.25)	0.00116*** (10.85)	0.00332*** (25.10)	0.00227*** (24.34)	0.00449*** (29.03)
Decile 7	126.9*** (12.18)	194.9*** (23.23)	163.2*** (15.69)	182.7*** (21.69)	0.00138*** (11.26)	0.00365*** (24.26)	0.00277*** (26.97)	0.00506*** (29.43)	0.00134*** (11.83)	0.00339*** (24.63)	0.00266*** (26.51)	0.00473*** (30.07)
Decile 8	159.1*** (15.36)	242.0*** (26.32)	197.3*** (19.36)	230.5*** (25.46)	0.00164*** (12.97)	0.00412*** (25.99)	0.00315*** (29.57)	0.00558*** (30.89)	0.00161*** (13.65)	0.00388*** (26.72)	0.00306*** (29.65)	0.00529*** (32.06)
Decile 9	218.1*** (20.88)	323.2*** (29.93)	258.1*** (26.33)	312.2*** (29.00)	0.00182*** (14.40)	0.00370*** (21.70)	0.00344*** (30.20)	0.00544*** (28.13)	0.00175*** (14.93)	0.00352*** (22.79)	0.00332*** (30.59)	0.00520*** (29.66)
Decile 10 (Highest)	318.6*** (29.33)	454.4*** (44.81)	357.0*** (35.06)	449.0*** (43.91)	0.00244*** (19.71)	0.00412*** (23.56)	0.00406*** (32.60)	0.00604*** (31.04)	0.00221*** (19.09)	0.00378*** (24.71)	0.00379*** (32.34)	0.00566*** (32.22)
Decile 1 (Lowest)	-125.1*** (-9.08)	-140.3*** (-16.53)	-214.0*** (-14.93)	-181.0*** (-21.22)	-0.000783*** (-6.15)	-0.00232*** (-15.51)	-0.00335*** (-29.97)	-0.00493*** (-28.19)	-0.000410*** (-3.46)	-0.00186*** (-14.11)	-0.00287*** (-25.87)	-0.00438*** (-28.37)
Nb observations	1671649	514720	1650762	514769	1632223	507039	1610782	506973	1669956	514506	1648969	514580
R-Square	0.011	0.020	0.012	0.019	0.006	0.008	0.015	0.020	0.005	0.008	0.014	0.019

## Table 2.5: Order choice

This table compares the main characteristics of the round-trips executed by each decile groups. It shows the proportion of limit-orders, the average amount, the average level of leverage and the average fees paid by each decile (from the 2nd to the 10th) relative to the bottom group. Standard errors are clustered at the individual level. \*, \*\*, \*\*\* denote significance at the 10%, 5% and 1% levels.

	Limit Orders	Amount	Leverage	Fees
Decile 2	0.0409 (1.64)	79.94 (0.36)	-0.0285 (-1.34)	-0.00161*** (-5.10)
Decile 3	0.0326 (1.15)	698.4* (2.38)	-0.105*** (-5.54)	-0.00250*** (-8.69)
Decile 4	0.0387 (1.45)	826.4** (3.13)	-0.126*** (-6.80)	-0.00265*** (-9.18)
Decile 5	0.0840** (2.78)	1381.3*** (4.70)	-0.114*** (-4.43)	-0.00349*** (-10.31)
Decile 6	0.121*** (4.78)	2319.4*** (7.60)	-0.145*** (-7.58)	-0.00368*** (-12.85)
Decile 7	0.178*** (5.13)	2631.3*** (5.99)	-0.144*** (-3.54)	-0.00414*** (-9.91)
Decile 8	0.145*** (4.97)	1376.2*** (5.16)	-0.318*** (-13.33)	-0.00312*** (-9.67)
Decile 9	0.0900** (2.87)	1182.8* (2.15)	-0.342*** (-15.95)	-0.00273*** (-8.34)
Decile 10 (Highest)	0.146*** (5.55)	394.0 (0.69)	-0.318*** (-18.31)	-0.00177*** (-5.34)
Decile 1 (Lowest)	0.546*** (26.97)	3600.0*** (27.26)	0.656*** (68.38)	0.00802*** (31.29)
<i>N</i>	6373919	8899226	8899226	6644667

## Table 2.6: Contrarian behavior

This table shows the results from the three following regressions

$$Type_{i,s,t} = \alpha + \sum_{d=2}^{10} Decile_{d,i} + \epsilon_i,$$

where  $Type_{i,s,t}$  is one of the possible three dummy variables  $Contr_{i,s,t}$ ,  $Mom_{i,s,t}$  or  $Round_{i,s,t}$  that identify whether an individual daily-imbalance is a contrarian, a momentum or a round-trip trade.  $Decile_{d,i}$  is a dummy variable that equals one only if individual  $i$  belongs to decile  $d$  and  $\epsilon_i$  is a error term that is clustered at the individual level. The bottom decile constitutes the omitted category. \*, \*\*, \*\*\* denote significance at the 10%, 5% and 1% levels.

	Contrarian	Momentum	Round Trip
Decile 2	0.00967 (1.14)	-0.0146 (-1.77)	0.00495* (2.29)
Decile 3	-0.00615 (-0.96)	-0.00473 (-0.73)	0.0109*** (4.38)
Decile 4	0.00946 (1.52)	-0.0176** (-2.91)	0.00815** (3.21)
Decile 5	0.0158* (2.32)	-0.0237*** (-3.34)	0.00784** (3.05)
Decile 6	0.0264*** (4.18)	-0.0346*** (-5.36)	0.00820** (3.02)
Decile 7	0.0238*** (3.68)	-0.0330*** (-4.61)	0.00923 (1.60)
Decile 8	0.0272*** (4.09)	-0.0277*** (-4.12)	0.000525 (0.09)
Decile 9	0.0519*** (9.02)	-0.0409*** (-7.25)	-0.0110*** (-3.72)
Decile 10 (Highest)	0.0949*** (18.78)	-0.0762*** (-14.84)	-0.0187*** (-11.45)
Decile 1 (Lowest)	0.579*** (144.93)	0.388*** (95.34)	0.0323*** (22.09)
<i>N</i>	6396831	6396831	6396831

## Table 2.7: Round-trips durations

This table shows some summary statistics about the duration (in days) of the round-trips submitted by each decile group.

	Round-Trip Length				
	mean	sd	p25	p50	p75
Decile 1 (Lowest)	46.04494	159.5033	1	7	26
Decile 2	49.57374	165.9789	2	8	30
Decile 3	48.35777	154.2482	2	8	30
Decile 4	58.10635	175.0067	2	11	39
Decile 5	57.45234	174.0326	2	9	37
Decile 6	58.00799	177.4783	2	10	38
Decile 7	61.63532	181.2242	2	10	42
Decile 8	69.93181	200.7007	2	11	48
Decile 9	95.19766	252.7459	3	16	69
Decile 10 (Highest)	135.3675	321.4881	4	24	105
Total	66.52783	199.1449	2	10	43
<i>N</i>	1046113				

## Table 2.8: Performance Persistence

I compute each year the Sharpe ratio net of trading costs for the round-trips of each investor in my sample. I focus on investors who submitted at least 10 round trips in a single year in order to obtain valid (yearly) Sharpe ratio estimates. An investor is classified as a winner in given year if his net Sharpe ratio during the same year is positive and is classified as a loser if his net Sharpe ratio is negative. The table is obtained as follows: rows indicates winning or losing investors in the *current* year, and columns indicates, conditional of their status in the current year, the number of investor that are either winners or losers the year after. The table gives the corresponding number of trades, percentage and cumulative percentage for each category of information. The total number of trades may be different from a category to another due to missing values.

Transition in 2003/2004										
	LOSE-STAY		LOSE-QUIT		WIN-STAY		WIN-QUIT		Total	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
LOSE	743	77.7	44	4.6	167	17.5	2	0.2	956	100.0
WIN	551	51.0	14	1.3	513	47.5	2	0.2	1080	100.0
Total	1294	63.6	58	2.8	680	33.4	4	0.2	2036	100.0

Transition in 2004/2005										
	LOSE-STAY		LOSE-QUIT		WIN-STAY		WIN-QUIT		Total	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
LOSE	724	61.4	29	2.5	420	35.6	7	0.6	1180	100.0
WIN	243	26.8	6	0.7	654	72.0	5	0.6	908	100.0
Total	967	46.3	35	1.7	1074	51.4	12	0.6	2088	100.0

Transition in 2005/2006										
	LOSE-STAY		LOSE-QUIT		WIN-STAY		WIN-QUIT		Total	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
LOSE	760	73.6	25	2.4	241	23.4	6	0.6	1032	100.0
WIN	722	47.0	15	1.0	783	51.0	16	1.0	1536	100.0
Total	1482	57.7	40	1.6	1024	39.9	22	0.9	2568	100.0

Transition in 2006/2007										
	LOSE-STAY		LOSE-QUIT		WIN-STAY		WIN-QUIT		Total	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
LOSE	1270	92.4	31	2.3	69	5.0	5	0.4	1375	100.0
WIN	942	82.5	11	1.0	183	16.0	6	0.5	1142	100.0
Total	2212	87.9	42	1.7	252	10.0	11	0.4	2517	100.0

Transition in 2007/2008										
	LOSE-STAY		LOSE-QUIT		WIN-STAY		WIN-QUIT		Total	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
LOSE	1345	87.2	45	2.9	149	9.7	3	0.2	1542	100.0
WIN	65	60.7	1	0.9	41	38.3	0	0.0	107	100.0
Total	1410	85.5	46	2.8	190	11.5	3	0.2	1649	100.0

Transition in 2008/2009										
	LOSE-STAY		LOSE-QUIT		WIN-STAY		WIN-QUIT		Total	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
LOSE	603	51.3	19	1.6	545	46.3	9	0.8	1176	100.0
WIN	48	26.1	3	1.6	132	71.7	1	0.5	184	100.0
Total	651	47.9	22	1.6	677	49.8	10	0.7	1360	100.0

Transition in 2009/2010										
	LOSE-STAY		LOSE-QUIT		WIN-STAY		WIN-QUIT		Total	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
LOSE	351	73.7	54	11.3	60	12.6	11	2.3	476	100.0
WIN	231	43.0	53	9.9	223	41.5	30	5.6	537	100.0
Total	582	57.5	107	10.6	283	27.9	41	4.0	1013	100.0

## Table 2.9: Trading performance and limits-to-arbitrage I

I run the following regression, separately for each decile of investors:

$$R_{s,t,k} = \alpha_s + \sum_{i=2}^{10} DecileVol_{s,t} + Year_t + \epsilon_{s,t}$$

Where  $R_{s,t,k}$  is the abnormal return net of fees of the k-th round-trip started on day t and stock s by an investor in a given decile.  $\alpha_s$  is a stock fixed effect,  $Year_t$  is a year fixed-effect and  $DecileVol_{s,t}$  is a dummy variable that classifies the round-trips into deciles of intraday volatility according to the volatility level of stock s on day t (the starting date of the round-trip).  $\epsilon_{s,t}$  is an error term that is doubly clustered by stock s and starting date t. \*, \*\*, \*\*\* denote significance at the 10%, 5% and 1% levels.

	Decile 1 (Lowest)	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10 (Highest)
Vol Decile 2	-0.000725*** (-3.87)	-0.000661*** (-5.01)	-0.000303*** (-3.34)	-0.000403*** (-4.40)	-0.000255*** (-3.27)	-0.000192** (-2.55)	-0.000183*** (-2.73)	-0.000313*** (-4.14)	-0.000230*** (-2.96)	-0.000430*** (-4.02)
Vol Decile 3	-0.00105*** (-5.30)	-0.000901*** (-6.36)	-0.000426*** (-4.38)	-0.000400*** (-4.00)	-0.000278*** (-3.11)	-0.000138* (-1.72)	-0.000190*** (-2.58)	-0.000252*** (-3.03)	-0.000269*** (-2.98)	-0.000413*** (-3.73)
Vol Decile 4	-0.00120*** (-5.61)	-0.000911*** (-6.32)	-0.000541*** (-5.23)	-0.000493*** (-4.77)	-0.000365*** (-3.82)	-0.000116 (-1.29)	-0.000134* (-1.72)	-0.000231*** (-2.60)	-0.000272*** (-2.74)	-0.000297** (-2.49)
Vol Decile 5	-0.00116*** (-5.13)	-0.000963*** (-5.96)	-0.000600*** (-5.42)	-0.000389*** (-3.80)	-0.000227** (-2.34)	-0.0000528 (-0.55)	-0.0000157 (-0.17)	-0.0000182 (-0.19)	-0.0000541 (-0.52)	-0.000128 (-0.98)
Vol Decile 6	-0.00121*** (-5.55)	-0.000946*** (-5.72)	-0.000613*** (-4.32)	-0.000292** (-2.32)	-0.000141 (-1.34)	0.0000606 (0.59)	0.0000478 (0.51)	0.000133 (1.31)	0.000118 (1.04)	0.0000300 (0.20)
Vol Decile 7	-0.00106*** (-4.40)	-0.000815*** (-4.72)	-0.000466*** (-3.26)	-0.000306** (-2.48)	-0.000000935 (-0.01)	0.000101 (0.90)	0.000198* (1.84)	0.000450*** (4.49)	0.000430*** (3.44)	0.000425*** (2.92)
Vol Decile 8	-0.00125*** (-4.56)	-0.000775*** (-4.08)	-0.000336** (-2.35)	0.0000432 (0.30)	0.000211* (1.66)	0.000358*** (2.83)	0.000320*** (2.63)	0.000648*** (4.95)	0.000596*** (4.30)	0.000752*** (4.40)
Vol Decile 9	-0.000817*** (-2.93)	-0.000270 (-1.31)	-0.000306* (-1.75)	0.000387** (2.53)	0.000361** (2.48)	0.000725*** (4.99)	0.000484*** (3.46)	0.00110*** (7.19)	0.00133*** (7.74)	0.00120*** (5.33)
Vol Decile 10	-0.000923*** (-2.78)	-0.000319 (-1.21)	-0.0000247 (-0.10)	0.000460** (2.29)	0.000744*** (3.68)	0.000885*** (4.38)	0.00118*** (6.32)	0.00164*** (8.47)	0.00167*** (7.79)	0.00161*** (5.73)
Nb observations	90387	121118	158794	190527	207841	231270	221841	171100	112480	82619
R-Square	0.028	0.020	0.011	0.008	0.008	0.006	0.005	0.007	0.012	0.021

## Table 2.10: Trading performance and limits-to-arbitrage II

I look at the time serie variation of the abnormal returns earned, on average, by each decile group of investors. To do so, I regress the abnormal returns earned by each decile group of investors on a set of year-dummy variables (from 2000 to 2010) and I double-cluster the standard errors at the day-stock level. \*, \*\*, \*\*\* denote significance at the 10%, 5% and 1% levels.

	Decile 1 (Lowest)	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10 (Highest)
2000	-0.00260*** (-13.88)	-0.00139*** (-7.69)	-0.00138*** (-8.14)	-0.000583*** (-3.53)	-0.0000443 (-0.26)	0.0000685 (0.42)	-0.0000307 (-0.20)	-0.000182 (-1.01)	0.000224 (1.39)	0.000496*** (2.99)
2001	-0.00437*** (-14.84)	-0.00283*** (-12.73)	-0.00164*** (-7.32)	-0.00136*** (-6.10)	-0.000586*** (-3.00)	-0.000567*** (-3.16)	0.0000461 (0.26)	0.000153 (0.93)	0.000506*** (3.13)	0.00164*** (8.00)
2002	-0.00471*** (-19.26)	-0.00362*** (-17.88)	-0.00298*** (-12.95)	-0.00210*** (-9.87)	-0.00121*** (-6.20)	-0.00120*** (-5.50)	-0.000476** (-2.51)	0.0000977 (0.51)	0.000440** (2.57)	0.00149*** (6.45)
2003	-0.00657*** (-20.81)	-0.00477*** (-18.66)	-0.00377*** (-14.78)	-0.00265*** (-13.43)	-0.00174*** (-9.01)	-0.00146*** (-9.88)	-0.000882*** (-6.48)	-0.000600*** (-4.70)	-0.000311*** (-2.71)	0.000485*** (2.72)
2004	-0.00672*** (-19.11)	-0.00478*** (-15.71)	-0.00355*** (-18.41)	-0.00266*** (-18.05)	-0.00207*** (-15.05)	-0.00156*** (-13.68)	-0.00114*** (-12.31)	-0.000654*** (-6.32)	-0.000650*** (-6.18)	-0.000180 (-1.09)
2005	-0.00650*** (-22.04)	-0.00410*** (-30.79)	-0.00300*** (-23.20)	-0.00239*** (-20.67)	-0.00172*** (-16.76)	-0.00123*** (-12.25)	-0.000811*** (-8.62)	-0.000575*** (-5.99)	-0.000574*** (-7.34)	-0.000153* (-1.73)
2006	-0.00588*** (-22.04)	-0.00325*** (-17.85)	-0.00307*** (-20.79)	-0.00205*** (-15.33)	-0.00152*** (-11.84)	-0.00125*** (-11.21)	-0.000608*** (-5.67)	-0.000401*** (-3.80)	-0.0000370 (-0.23)	0.000653*** (3.69)
2007	-0.00468*** (-20.27)	-0.00342*** (-17.25)	-0.00209*** (-16.07)	-0.00181*** (-14.98)	-0.00151*** (-13.16)	-0.000967*** (-8.84)	-0.000609*** (-5.95)	-0.0000493 (-0.47)	0.000175* (1.69)	0.00110*** (8.35)
2008	-0.00661*** (-24.80)	-0.00417*** (-19.56)	-0.00294*** (-14.29)	-0.00160*** (-9.72)	-0.00153*** (-8.85)	-0.000606*** (-3.55)	0.0000378 (0.23)	0.000750*** (3.82)	0.00169*** (6.77)	0.00354*** (12.06)
2009	-0.00706*** (-17.41)	-0.00579*** (-24.72)	-0.00395*** (-17.29)	-0.00255*** (-15.57)	-0.00220*** (-14.71)	-0.00152*** (-9.24)	-0.00106*** (-7.49)	0.0000153 (0.10)	0.000377** (2.11)	0.00148*** (6.50)
2010	-0.00614*** (-13.57)	-0.00506*** (-20.05)	-0.00280*** (-16.55)	-0.00224*** (-18.01)	-0.00208*** (-16.15)	-0.00117*** (-9.78)	-0.00129*** (-9.57)	-0.000127 (-0.85)	0.000538*** (3.37)	0.00131*** (8.76)
Nb observations	70780	106168	146447	177208	196432	220336	212312	162049	103377	68869
R-Square	0.148	0.100	0.071	0.045	0.029	0.017	0.009	0.003	0.007	0.031

# 3 Chapter Three

## **Retail trading strategies around earnings announcements**

I show that retail investors increase market participation and trading activity around earnings announcements because they are able to forecast the stock market reaction following the announcement. To obtain this result, I exploit a large brokerage dataset that allows me to identify the trading strategies executed by a given investor before, during and after the event. Round-trips *initiated* one day before the announcement are much shorter in duration and earn a higher risk-adjusted return (net of transaction costs) relative to the one initiated in the non-announcement period. These results suggest that a subset of the retail population actively exploit some private information before the event, by buying before good news and selling before bad news. On the event day, these investors completely unwind their positions to cash-in their trading gains, thus generating a contrarian price pressure on that day.

## 3.1 Introduction

There is recent evidence that the aggregated order flow generated by individual investors can predict stock returns around stock earnings announcements (Kaniel, Liu, et al. [2012], Kelley and Tetlock [2013]). Yet, these papers exploit large brokerage datasets that do not allow to follow a single investor over time. That is, the authors of these two studies simply cannot identify which strategies are actually pursued by individual investors before, on and after the announcement.

However, keeping track of the trading strategies implemented by individuals around the announcement event is important, for at least two reasons. First, identifying those trading strategies and evaluating their performance can bridge the gap between the recent findings in Kelley and Tetlock [2013] and Kaniel, Liu, et al. [2012], that shed a positive light on the trading skill of retail traders, and the previous literature that generally depicts the individual investors as naive “noise traders”. Specifically, the fact that the aggregate retail order flow has some predictive power for future stock returns around announcement events does not necessarily imply that some retail investors are able to generate trading gains by trading around the event. In other words, whether individuals are able to profit from specific trading strategies around earnings announcement days, after risk-adjustment and after the inclusion of transaction costs, is still an open question in the literature.<sup>1</sup>

Second, identifying which strategies are implemented around earnings announcements may help to explain several puzzling price patterns that have been identified in the literature. For instance, both the “*earnings announcement premium*” (Frazzini and Lamont [2007]) and the “*earnings announcement drift*” (Taylor [2010] and Kaniel, Liu, et al. [2012]) seem to be driven

---

<sup>1</sup>Consistent with the noise trader hypothesis, Barber and Odean [2011] argue that “*the short-run return predictability and the poor performance of individual investors are easily reconciled, as the average holding period for individual investors is much longer than a few weeks.*”. Alternatively, Kelley and Tetlock [2013] suggest that “*through learning or attrition, the aggregate skill of retail traders may have changed over time*”, so that some retail investors are effectively able to generate net trading profits by trading around the announcement events. However, Kelley and Tetlock [2013] cannot conclude because they observe neither the trader identities nor the transactions costs.

in part by the trading behavior of individual investors.<sup>2</sup> As in [Kaniel, Liu, et al. \[2012\]](#), severe data-limitations do not allow the authors of these studies to follow the individual traders over time, and thus confirm that some retail trading strategies can exert enough price pressure to generate the two puzzles above.

The scope of this paper is to identify directly the trading strategies implemented by retail investors around earnings announcements days, and to assess their net performance while controlling for unobserved and time-invariant individual fixed effects. In other words, I am able in this paper to overcome the previous data limitations found in the literature by keeping track of the trading strategies implemented by any given individual before, on and after any earnings announcement. To do so, I exploit a large and novel brokerage dataset which contains the detailed trades, along with the trading fees incurred, of more than 140,000 French investors at the individual-stock-day level.

I obtain three main results. First, I find that retail investors sharply increase both market participation and trading volume before, on and shortly after a given announcement day. This evidence is consistent with [Vieru et al. \[2006\]](#), [Taylor \[2010\]](#) and [Etheber, Hennen, and Meyer \[2012\]](#), but the pattern I document is much more concentrated around the announcement event than previously found in the literature.

Second, I show that retail investors *actively* try to exploit some private information before the announcement by executing very short-term round-trips around the event. As explained in [Kelley and Tetlock \[2013\]](#) the analysis of the actual individual trading strategies implemented before, on and after the announcement day “*is particularly interesting because, if individuals trade on information prior to the announcements, it could be the case that they reverse their positions after the announcements. Unfortunately, our data do not permit us to observe the strategy of specific individuals and hence we cannot unequivocally identify such trading patterns*”. I find indeed that

---

<sup>2</sup>The “*earnings announcement premium*” is the fact that stock prices rise around scheduled earnings announcement dates (see for instance [Barber, De George, et al. \[2012\]](#)). The “*earnings announcement drift*” is the tendency for a stock’s cumulative abnormal returns to drift in the same direction of the earnings surprise after an earnings announcement (see also [Hirshleifer, Myers, and Myers \[2008\]](#)).

during the very few days *before* the announcement day, the duration in days of the round-trips executed by individual investors in my sample sharply *decrease*, thus suggesting that individual investors unwind their speculative trades shortly after the earnings announcement to cash-in their gains. Indeed, round-trips initiated exactly one day prior to the announcement earn an additional gross return of about 11 basis points (about 7 basis points after the inclusion of transaction costs) relative to those initiated in the non-announcement period.

Third, I show that while the round-trips generated one day before the event are more profitable than those started during the non-announcement period, this is not the case for round-trips that are initiated *on* the announcement day and during the following days. Round-trips initiated on the announcement day and the day after actually *decrease* the round-trip returns by about 4 basis points relative to the returns earned during the non-announcement period. This difference in performance suggests that individual investors may take advantage of some private information before the release of the earnings news, but they are less able to process and react efficiently to new information that has just been released after the earnings announcement.

This trading behavior has the potential to slow-down the adjustment of prices after the news release. Specifically, I find that retail investors are strongly contrarians with respect to the *intraday return* on the day of the announcement: they massively sell when there is a very high intraday return and they massively buy when there is a very low intraday return on the announcement day, which is the direct consequence of a profit-taking behavior after the event. Indeed, I show that, conditional on being completely reversed during the very few days after the earnings announcement, the average performance of the round-trips initiated a few days before the event are *positive* even after risk-adjustment and the inclusion of transaction costs. This pattern confirms thus that retail investors cash-in their gains from private information just after the event.<sup>3</sup>

To obtain these results, I exploit a very large panel brokerage dataset that covers almost

---

<sup>3</sup>The unconditional average performance is negative, however. This result suggests that some retail investors are prone to the “disposition effect” after the event.

a decade of complete trading records for more than 140,000 individual investors in France. I adopt a standard event study methodology (MacKinlay [1997], Frank and Goyal [2007]) that uses the trades executed by the individual investors in my sample on a 50 trading day window centered on any given announcement day. Then, to study how the individual trading patterns vary before, on and after the announcement, relative to the non-announcement period, I compute abnormal measures of market participation, market activity and trading performance as in Nofsinger [2001] and Christophe, Ferri, and Angel [2004].

To assess the profitability of the trading strategies executed by individuals, I adapt the methodology described in Schlarbaum, Lewellen, and Lease [1978]. This methodology is based on investors' round-trips: a round-trip on a stock, for any given investor, starts and ends with a zero net inventory position on that stock. As Schlarbaum, Lewellen, and Lease [1978] argue, *“the dates, prices, and commissions recorded in the transactions file are “hard” data; they represent actual investment activities. As such, they eliminate any need to choose an arbitrary periodic-return measurement interval, to adjust for portfolio changes during that interval, or to allocate transactions costs across intervals by some amortization rule. A realized-return, investment-round-trip orientation, therefore, is persuasively clean.”* This methodology allows me to control for the difference in duration between round-trips (this duration is the number of days between those two events), and to include the transaction costs incurred. That is, as in Shapira and Venezia [2001] and Chakrabarty, Moulton, and Trzcinka [2013], I obtain risk-adjusted returns for each individual round-trip by subtracting the return that the individual investor would have obtained from a passive investment in the market index (over the same duration period) to the (gross or net) return of the round-trip.

This article proceeds as follows. Section 3.2 presents my brokerage dataset and Section 3.3 presents my methodology. Results are discussed in Section 3.4.

## 3.2 Data

I describe in this section the data used in this paper, and my methodology for assessing the investors' trading performance and investors' behavior around earnings announcements.

### 3.2.1 The brokerage dataset

The data used in this paper comes from a leading French online broker.<sup>4</sup> The raw dataset contains at the daily level all the executed trades sent by the 145801 distinct broker's customers from 1999 up to 2010, which represents more than 15 million trades.<sup>5</sup>

Each trade comes with the following information: the asset type (equity, bonds, etc), the trading exchange identifier (the ISIN), the trading date, the quantity, the order type (limit, market and other minor orders types), the amount traded in euro, the trading place, the transaction fees incurred (from 2006 onward) and the trading channel used to submit the order. I match the trades in my dataset with market data provided by Eurofidai, the European financial data institute. Trades are matched by ISIN code, trading day and trading exchange code. Trades for which no information is available from Eurofidai are discarded from the sample.

Summary statistics of the raw brokerage dataset are provided in Table 3.1. This table shows the corresponding number of trades, percentage and cumulative percentage for several categorical variables that describe the nature of my data well. Panel A shows for instance that trades are in the majority of cases limit orders (62%) and market orders (28%). The other minor orders types are only used marginally by investors. The information on order type is completely missing for 1999 and 2000 and may be marginally missing for up to 2004. Therefore these figures do not take into account the missing values.

---

<sup>4</sup>A subsample of this database covering the period 1999-2001 has already been used, to address another research question, in [Foucault, Sraer, and Thesmar \[2011\]](#).

<sup>5</sup>This section is taken from my first chapter.

Investors could submit an order by using the telephone to speak to a broker official (Telephone), by calling a voicemail service and typing instruction using the telephone's keys (Phone+), by using a web navigator to connect to the broker website (Web), by using an old French Videotex online service accessible through the telephone lines (Minitel) or by using online basic computer software (Online+ and Trader+). Panel B of the table highlights that a large majority of trades are submitted using the internet or trading software (Online+ and Trader+). Thus, my sample contains investors in the post-online trading era.

Panel C and Panel D of the table show that most of individual trading activity consists of buying or selling common stocks on the NYSE Euronext Paris trading exchange. This dataset, therefore, shows patterns of individual investment behavior that are similar to other recent databases used in the literature. For instance, Finnish investors in [Linnainmaa \[2010\]](#) also use limit orders for most of their trades on the Helsinki Stock Exchange.

<Insert [Table 3.1](#) here>

As I will use in the next sections some standardized measures of abnormal buying and selling activity, I also show in [Table 3.2](#) the number of buy and sell orders separately, year after year. This table simply shows that the ratio is roughly equivalent between buy and sell orders.

<Insert [Table 3.2](#) here>

### 3.2.2 Earnings Announcements

I explain in this section how I collect the data on announcement dates for the stocks in my sample.

As explained in [Barber, De George, et al. \[2012\]](#), obtaining accurate announcement dates for European stocks is very difficult because the leading provider of such piece of information, IBES, often reports those dates with several days of delay. I therefore use Bloomberg to recover

the announcement dates for all the firms in my sample, as the accuracy of Bloomberg is very high for european stocks (Barber, De George, et al. [2012]).

Moreover, as the exact timing of the announcement is critical in this paper, I also collect from Bloomberg the exact announcement timestamp of the earnings release. As in Christophe, Ferri, and Angel [2004], I then classify the trading day  $t$  as being *on the announcement day* ( $t = 0$ ) if the earnings announcement has been released *before* the market opening. If the announcement has occurred after the market closing, I mark this day as  $t = -1$ . Most of the announcements in my sample occur before the market opening, but still a significant part of them occur after the market closing, so this differentiation is important.

For some robustness checks, I also obtain the earnings surprise for any announcement in my sample using the IBES surprise history. To overcome the accuracy problems above, I require that both Bloomberg and IBES agree on the closest announcement date. Furthermore, I restrict my sample to the announcements events where the exact timestamp from Bloomberg is obtained.

I show in Table 3.3 the number of trades in my sample that occur on earnings events, depending on whether the announcement has been made *before or after* the market open session. It appears that a majority of the announcement events happens before the opening price (and retail investors also trade more on those events), but it is still important to correctly mark as  $t=-1$  the days where the announcement has been made after the market closure.

<Insert Table 3.3 here>

## 3.3 Methodology

### 3.3.1 Event Study

I describe in this section the methodology I use in this paper to focus on the retail trading patterns around the announcement events.

As in [Nofsinger \[2001\]](#), I use an event study methodology. Specifically, I first aggregate transactions in my dataset at the id-day-stock level and at the stock-day level. Then I obtain measures of abnormal market participation and abnormal trading volume as follows.

I define the standardized individual trading measure  $X$  for stock  $s$  on day  $t$  as:

$$AbnX_{s,t} = \frac{X_{s,t}}{M_{t=-50}^{-5} X_{s,t}} - 1$$

where  $M_{t=-50}^{-5} X_{s,t}$  is the median of the individual trading measure  $X_{s,t}$  computed with data between -50 days to -5 days relative to the next announcement day ( $t = 0$ ). The idea behind this computation is simple: the denominator captures the *normal* level of trading volume for firm  $s$  in the non-announcement period. So this measure of trading activity expresses to which extent the trading volume around the earning announcement is different from that normal period.

$X_{s,t}$  in this paper can be the number of (executed) buy orders, sell orders or total executed orders on stock  $s$  at day  $t$ . I also compute the distinct number of market and/or limit orders executed on stock  $s$  and day  $t$  and the number of orders executed on a margin/spot trading account. The net imbalance is obtained as follows:

$$AbnNet_{s,t} = AbnBuy_{s,t} - AbnSell_{s,t}$$

I also compute an abnormal measure of abnormal market participation by counting the number of distinct investors trading on stock  $s$  on a given day, as in [Vieru et al. \[2006\]](#). For instance, when  $X$  represents the number of Buy orders, a large positive value for  $Abn_{Vol}$  on day  $t$  indicates that there is an abnormally high buying pressure that day. I assess how large these deviations are from normality using OLS regressions, as described in the next paragraph.

To have meaningful measures of abnormal and normal trading volume, I require at least 15 days with non-zero trading volume in the  $[-50, -5]$  relative period interval. I also require

at least one trade during the announcement period. Then, as in [Taylor \[2010\]](#) and [Etheber, Hennen, and Meyer \[2012\]](#) I create multiple dummy variables that isolate the relative position of a particular trading day-stock relative the to closest announcement day on that stock. I thus generate 11 dummy variables, spanning from -5 days prior to the announcement to +5 days after the announcement, and I use the following regression framework.

$$AbnX_{s,t} = \alpha + \sum_{i=-5}^{+5} \beta_i * I_{t=i} + \theta_t + \gamma_s + \epsilon_{s,t}$$

,

Where  $I_{t=0}$ , for instance, is a dummy variable that is one if the trading day  $t$  is on the day of the announcement, and zero elsewhere. Similarly,  $I_{t=-4}$  is a dummy variable that is one only if  $t = -4$ , that is the trading day  $t$  is exactly 4 days before the next announcement. I include a full range of year fixed-effects  $\theta_t$  and stock fixed effects  $\gamma_s$ . The coefficient  $\beta$  capture the effect of a particular day around the announcement date on my abnormal trading measures.

### 3.3.2 Investors' round-trips

To understand the strategies adopted by retail investors around the earnings announcements, I also need to accurately assess their trading performance. This performance evaluation is in itself a difficult methodological issue.<sup>6</sup>

Indeed, investors may trade for various reasons and may have very heterogeneous trading horizons that are not captured by the standard fixed-horizon performance measures used in the literature. I overcome this methodological challenge in this paper by following [Schlarbaum, Lewellen, and Lease \[1978\]](#): I aggregate investors' single trades into round-trip trades that fully reflect the actual cash-flows received or paid.

---

<sup>6</sup>This section is taken from my second chapter.

To do so, I first compute the daily net quantity traded by investor  $i$ , on stock  $s$  and day  $t$  as:

$$Q_{i,t,s} = Qtit_{i,t,s}^{buy} - Qtit_{i,t,s}^{sell},$$

where  $Qtit_{i,t,s}^{buy}$  (resp.  $Qtit_{i,t,s}^{sell}$ ) represents the actual quantity of stock  $s$  bought (resp. sold) by investor  $i$  on day  $t$ .

Then, I sort all the daily net quantities in my dataset by trader, stock and trading day and I keep track of the cumulative stock quantity held by a trader day after day. A round-trip starts and ends with a zero net cumulative quantity. Round-trips are thus trading positions that are fully unwound: stocks previously bought are entirely sold and stocks previously sold short are completely bought back. I provide in [Table 3.4](#) a detailed example of this procedure, and summary statistics are shown on [Table 3.5](#).

<Insert [Table 3.4](#) here>

The main advantage of computing round-trips over a traditional fixed-horizon approach is clearly stated in [Schlarbaum, Lewellen, and Lease \[1978\]](#):

*The dates, prices, and commissions recorded in the transactions file are hard data; they represent actual investment activities. As such, they eliminate any need to choose an arbitrary periodic-return measurement interval, to adjust for portfolio changes during that interval, or to allocate transactions costs across intervals by some amortization rule. A realized-return, investment-round-trip orientation, therefore, is persuasively clean*

This methodology is still rare in the literature because it requires brokerage datas that have a panel structure at the individual level, which is in general not possible in the US (see for instance [Kaniel, Saar, and Titman \[2008\]](#) or [Kelley and Tetlock \[2013\]](#)). My dataset thus provides an opportunity to enrich the inference previously drawn on investor trading behavior and performance.

Once a round-trip is identified, I compute several measures that will serve to assess the profitability of investors' trades. The duration of a round-trip is the number of days between the start and the end of a round-trip. I thus can obtain a proxy for investor  $i$ 's trading horizon, in a given month, by taking the average duration of all the investor  $i$ 's round-trips *started* that month.

The gross profit generated by a round trip is the (signed) sum of all the cash-flows arising within a given round-trip. A buy order represents a negative cash-flow corresponding to the amount paid, while a sell orders generate positive cash-flow corresponding to the amount received from the sell. I thus compute:

$$\text{Gross Profit}_{i,s}(t, T) = \text{Sell}_{i,s}(t, T) - \text{Buy}_{i,s}(t, T),$$

where  $\text{Gross Profit}_{i,s}(t, T)$  stands for the gross profit of a round-trip initiated on day  $t$ , closed on day  $T$ , on stock  $s$ , by investor  $i$ .  $\text{Sell}_{i,j}(t, T)$  corresponds to the sum of all the sell orders' euro amount executed within that round-trip while  $\text{Buy}_{i,j}(t, T)$  corresponds to the sum of all the buy orders' euro amount executed within the same round-trip. I compute net profits by subtracting the amount of transactions fees paid by the traders to total buy and sell cash-flows.

$$\text{Sell}_{i,s}^{\text{net}}(t, T) = \text{Sell}_{i,s}(t, T) - \text{Fee}_{i,s}^{\text{sell}}(t, T)$$

$$\text{Buy}_{i,s}^{\text{net}}(t, T) = \text{Buy}_{i,s}(t, T) + \text{Fee}_{i,s}^{\text{Buy}}(t, T)$$

Similarly, I compute the (gross and net) returns for individual's  $i$  round-trip on stock  $j$ , started on day  $t$  and closed on day  $T$  as:

$$R_{i,j}^{\text{gross}}(t, T) = \frac{\text{Sell}_{i,j}(t, T) - \text{Buy}_{i,j}(t, T)}{\text{Buy}_{i,j}(t, T)},$$

$$R_{i,j}^{net}(t, T) = \frac{Sell_{i,j}^{net}(t, T) - Buy_{i,j}^{net}(t, T)}{Buy_{i,j}^{net}(t, T)},$$

Importantly, following [Shapira and Venezia \[2001\]](#) and [Chakrabarty, Moulton, and Trzcinka \[2013\]](#), I keep track of the starting and ending date of each round-trip to compute risk-adjusted returns. To do so, I compare the (gross or net) return earned on a given round-trip to the return that would have been *passively* achieved over the *same* holding period by investing in the market index. :

$$R_{i,j}^{abnormal}(t, T) = R_{i,j}(t, T) - \frac{Index(T) - Index(t)}{Index(t)},$$

$$R_{i,j}^{abnormal-net}(t, T) = R_{i,j}^{net}(t, T) - \frac{Index(T) - Index(t)}{Index(t)},$$

where  $Index(t)$  stands for the general French market index computed by Eurofidai, or the French CAC40 index value at the end of day t. Adjusting returns using French indexes as market benchmarks is appropriate in my setting because the large majority of trades in the database are executed on French stocks.

Last, as different round-trips may have different durations, I follow [Shapira and Venezia \[2001\]](#), and I convert each (gross or net) abnormal return into an equivalent *daily-return* using the formula:<sup>7</sup>

$$R_{i,j}^{daily}(t, T) = (1 + R_{i,j}^{abnormal-net}(t, T))^{(1/T)} - 1$$

---

<sup>7</sup>Instead, annualizing return could be problematic as it could lead to extremely high figures.

## 3.4 Results

### 3.4.1 Abnormal trading volume and participation around earnings announcements

I explain in this section how the retail trading volume and the individual market participation are affected around earnings announcement dates. The results of the regression described in [Section 3.3.1](#) are shown on [Table 3.6](#).

<Insert [Table 3.6](#) here>

[Table 3.6](#) shows that all the measures of abnormal trading activity and market participation experience large positive shocks starting one day before the announcement and lasting for a couple of days after the announcement. In terms of the magnitude of the increase, most of the abnormal trading activity I document is located on the day of the announcement ( $t=0$ ), that is during the first trading session that immediately *follows* the earnings' release.

For instance, column "Abn. Volume" shows how the abnormal volume, measured as the raw number of executed trades by individual investors on a given day, changes around earnings announcements. In this regression specification, the coefficient on the dummy variable "Announcement Day" ( $t=0$ ) is 1.105 and is highly significant. This coefficient shows that the abnormal volume activity is more than two times higher on the day of the announcement than during the non-announcement period (the constant is 0.49).

This increase in trading activity prior to the announcement is consistent with [Etheber, Hennen, and Meyer \[2012\]](#). They also find that the net Buy/Sell ratio (column "Abn Imbalance in [Table 3.6](#)) increases around the announcement, suggesting a possible "attention-grabbing" mechanism as in [Barber and Odean \[2007\]](#).

Importantly, the column labeled "Abn ID" shows that market participation also increases around earnings announcements. This result is important because it may be the case that the same number of investors just become more active around announcements days. The highly significant coefficients on the day before and on the announcement day suggest instead that there are more unique investors around the announcements, so that part of the abnormal trading activity I find is due to an increased market participation around those events.

### 3.4.2 Individual trading performance and trading patterns around earnings announcements

I study in this section the trading performance of the individual investors in my sample around the earnings announcement days. The results of the regression presented in [Section 3.3.1](#) and [Section 3.3](#) are shown in [Table 3.7](#).<sup>8</sup>

<Insert [Table 3.7](#) here>

[Table 3.7](#) shows that across all the measures of performance, round-trips started exactly one day before the announcement earn higher returns than round-trips that are started during the non-announcement period. The first column of the table shows for instance that the round-trip returns gained by individuals in my sample, after controlling for time and individual fixed effects, are non statistically different from zero in the non-announcement period but earn an additional return of about 11 basis points when initiated one day before the announcement.

This positive performance should however be contrasted with the regression in column two, that uses round-trip returns net of transaction costs. As in the previous case, the returns of the round-trips started one day before the announcement earn a positive premium, which suggests that the retail investors can predict stock returns in the short run right after a news release

---

<sup>8</sup>I use in this section individual fixed effects instead of stock fixed effects, to control for unobserved heterogeneity at the individual level.

(consistent with [Kelley and Tetlock \[2013\]](#)). However, when considering the *total average profit* this increase in profitability seems not to be sufficient to overcome the average losses of about -0.1% that are experienced during the non-announcement period. Recall that the constant in the regression captures the average individual fixed effects, so the (average) overall negative performance I document hides probably a large heterogeneity in trading performance: it may be the case that only a subset of the retail population is able to earn abnormal returns - net of fees- by trading around earnings release.

In any case, an additional piece of information is provided by the study of the round-trip lengths around the earnings release. This analysis is done on the last column of [Table 3.7](#). I use the length of a round-trip as the dependent variable and I restrict my sample to the round-trips that have a duration of less than 100 days. As one can see from the regression, the duration of the round-trips sharply decreases before and on the announcement, suggesting that retail investors try to take advantage of their short-term predictive power by implementing some very short-term strategies prior and on the announcement date.

This pattern is interesting. Indeed, it is not clear in the literature whether retail investors take profit from their unexplained ability to predict stock prices in the very short run. For instance, [Barber and Odean \[2011\]](#) state that:

*The one caveat to this general finding [that the long-term average performance of individual investors is poor] is the intriguing evidence that stocks heavily bought by individuals over short horizons in the U.S. (e.g., a day or week) go on to earn strong returns in the subsequent week, while stocks heavily sold earn poor returns. It should be noted that the short-run return predictability and the poor performance of individual investors are easily reconciled, as the average holding period for individual investors is much longer than a few weeks.*

My results would instead suggest the opposite. Indeed, [Table 3.7](#) (column "horizon") clearly shows that retail investors engage actively in short-term speculative activities just before the

earning announcement day ( $t=-1$ ). This pattern is consistent with individual investors trying to profit from their short-term predictive ability, an ability that is confirmed by the positive coefficient on the dummy variable Before in the other columns of [Table 3.7](#).

I provide visual evidence of this pattern in [Figure 3.2](#).

<Insert [Figure 3.2](#) here>

Subfigure b in [Figure 3.2](#) shows indeed that the round-trip duration quickly falls just before the announcement day, and then reverts to its normal level after the announcement, suggesting that retail engage in trading strategies that are specific to the earnings announcement period.

In summary, [Table 3.7](#) shows essentially that, on average, it seems that the predictive power of retail investors is not sufficient to cover the transactions costs paid. That is, they lose *less* when trading before the announcement but they still lose on average. However, this is an unconditional analysis that uses the performance of *all* the round-trips executed around the announcement day, irrelevant of the duration of the round-trip. Understanding how the performance varies around the event can be a way to see whether individual investors may have an effect on prices after the news-release. For instance, [Kaniel, Liu, et al. \[2012\]](#) state that:

*This analysis [individual trading strategies before and after the announcement day] is particularly interesting because, if individuals trade on information prior to the announcements, it could be the case that they reverse their positions after the announcements. Unfortunately, our data do not permit us to observe the strategy of specific individuals and hence we cannot unequivocally identify such trading patterns.*

Hence, I compute the gross and net returns of the executed round-trips by conditioning on the closing date. Indeed, if investors reverse their winning trading bets after the announcement day and hold on on losses (a behavior consistent with the disposition effect) then one should observe that returns are particularly high for round-trips initiated before the announcement day and completely unwinded shortly after. This pattern can be seen in [Figure 3.2](#).

To obtain the subfigure a in [Figure 3.2](#), I first restrict my sample to the round-trips that have been initiated during the period [-4,-1] relative to the next announcement day. Then I compute the average gross and net round-trip return *conditional* on the closing date. It appears quite clearly that there is a spike in performance for the round-trips unwinded on the announcement day and during the few days after the event. Interestingly, both gross and net returns show a positive performance on the event day. This behavior is of course consistent with the disposition effect (see [Shefrin and Statman \[1985\]](#)), that is retail investors reverse their speculative *winning* positions right after the event while holding on to losses.<sup>9</sup> [Frazzini \[2006\]](#), for instance, shows that the disposition effect among investors can generate underreaction to the earning news but the author cannot separate between retail and institutional investors.

To have a destabilizing effect on prices, for instance by slowing down the adjustment of prices after the earning-release, it should be the case that retail investors trade in a contrarian manner after the event. While the direct scope of this paper is to focus on the trading strategies of retail investors around the event, I show in [Figure 3.3](#) how the net daily imbalance (buys-sells) changes around the announcement event, *conditional* on the *intraday return* that day.

<Insert [Figure 3.3](#) here>

[Figure 3.3](#) shows a very clear pattern. In this graph I separate between announcement events with a very high and a very low intraday return, defined as:

$$R_{\text{intraday}} = \frac{\text{closing price} - \text{opening price}}{\text{opening price}}$$

[Figure 3.3](#) suggests that, on announcement events with a positive intraday return, investors sell because they unwind previously bought positions. Interestingly, this pattern is mirrored for

---

<sup>9</sup>The disposition effect bias is not a specificity of US investors and has been found in many other countries. See in particular [Boolell-Gunesh, Broihanne, and Merli \[2009\]](#), and [Boolell-Gunesh, Broihanne, and Merli \[2012\]](#) for recent studies on the disposition effect using a French dataset similar to the one used in this paper.

very negative returns on announcement days.<sup>10</sup> By trading massively in the opposite direction of the stock price on the announcement day, retail investors have the potential to slow down the adjustment of prices to new information.

### 3.5 Conclusion

I show in this paper that retail investors initiate specific round-trip strategies one day prior to an earnings announcement day. Those round-trips have a duration (in days) much shorter than usual and are more profitable than round-trips started during the non-announcement period, even after including transaction costs. This suggests that retail investors have private information prior to the event, trade in the same direction as the market reaction on the event day *before* the event, and subsequently reverse their position *after* the event. This trading behavior creates a strong contrarian price pressure on the announcement day.

---

<sup>10</sup>I also condition on announcement events having an extreme earning surprise (according to IBES). It seems that before events that have a very high *positive* surprise, retail buys and then massively sell after the event. This pattern seems not mirrored in the case of a very negative surprise. This asymmetry is probably best explained by the simple fact that retail investors do not react to earnings surprise, but are influenced instead by stock returns around the event (Kaniel, Liu, et al. [2012]).

### 3.5 References

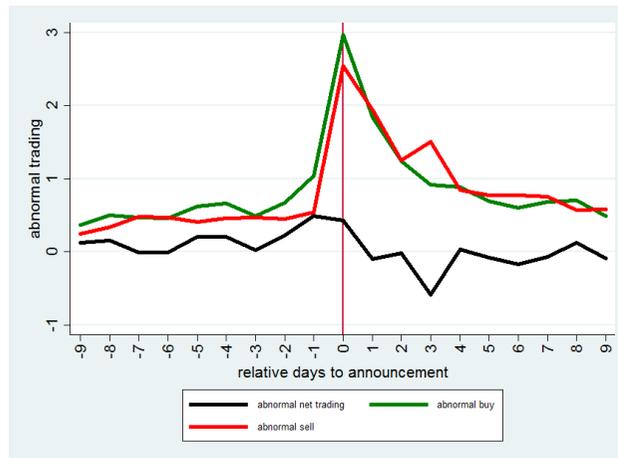
- Barber, B. and Odean, T. (Dec. 2007). "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors". *Review of Financial Studies* 21.2.
- Barber, B. and Odean, T. (2011). "The behavior of individual investors". *Working Paper*.
- Barber, M., De George, E., Lehavy, R., and Trueman, B. (Oct. 2012). "The earnings announcement premium around the globe". *Journal of Financial Economics*.
- Boolell-Gunesh, S., Broihanne, M., and Merli, M. (2009). "Disposition effect, investor sophistication and taxes: Some French Specificities". *Finance* 30.
- (2012). "Sophistication of Individual Investors and Disposition Effect Dynamics". *Finance* 33.
- Chakrabarty, B., Moulton, P., and Trzcinka, C. (2013). "Institutional Holding Periods". *Working Paper*.
- Christophe, S., Ferri, M., and Angel, J. (2004). "Short-Selling Prior to Earnings Announcements". *The Journal of Finance* LIX.4.
- Etheber, T., Hennen, D., and Meyer, S. (2012). "Earnings Announcements, Miller's Stocks and Private Investors ~ Trading Patterns and Performance". *Working Paper* January.
- Foucault, T., Sraer, D., and Thesmar, D. J. (2011). "Individual Investors and Volatility". *Journal of Finance* LXVI.4.
- Frank, M. Z. and Goyal, V. K. (2007). *Handbook of Empirical Corporate Finance*. Elsevier, pp. 135–202.
- Frazzini, A. (2006). "The disposition effect and underreaction to news". *The Journal of Finance* LXI.4.
- Frazzini, A. and Lamont, O. (2007). *The earnings announcement premium and trading volume*. May.
- Hirshleifer, D., Myers, J., and Myers, L. A. (2008). "Do Individual Investors Cause Post-Earnings drift". *The Accounting Review* 83.6.

- Kaniel, R., Liu, S., Saar, G., and Titman, S. (2012). "Individual investor trading and return patterns around earnings announcements". *The Journal of Finance* LXVII.2.
- Kaniel, R., Saar, G., and Titman, S. (2008). "Individual investor trading and stock returns". *The Journal of Finance* LXIII.1.
- Kelley, K. and Tetlock, P. (2013). "How wise are crowds? Insights from retail orders and stock returns". *The Journal of Finance* January.
- Linnainmaa, J. (2010). "Do limit orders alter inferences about investor performance and behavior?" *The Journal of Finance* LXV.4.
- MacKinlay, A. (1997). "Event studies in economics and finance". *Journal of economic literature* XXXV.March.
- Nofsinger, J. R. (July 2001). "The impact of public information on investors". *Journal of Banking & Finance* 25.7.
- Schlarbaum, G., Lewellen, W., and Lease, R. (1978). "Realized returns on common stock investments: The experience of individual investors". *Journal of Business* 51.2.
- Shapira, Z. and Venezia, I. (2001). "Patterns of behavior of professionally managed and independent investors". *Journal of Banking & Finance* 25.
- Shefrin, H. and Statman, M. (1985). "The disposition to sell winners too early and ride losers too long: Theory and evidence". *The Journal of finance* XL.3.
- Taylor, D. (2010). "Individual investors and corporate earnings". *Working Paper*.
- Vieru, M., Perttunen, J., Schadewitz, H., and Chadewitz (Jan. 2006). "How investors trade around interim earnings announcements". *Journal of Business Finance and Accounting* 33.March.

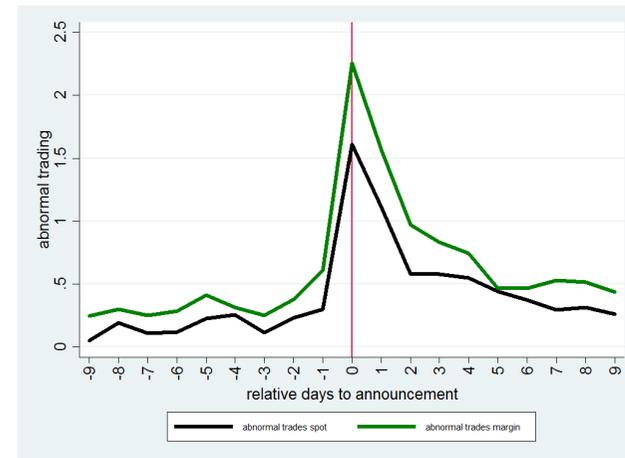
## 3.6 Figures

## Figure 3.1: Abnormal trading activity and abnormal market participation around earnings announcements

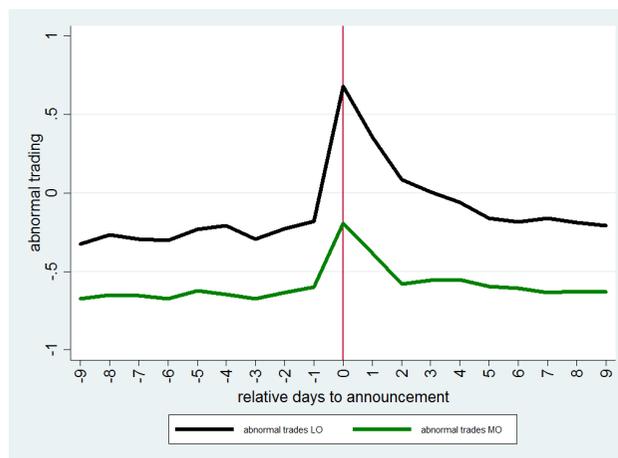
I show how the measures of abnormal trading volume and abnormal market participation used in this paper change around the earnings announcement date. On the x axis I plot the relative time to the closest earning announcement,  $t=0$  being the day of the announcement. In subfigure a I plot the abnormal net trading (buys-sells) and the number of executed buy and sell orders. In subfigure b, I separate the orders executed on a leveraged margin account from those executed on a cash-spot account. In subfigure c I separate between the executed limit orders and the executed market orders. Last, in figure d is plot the abnormal number of unique investors trading on a single day around the announcement date.



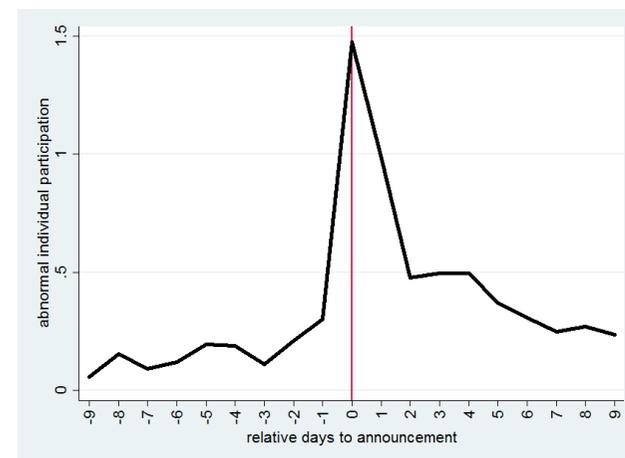
(a) Abnormal Trading Volume



(b) Abnormal Trading Volume - Spot vs. Margin accounts



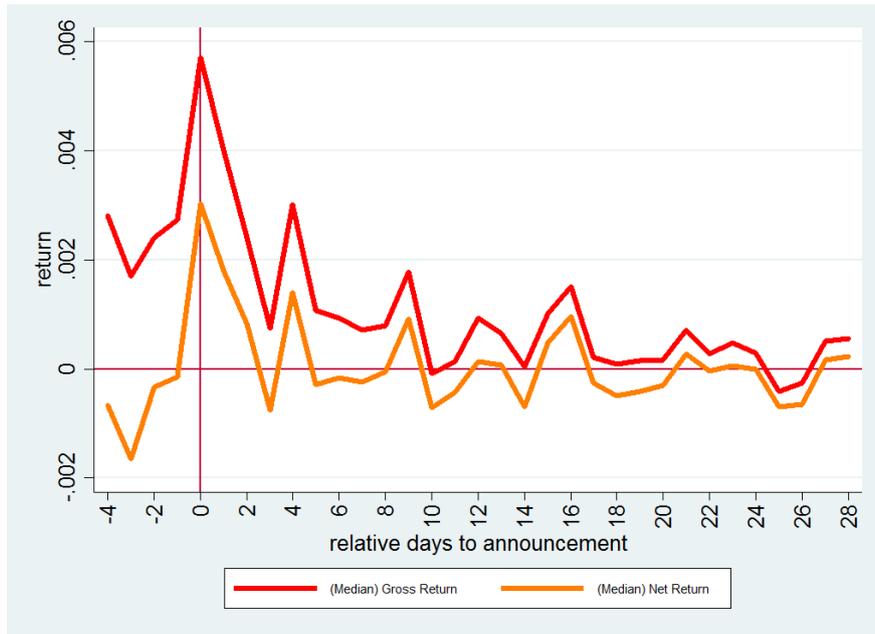
(c) Abnormal Trading Volume: Limit vs. Market orders



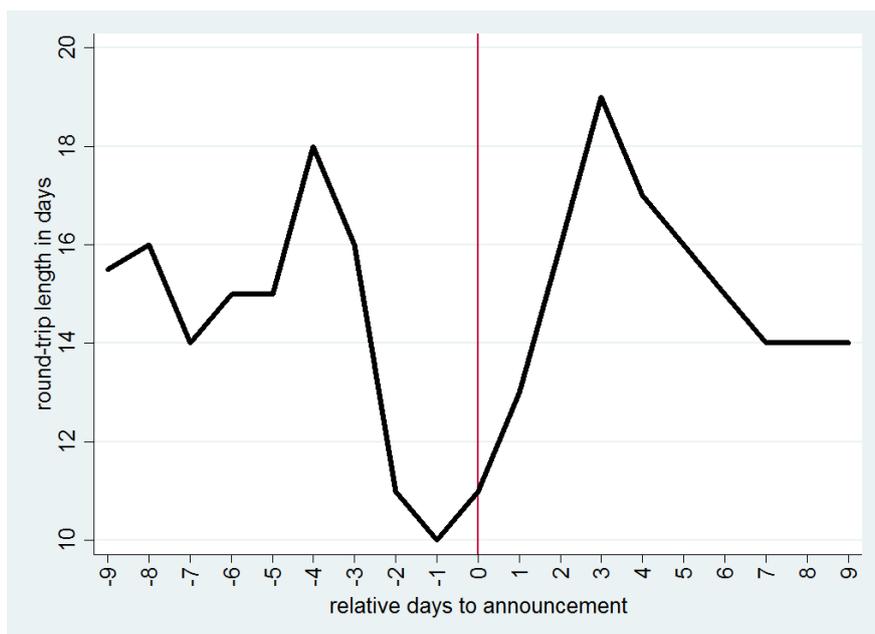
(d) Abnormal Number of Unique Investors

## Figure 3.2: Profit-Taking patterns on and after the announcement date

I recover the round-trips executed by retail investors around any earnings announcement day. A round-trip starts and ends with a zero net cumulative quantity. Round-trips are thus trading positions that are fully reversed: stocks previously bought are entirely sold and stocks previously sold short are completely bought back. In subfigure a, I restrict my sample to round-trip that are *initiated* during the time interval  $[-5,-1]$  relative to the next announcement day. Then I plot (on the y axis) the corresponding mean gross and mean net abnormal daily return *conditional* on *reversing* the round-trip on a specific day (on the x axis) relative to the announcement. On subfigure b I show how the median round-trip duration (on the y axis), defined as the time in days between the opening and closing time of any round-trip, changes around the earning announcement day.



(a) Profit taking



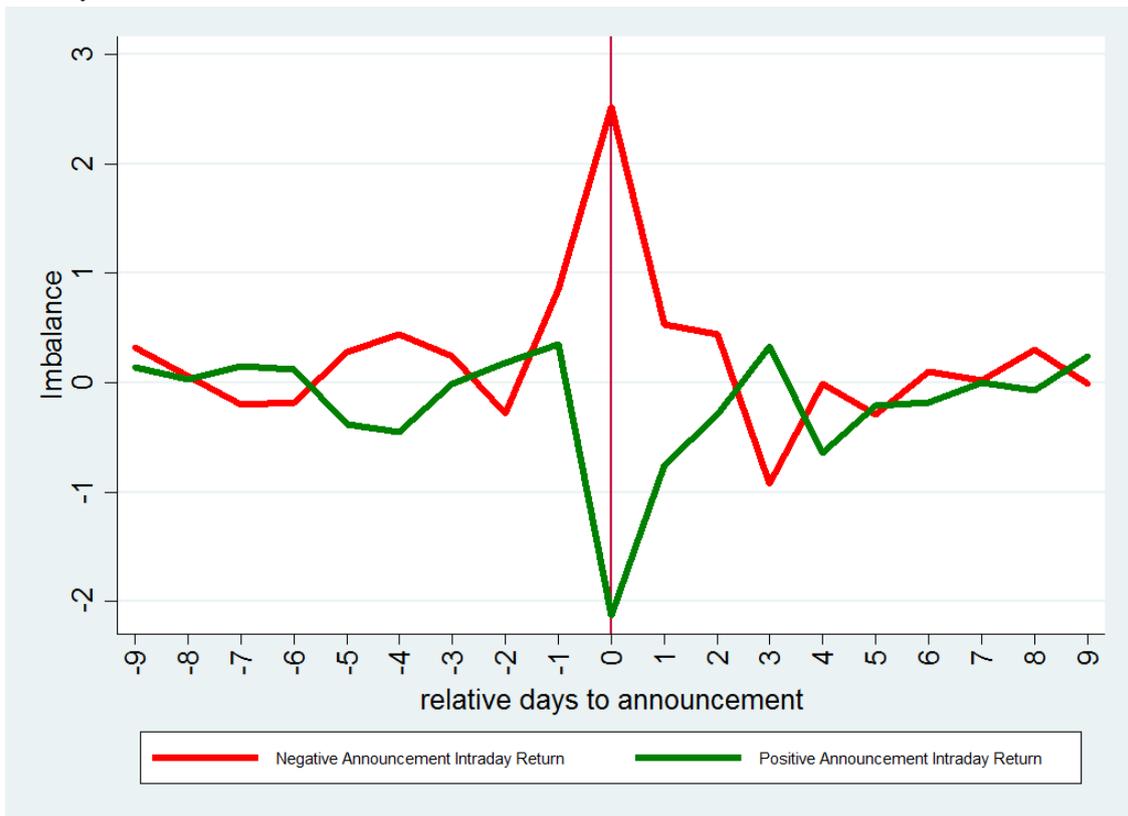
(b) Round-Trip Duration

### Figure 3.3: Net Individual Imbalance around earnings announcements

I show in this Figure how the abnormal individual imbalance on a given stock, defined as (Abnormal Buys - Abnormal Sells), changes around the earnings announcement date. A high and positive abnormal net trading imbalance value on a given day expresses an abnormally high buying activity on that stock, that day. I classify the announcement events in quintiles according to the *intraday* return

$$R_{\text{intraday}} = \frac{\text{closing price} - \text{opening price}}{\text{opening price}}$$

on the announcement day. Here, I use events in the first quintile (very low intraday return) and fifth quintile (very high intraday returns).



## 3.7 Tables

### Table 3.1: Summary statistics of the brokerage dataset

This table describes the main characteristics of the brokerage dataset used in this paper. The table gives the corresponding number of trades, percentage and cumulative percentage for each category of information (investors' order choices, investors' use of trading channels, investors' use of asset classes and investors' use of trading exchanges). The total number of trades may be different from one category to another due to missing values. The sample period is from 1999 to 2010.

Panel A: Order Choice			
	Frequency	Percentage	Cum. Percentage
Limit Order	6434646	62.6	62.6
Market Order	2918369	28.4	91.0
Others/Missing	920520	8.96	100
Total	10273535	100	

Panel B: Trading Channels			
	Frequency	Percentage	Cum. Percentage
Web	5785047	42.1	42.1
Online+	4103445	29.8	71.9
Trader+	2049605	14.9	86.8
Minitel	782961	5.69	92.5
Telephone	686728	4.99	97.5
Phone+	181141	1.32	98.8
Manual	165430	1.20	100
Total	13754357	100	

Panel C: Asset Classes			
	Frequency	Percentage	Cum. Percentage
Common Stocks	12986031	86.2	86.2
Others	824604	5.47	91.6
Equity Warrants	488490	3.24	94.9
Index Warrants	365357	2.42	97.3
Turbo Warrants	270680	1.80	99.1
ETF	134270	0.89	100.0
Total	15069432	100.0	

Panel D: Trading Exchanges			
	Frequency	Percentage	Cum. Percentage
NYSE Euronext Paris	13888529	92.16	92.16
Tradegate	669061	4.44	96.60
Others	437689	2.90	99.51
Nasdaq	74153	0.49	100.00
Total	15069432	100.00	

Nb. of distinct investors	145801
Nb. of buy order	7910252
Nb. of sell orders	7159180

## Table 3.2: Summary Statistics - Sample

This table shows, each year, the number of buy and sell orders in my sample.

<u>Year</u>	Sell	<u>Buy/Sell</u> Buy	Total
1999	18897	18733	37630
2000	45746	50690	96436
2001	198794	232232	431026
2002	169410	219365	388775
2003	152744	164381	317125
2004	150432	163514	313946
2005	97728	96733	194461
2006	264737	260801	525538
2007	203972	208670	412642
2008	305087	350474	655561
2009	308644	333582	642226
2010	282496	306860	589356
Total	2198687	2406035	4604722

### Table 3.3: Summary Statistics - Announcement Time

This table shows, each year, the number of orders executed (and the corresponding row frequency) on an earnings announcement day, separately for announcement events where the announcement has been made before the market opening and for announcement days where its has been made after the market closure.

<u>Year</u>	<u>Announcement Time</u>		Total
	Before Market	After Market	
1999	76.8 390	23.2 118	100.0 508
2000	44.2 278	55.8 351	100.0 629
2001	87.5 6614	12.5 948	100.0 7562
2002	71.7 6613	28.3 2604	100.0 9217
2003	75.2 4877	24.8 1609	100.0 6486
2004	84.1 7086	15.9 1341	100.0 8427
2005	70.2 3415	29.8 1452	100.0 4867
2006	89.6 14068	10.4 1635	100.0 15703
2007	83.6 9577	16.4 1875	100.0 11452
2008	95.0 14750	5.0 779	100.0 15529
2009	98.1 15100	1.9 291	100.0 15391
2010	94.5 14821	5.5 868	100.0 15689
Total	87.6 97589	12.4 13871	100.0 111460

### Table 3.4: Identification of Trading Strategies

I provide here an illustration of the round-trip identification used in this paper. In the example below, the investor completes two round-trips on the same security ABC. The first one started on January 1 and ended in January 18. The second one consists of a buy of 5 ABC on February 20 followed by a sell of 5 ABC the next day. Quantity stands for the quantity of stocks purchased or sold, Cash-Flow stand for the amount of money (in euro) actually received (in case of a sell) or paid (in case of a buy) by the investor. Fees is the daily amount of commissions fees incurred by the trader. CumQ keeps track of the inventory level on security ABC. As I identify two round-trips, the inventory level (or cumulative position held by the trader) revert accordingly to 0 two times.

Id	Stock	Date	Order	Quantity	Cash Flow	Fees	Daily Delta	CumQ
1	ABC	01/01/2007	BUY	10	-500	4,5	10	10
1	ABC	15/01/2007	BUY	11	-510	4,5	11	21
1	ABC	16/01/2007	SELL	10	550	4,5	-10	11
1	ABC	18/01/2007	SELL	11	560	4,5	-11	0
1	ABC	20/01/2007	BUY	5	-200	5	5	5
1	ABC	21/02/2007	SELL	5	170	4,5	-5	0

### Table 3.5: Round-Trips

This table shows the main summary statistics for the round-trips executed by investors in my sample. Gross and Net Cash are the actual Euro amount earned or lost after unwinding a round-trip. The net cash computation adds the real transaction fees involved in a given round-trip. The round-trip length is the duration in days between the opening and the closing of a round-trip. Gross and net returns are defined similarly. Number of trades per round-trip corresponds to the number of executed daily-trades in a given round-trip; the minimum number of day-trades is two (one buy and one sell day-trades).

	mean	sd	p25	p50	p75
Gross Cash (euro)	-46.15656	2139.018	-107.0205	28.2998	169.8499
Net Cash (euro)	-102.8993	2161.354	-144.2997	4.439074	131.75
Round-trip Length (days)	117.1709	324.7717	3	14	68
Round-trip gross return	-.0042238	.1947259	-.0342474	.0082391	.050471
Round-trip net return	-.0167317	.1946513	-.0451988	.001205	.0389986
Nb. of Trades per Round-trip	3.224945	6.748087	2	2	3
<i>N</i>	351459				

**Table 3.6: Abnormal Trading Activing around Earnings Announcements**

This table shows the results of the following regression:

$$AbnX_{s,t} = \alpha + \sum_{i=-5}^{+5} \beta_i * I_{t=i} + \theta_t + \gamma_s + \epsilon_{s,t},$$

where the abnormal measure  $AbnX_{s,t}$  is the abnormal buy-sell imbalance, the abnormal daily volume in numbers of executed trades, the abnormal trading volume executed on cash accounts, the abnormal trading volume executed on leveraged margin and cash accounts, the abnormal trading volume of limit orders only, the abnormal trading volume of market orders only and the abnormal number of unique investors trading on stock  $s$  on day  $t$ . I include a full range of year fixed-effects  $\theta_t$  and stock fixed effects  $\gamma_s$ . The coefficients Before(i) After(i) and Announcement Day corresponds to the  $\beta_i$  above and capture the effect of a particular day around the announcement date on my abnormal trading measures. \*, \*\*, \*\*\* denote significance at the 10%, 5% and 1% levels.

	Abn Imb.	Abn Volume	Abn Margin	Abn Spot	Abn LO	Abn MO	Abn ID
Before(5)	0.120 (0.144)	0.000854 (0.0421)	0.0854 (0.0730)	-0.0691 (0.0502)	-0.0226 (0.0241)	-0.0229* (0.0121)	-0.0848** (0.0411)
Before(4)	-0.00468 (0.244)	0.0303 (0.0510)	0.0360 (0.0890)	0.0667 (0.0648)	0.00652 (0.0286)	-0.0141 (0.0131)	-0.000632 (0.0501)
Before(3)	0.103 (0.149)	0.00402 (0.0362)	0.0323 (0.0659)	-0.0250 (0.0435)	0.00600 (0.0240)	0.00988 (0.0132)	0.0215 (0.0392)
Before(2)	0.147 (0.131)	0.0418 (0.0326)	0.118** (0.0561)	0.0340 (0.0383)	0.0240 (0.0210)	0.0196* (0.0115)	0.0611* (0.0340)
Before(1)	0.603*** (0.123)	0.167*** (0.0325)	0.324*** (0.0555)	0.167*** (0.0391)	0.0997*** (0.0203)	0.0410*** (0.0101)	0.183*** (0.0322)
Announcement Day	0.613*** (0.202)	1.105*** (0.0470)	1.735*** (0.0858)	1.196*** (0.0574)	0.595*** (0.0286)	0.279*** (0.0142)	1.141*** (0.0475)
After(1)	-0.183 (0.164)	0.740*** (0.0444)	1.196*** (0.0843)	0.796*** (0.0539)	0.455*** (0.0278)	0.200*** (0.0144)	0.732*** (0.0451)
After(2)	-0.168 (0.207)	0.473*** (0.0524)	0.600*** (0.0881)	0.403*** (0.0602)	0.278*** (0.0315)	0.0836*** (0.0152)	0.455*** (0.0513)
After(3)	-0.137 (0.222)	0.340*** (0.0592)	0.378*** (0.100)	0.465*** (0.0714)	0.200*** (0.0344)	0.0826*** (0.0181)	0.421*** (0.0612)
After(4)	0.00305 (0.164)	0.307*** (0.0449)	0.349*** (0.0733)	0.431*** (0.0582)	0.183*** (0.0282)	0.126*** (0.0163)	0.376*** (0.0479)
After(5)	-0.211 (0.129)	0.258*** (0.0395)	0.230*** (0.0614)	0.315*** (0.0472)	0.141*** (0.0247)	0.0636*** (0.0122)	0.260*** (0.0383)
Constant	-0.715** (0.328)	0.494*** (0.0595)	0.309*** (0.0754)	0.432*** (0.0641)	-0.189*** (0.0411)	-0.757*** (0.0201)	0.279*** (0.0576)
Nb observations	51032	62707	42905	62767	81927	82035	64264
R-Square	0.020	0.067	0.063	0.054	0.136	0.123	0.061

## Table 3.7: Trading performance around Earnings Announcements

This table shows the results of the following regression:

$$Y_{i,t} = \alpha + \beta_1 * \text{Before}_{i,t} + \beta_2 * \text{Announcement}_{i,t} + \beta_3 * \text{After}_{i,t} + \delta_i + \theta_t + \epsilon_{i,t},$$

Where  $Y_{i,t}$  is, in the first six columns of this table, a performance measure computed on the round-trips executed by investors in my sample. Each measure (including the risk-adjusted ones) is shown twice: with or without the transactions fees incurred. Horizon is the length in days of the executed round-trips of less than 100 days.  $\text{Before}_{i,t}$  (resp.  $\text{Announcement}_{i,t}$ ,  $\text{After}_{i,t}$ ) is a dummy variable that is one only if  $t = -1$  (resp.  $t=0$ ,  $t=+1$ ) relative to the next earning announcement. I include a full range of year fixed-effects  $\theta_t$  and individual fixed effects  $\delta_i$ . The coefficients  $\beta_i$  capture the effect of a particular day around the announcement date on the individual trading performance. Standard errors are clustered at the individual level and shown in parenthesis. Returns are multiplied by 100. \*, \*\*, \*\*\* denote significance at the 10%, 5% and 1% levels.

	Gross return	Net Return	Gross Abn Return	Net Abn Return	Gross Abn (CAC)	Net Abn (CAC)	Horizon
Before	0.116*** (0.014)	0.0760*** (0.014)	0.126*** (0.014)	0.0878*** (0.014)	0.116*** (0.013)	0.0830*** (0.014)	-2.170*** (0.23)
Ann. Day	-0.0121 (0.0083)	-0.0313*** (0.0082)	-0.0121 (0.0082)	-0.0293*** (0.0080)	0.00692 (0.0080)	-0.00989 (0.0079)	-0.458*** (0.16)
After	-0.0160 (0.010)	-0.0237** (0.010)	-0.0197** (0.0099)	-0.0253** (0.0099)	-0.00929 (0.0096)	-0.0164* (0.0098)	-0.405* (0.21)
Constant	-0.0161 (0.0098)	-0.107*** (0.0099)	-0.0260** (0.011)	-0.120*** (0.011)	0.0208** (0.0095)	-0.0681*** (0.0096)	29.14*** (0.91)
Nb observations	344313	341886	335261	332798	343659	341315	280407
R-Square	0.003	0.002	0.003	0.002	0.002	0.001	0.011

# Executive Summary

This thesis is made of three distinct chapters. In the first chapter, I test whether the display format of financial information matters for the individual investor. I find that a more efficient information display allows investors to increase returns on their limit orders, because it becomes easier for them to mitigate the risk of adverse selection when trading with those orders. My findings suggest that retail investors have bounded rationality. In the second chapter I test whether liquidity provision to the market can be a profitable strategy, after fees, for active retail investors. I find that only individuals ranked in the top decile of performance can persistently beat the market using highly contrarian limit order strategies. Limits-to-arbitrage seem to explain why these top retail investors exploit trading opportunities before other more sophisticated arbitrageurs. In the third chapter, I study the retail trading strategies around stock earnings announcements. I find that round-trips started one day before an announcement are more profitable and much shorter in duration than those started during the non-announcement period. Retailers reverse their winning trades on the event date, which can slow down the adjustment of prices to new information.

**Keywords:** *individual investor, bounded rationality, limits-to-arbitrage, trading strategies*

Cette thèse consiste en trois chapitres distincts. Dans le premier chapitre, je teste l'hypothèse selon laquelle le format d'affichage de l'information financière affecte les décisions des investisseurs individuels. Je montre qu'un affichage plus efficace permet aux individus de mieux gérer leurs ordres à cours limité en minimisant le risque de sélection adverse encouru en utilisant ces ordres. Cela suggère que les investisseurs individuels ont une rationalité limitée. Dans le second chapitre, je teste si les stratégies de trading apportées de liquidité peuvent générer des profits, après coûts de transactions, pour les traders actifs qui les implémentent. Je montre que seuls les individus situés dans le plus haut décile de performance peuvent battre le marché de façon persistante en utilisant des stratégies hautement contrariantes qui nécessitent l'utilisation massive d'ordres à cours limité. Les limites-à-l'arbitrage semblent expliquer ce phénomène. Dans le troisième chapitre, j'étudie les stratégies des individus autour des annonces de résultats. Je montre que les aller-retours qui sont implémentés un jour avant une annonce génèrent en moyenne des profits plus élevés et sont plus courts en durée que ceux implémentés en temps normal. Les individus clôturent leurs positions gagnantes le jour de l'annonce, ce qui peut ralentir l'ajustement des prix suite à l'annonce.

**Mots-clés:** *investisseur individuel, rationalité limitée, limites à l'arbitrage, stratégies de trading*